Fostering Balanced Contributions among Children through Dialogue Visualization

Mehmet Celepkolu, Aisha Chung Galdo, Kristy Elizabeth Boyer

***Abstract*—As children develop conversational skills such as taking turns and openly listening to ideas, they often experience conflicts and inequity within collaborative dialogue for learning. Previous research suggests that increasing children's awareness about their own behaviors during collaboration may help them adjust their behaviors and become better partners. Despite this promise, there are currently no educational technologies designed to support children in visualizing and reflecting on their collaborative dialogues. This paper reports on an application that generates interactive visualizations of children's dialogue illustrating their word counts, questions counts and types, dialogue content, keywords from their dialogue, and a video recording of their interaction. We evaluated the application by conducting a study with 20 children who were completing computer science (block-based coding) tasks collaboratively and examined how they changed their dialogues in a subsequent dialogue after interacting with the visualizations of their dialogues. Results show that after viewing their dialogue visualizations, children engaged in more balanced dialogues and that less-engaged students talked more and asked more questions. This research provides evidence that dialogue visualization tools have a great potential for supporting young learners as they deeply think about their own dialogue and improve their collaborative behaviors.**

***Index Terms*—Computer Supported Collaborative Learning, Collaborative Dialogue, Data Visualization, Middle School**

I. Introduction

Collaborative learning has numerous benefits for children such as increasing engagement in the learning process [1], [2], improving critical thinking skills [3], [4], and developing social and communication skills [5]. However, children do not always follow productive collaboration norms such as being open to different ideas, taking turns, and asking meaningful questions [6]. Previous research suggests that the type of dialogue during teamwork impacts the success of the teamwork [7], [8] and the dialogue between students should be like a dance of voices and perspectives [9]. Bakhtin's dialogic theory [10] also highlights the importance of a balance between partners and defines dialogue as an active double-voiced discourse, where sides do not "dominate the other's thought." If equal participation is not encouraged and children are left to collaborate without any structure, the interaction may result in some children dominating the conversation or some children becoming less active in the learning task [11]. These challenges can lead to conflicts and marginalization among partners [12], and can discourage students from working in groups for future activities [13]. While "more balanced" dialogues are not always more productive or more effective, dialogue balance indicates each team member is actively contributing to the conversation [14], [15], and their opinions are not neglected [16].

A large body of research on group interaction has established that additional scaffolding can improve students collaborative behaviours [17]. One effective scaffolding approach is utilizing *mirroring* tools that reflect data about students' interactions back to them (*e.g.*, visualizations of student contributions) [18], [19]. These tools provide useful information to the learners, through which they can monitor and adjust their collaborative behaviors [20], [21]. Previous studies conducted with adult learners provided promising evidence that these applications can help learners balance their dialogue, enjoy the learning process and increase their learning outcomes [22], [14]. However, current collaborative dialogue visualization technologies are mostly designed for adult learners, and their benefits for younger populations are still not known.

To address this research gap, our previous work investigated children's needs and expectations from dialogue visualization applications: we conducted a series of iterative design based research studies, and co-created a dialogue visualization application prototype for children [23]. The findings in our prior paper established that children perceive the dialogue visualizations as valuable and are eager to use them in their collaborative learning activities. Compared to existing systems that mostly visualize high level features, such as word counts, our visualization application utilizes computational methods and natural language processing techniques to extract semantic information from the content of the dialogue, such as question types.

As shown on Figure 1, the interactive visualization application has seven components: (1) a time-series line chart illustrating the moving average of the word spoken by each user during the activity, (2) a dialogue content section showing the transcription of the children's dialogues during the activity, (3) a pie chart illustrating the total number of words spoken by each user, (4) a pie chart illustrating the total number of questions asked by each user, (5) question types (open and closed) generated by machine learning models, (6) automatically generated keywords from children's dialogue, (7) and video recording of their activity interaction.

Our overarching goal in this study is to examine the change in children's future dialogues after interacting with the visualization application. In particular, this paper focuses on the following research question: *"In what ways do children's collaborative dialogues change after interacting with their own dialogue visualizations?"*. To investigate this research question, we conducted a three-day study with 20 children (remotely, due to the COVID-19 pandemic) and examined how they changed their dialogues in subsequent activities after interacting with the visualizations of their dialogues. Children

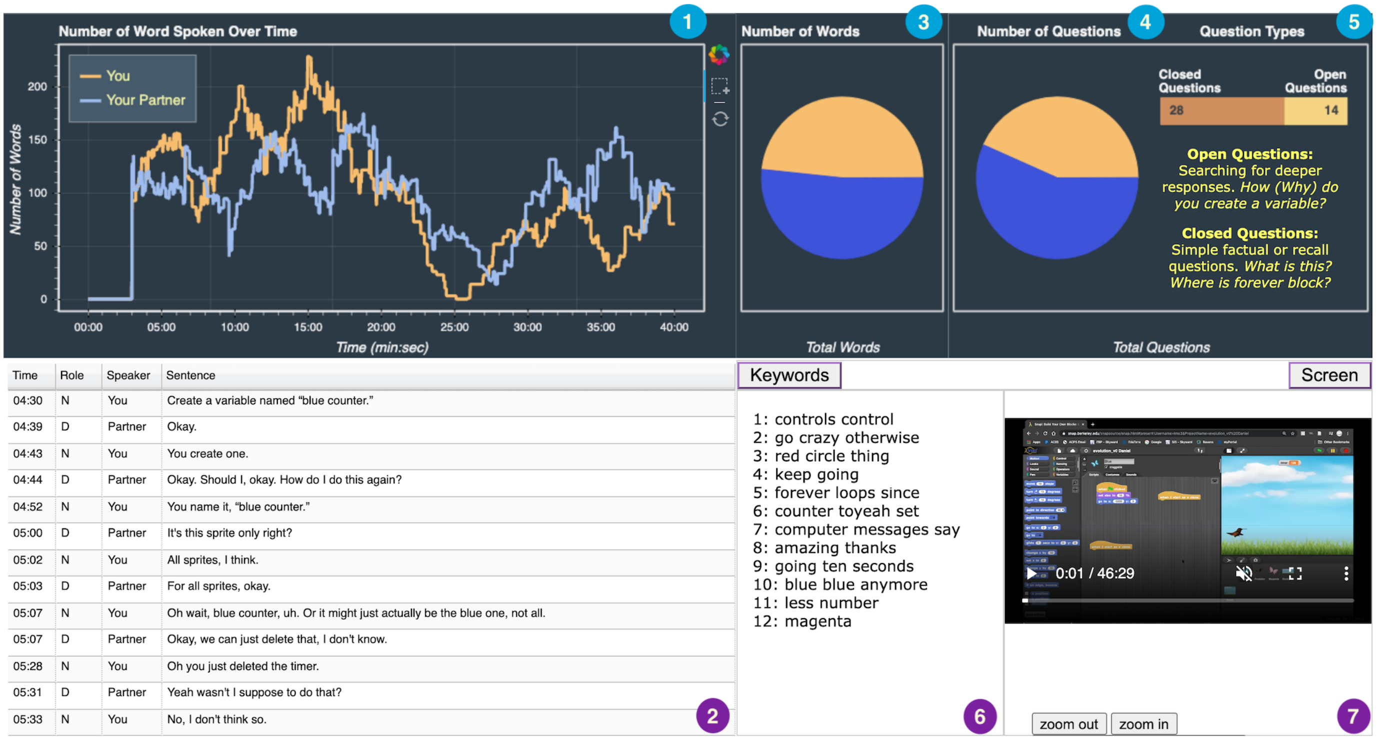

Fig. 1: Dialogue visualization application, which illustrates children's word counts, question counts, question types, dialogue content, keywords from their dialogue, and a video recording of their learning activity.

self-reported that they changed their dialogues in the final collaborative activity. Moreover, the quantitative analysis of student dialogues indicated that children engaged in more balanced dialogues, and that less-engaged students talked more and asked more questions after viewing their dialogue visualizations. This study provides evidence that dialogue visualization tools have great potential for supporting young learners to reflect on their own dialogue and improve future collaborative behaviors.

II. Related Work

This research builds upon a body of prior research on supporting learners during collaborative problem solving, implementing intelligent dialogue analysis techniques on learners' collaborative dialogues and using visualization applications to support collaborative problem solving. In this section, we first present the literature on collaborative dialogue and team dynamics, and then present the existing tools for supporting children's collaborative dialogue. Finally, we examine the existing dialogue visualization applications and present the empirical results, where available.

A. Collaborative Dialogue: Opportunities and Challenges

A wide body of research reports numerous benefits of collaborative dialogue for children, such as exchanging information [24] and learning new perspectives [3]. However, not all collaborative interactions produce desirable outcomes, and the type of dialogue during interaction defines the effectiveness of the collaboration [7], [8]. For example, the "Exploratory Talk" model [25], [26] suggests that in a productive collaborative dialogue, each member of the team adds relevant information to dialogue; members make sure they reach consensus and are on the same page before moving the other tasks; and they value each other's knowledge and skills. A similar model, "Accountable Talk" [27], suggests that team members should "use evidence to support your opinions, ideas, predictions, and inferences." In this model, there are three types of accountability: First, participants interact and create a contribution in response to others (accountability to the learning community), in which the talk can be on elaborating someone else's argument or requesting others to elaborate an argument. Second, the talk focuses on logical and reasonable connections (accountability to accepted standards of reasoning), which involves explanation and self-regulation (or correction) rather than simply supporting or rejecting proposals. Third, the facts are checked (accountability to knowledge) and evidence for claims is provided. This process helps in revealing and repairing misconceptions.

Previous literature also suggests that dialogue and engagement are both built on the same fundamental components: interactivity, responsiveness, and relationship initiations [28]. Kent and Taylor [29] put forth the principle of dialogic propinquity. According to this principle, the participants must be actively present and accessible during the interaction. A productive collaboration requires more than participants being present during the interaction [30], and the type of dialogue is important for successful collaborative learning [7].

B. Supporting Children's Collaborative Dialogue

Previous research has often emphasized the importance of providing scaffolding during collaborative learning activities for better outcomes [17]. In line with this finding, there are various technologies for supporting children's collaborative dialogues. For example, Melonio and Rizvi [31] presented TurnTalk, a tangible interactive device for orchestrating turn taking actions among children in group conversations. It is a pentagon-shaped tabletop device that flashes green and red lights to teach children about turn taking dynamics and moderate over- or under-participation. TurnTalk does not perform any analysis on children's dialogues; it only aims to teach primary school children some specific conversation norms, such as taking turns during conversations. In a later study, Gennari et al. [11] developed ClassTalk, an extension of TurnTalk designed for class conversations, and conducted a classroom study with 102 primary school children to evaluate the usability of the tool and children's engagement. The results showed that children could easily understand the purpose of the tool and their engagement in the conversation increased. Different than TurnTalk and ClassTalk, our visualization application does not act as a conversation mediator; rather, it provides information to increase children's awareness of their own dialogue and help them reflect on their collaborative behaviors. Additionally, our application analyzes children's dialogues and provides interactive visualizations about the content of these dialogues.

C. Existing Dialogue Visualization Applications

Visualizations, as opposed to verbal or numeric representations, facilitate an easier, structural way to make sense of information [32]. Visualizations have widely been used in learning environments to support collaborative learning sessions in classrooms [33], increase students' awareness for paying attention to their own collaborative behavior, [34], support adviser-student dialogues during meetings [35], and increase retrospective awareness of learners' emotions in online learning [36]. There have also been various visualization applications for orchestrating group conversations and raising learners' awareness about their actions; yet, these technologies have mainly been created for adult learners [37], [38], [15]. For example, Dimicco et al. [14] presented participation visualizations to adult learners (mean age = 25) during team activities, and the results showed that the visualizations led over-participators to decrease their comments while it did not make any changes in under-participators' participation levels. In another study, Kim et al. [39] extracted various features such as length and speed of talking, number of turns taken, and average turn length from adult learners' dialogues and provided these features to the learners as feedback. The results showed that the feedback made the participants more cooperative and increased their performance in group work.

Despite the promising empirical results for adult learners, the potential benefits of dialogue visualizations for children are still not known. Most visualization studies conducted with children have only focused on how children understand and generate visualizations [40] or how they use them for visualizing scientific processes [41]. A recently conducted systematic review [42] on mirroring tools and interaction indicators in collaborative work also did not report any visualization applications for supporting children's collaborative dialogue. Our initial studies provided evidence that dialogue visualizations serve children well and that children are capable of reflecting and using mirroring tools [23]. In this study, we extend this body of research by reporting quantitative evidence for how children change their dialogue after interacting with visualizations illustrating their own collaborative dialogue.

III. DIALOGUE VISUALIZATION APPLICATION

To investigate the research question, we have developed a novel web-based interactive dialogue visualization application with the Python Bokeh visualization library[1] and Flask web framework[2]. This section presents the design process, the design goals and principles, and finally a description of the dialogue visualization application features.

A. Design Process

We utilized a design-based research approach to iteratively develop and evaluate the usability of the dialogue visualization application with children. Throughout three iterations, we conducted individual think-aloud interviews to examine whether the information illustrated via visualizations was understandable by children, investigated whether they found the visualizations useful, explored their expectations from dialogue visualizations, and received their feedback for improvements.

In the first iteration, we created two different types of visualizations: Three pie charts illustrating the (1) total number of times a child spoke, (2) total number of words they spoke with their partner, and (3) total number of questions they asked during the activity. After designing the application, we conducted a think-aloud study with 18 middle school students to receive their feedback, which informed the design of the new features in the second iteration.

In response to the feedback we received in the first iteration, we re-designed the line charts and the pie chart, and added video recordings of the collaborative problem solving activity and student interactions. Then we conducted another user study with 18 middle school students. We placed all the visualizations on the same page, and connected the line chart and the pie charts. When a child highlighted a sub-region of the line chart, the pie chart was updated to reflect counts from the highlighted region in a dynamic way. The purpose of this approach was to allow children to benefit from both types of charts and give them the freedom of exploring the same data on their preferred chart. We also added the screen recording of the children's coding activity and the video recordings of children's group interactions with each other. Our previous work describes the design process of the first two iterations in detail and presents the qualitative investigation of children's understanding and expectations from the dialogue visualizations [23].

[1]https://bokeh.pydata.org/en/latest/
[2]https://palletsprojects.com/p/flask/

In the last iteration of the application, the focus of this article, we included new features, more specifically related to the content of the dialogue based on children's suggestions in the second iteration. During the think-aloud sessions, some students had suggested that the number of questions only present some shallow information without providing any details about the quality of the question; thus, they wanted to see more information about the "types of the questions" asked during their interactions. As a result, we added a bar chart illustrating the frequency of question types (*i.e.*, open versus closed). In addition, children suggested providing keywords extracted from the dialogue to make it easier to navigate through the dialogue content. We added a new section presenting a list of the most related keywords based on the dialogue part selected by the student. We provide the implementation details for each of these features in the Dialogue Visualization Application Features section.

B. Design Goals and Principles

Throughout the application design process, we continuously explored ways to improve usability of the application and provide a smooth experience for the children while interacting with the application. More specifically, our design processes were centered around the following six main design principles: (1) interactivity, (2) simplicity, (3) self-explanatory labels, (4) aesthetics and ease of use, (5) personalized experience, and (6) multiple types of information presentation.

1) **Interactivity:** Interactivity has been shown as one of the most important design principles for improving user satisfaction on web based applications, which attracts user attention and creates a playful feeling toward the application [43]. Aligned with this idea, the students also often indicated in the think-aloud sessions that the visualizations should be dynamic and interactive, like a game. As a result, the application updates all the charts based on the selected sub-region of the line chart, which allows users to focus on different parts of the dialogue in a dynamic and interactive way.
2) **Simplicity:** We have prioritized the simplicity of the application and attempted to ensure that each feature presented on the application is easily understandable by children. As suggested by interaction design research [44], the application presented the information in a simple way but relevant, provided direct and fast access, made the visuals clear and easy to understand.
3) **Self-Explanatory Labels:** Previous research emphasizes the importance of self-explanatory labels that provide detail on the information presented [45]. Given the possibility that most children might have never used such application before, we added detailed information and labels for each chart. During the previous think-aloud sessions, children also often used the labels to understand the purpose of the visualizations presented.
4) **Aesthetics and Ease of Use:** Children are often more receptive to vivid colors and our application used various contrasting colors to appeal the users as well as to make it easy to use. The application provides a smooth experience (without lagging) while interacting with the features. When students select a certain area on the line chart, all the other visualizations are automatically updated accordingly in less than a second.
5) **Personalized Experience:** Each child sees the dialogue from their own perspective. We use the labels "You" and "Your Partner" to increase the personalized experience of the visualizations for each user.
6) **Multiple Types of Information Presentation:** During the previous think-aloud sessions, the children expressed their desire to see different types of visualizations for the same information. For example, we used line charts as well as pie charts to illustrate the word counts because children sometimes prefer different charts to examine the word counts. We also included both dialogue content and the keywords to navigate through the content of the dialogue.

C. Dialogue Visualization Application Features

Dialogue between children provides a rich data source and there is a great deal of information that can be extracted from dialogue and be visualized for the users. However, there is a fine balance between what dialogue features can be visualized and how much of that can be understood and used by children. The complexity of the presented information needs to be adjusted based on their knowledge of visualizations so that they can comprehend the information presented. If the presented information is not tailored to their understanding, children may not be able to utilize the visualizations to reflect on their dialogue. During the iterative design process, we deeply examined children's suggestions related to the application and developed seven main features which provided useful information for children about their interaction with their partner. In this section, we present a detailed description of each application feature and the logical reasoning behind it.

1) Time Series Line Chart Illustrating Dialogue Flow Over Time: The purpose of this visualization is to allow children to observe how their dialogue evolves over time. Temporal dynamics in dialogue flow has been suggested as an essential aspect of dialogue data representation [46], and cumulative charts, such as pie charts, may obscure sequential information [43], [47]. We generated the time series-based line chart using a sliding-window approach [48] showing the number of words spoken over time. The x-axis of the chart showed the time and y-axis showed the number of words spoken.

2) Dialogue Content: We also integrated a dialogue exploration functionality connected to the line chart, in which the user selects certain areas of the line chart and examines the transcription of the their dialogue with their partner below. The dialogue content feature has four columns: (1) The *Time* column shows the timestamp of the beginning of each speaker's utterance. (2) The *Roles* column shows children's pair programming roles (driver or navigator) during the activity at a given time. (3) The *Speaker* column shows who spoke at a given time. (4) The *Sentence* column shows the dialogue between children. The dialogues are generated from the manually transcribed (verbatim) videos, and we presented the dialogue without making any changes.

3) Word Counts Pie Chart: When using the previous iterations of the application, the majority of children found the time series line charts more informative than the summative pie chart for illustrating the number of words spoken. However, some students expressed their desire to have access to both chart types, as pie charts show the total number of words in a simpler way. As a result, we decided to keep both versions to provide alternative ways to convey similar information. The pie charts in our visualization tool have interactive features, and children can see the details such as total number of words for each speaker when they hover over the charts.

4) Question Counts Pie Chart: Questions play a critical role keeping the collaboration process active by indicating interest in the topic and taking a first step to access information and resolve confusion [49], and question asking has been shown as an important sign of productive dialogue [50]. The question count was presented in a pie chart format, showing the total number of questions for a given duration of time (as selected on the line chart). These questions are manually extracted from the dialogue transcriptions.

5) Question Types Bar chart: During the think-aloud interviews, children often suggested that they would like to see more information about the content or type of the questions because the question count pie chart does not distinguish between low and high level questions. Aligned with the "simplicity" design principle, we classified children's questions into two simple categories: open and closed questions. Closed questions are similar to factual recall questions, which have clear and obvious answers. Closed questions usually begins words such as 'Do', 'Can', or 'Will' and they do not necessarily have a 'yes' or 'no response; yet, the responses are usually very short and aim to obtain a specific answer. For example, "Is that a variable?" and "Can you change the color of that circle?" are closed questions. These questions usually lead to limited answers with no expectation for deep responses. On the other hand, open questions lead to more dialogue. These questions do not have simple responses and require children to perform higher level thinking, such as providing logical justifications for a claim, explaining concepts with their own words, describing a process step by step, or comparing and contrasting different options. Some sample open questions are, "How do you create a variable?", "What happens when you change the forever block with an if-else block?", and "How about changing this part of the code completely?". Even though both types of questions can play an important role, open questions may be more valuable because they stimulate creativity and inquiry [51].

The question classification approach we used in this application is a form of text classification, which we created with supervised machine learning models. We had previously conducted many studies with middle school students in which we collected and transcribed 38 pairs' interactions. Given that this older data set was collected from children working on similar activities, we used it to train machine learning models to classify and display the question types to children for the final study. Our models achieved 0.87 accuracy (Precision: 0.85, Recall: 0.9, F1: 0.87), which indicated that our automated model classified about 87% of questions correctly. After creating the machine learning model with acceptable results, we saved the best machine learning model and created a question classification API that would categorize new questions based on this trained model. The purpose of this API was to classify the children's questions automatically in the next group work. The technical details of the classification task are beyond the scope of this paper and we omit further details in this paper.

6) Keywords: During the previous think-aloud interviews, children often expressed the need for creating visualizations that help them navigate through the dialogue and provide them some ideas about the semantics of the dialogue. Even though the dialogue content feature allows children to go through the verbatim content of the dialogue, some children indicated that some keywords extracted from the dialogue can help them first scan through the keywords and then focus on the dialogue content when needed. For a smooth user experience (without lagging), we implemented a lightweight keyword extraction algorithm, RAKE (Rapid Automatic Keyword Extraction) [52], and used a pre-trained Word2Vec model [53] to group the semantically similar keywords in order to rank them for better user exploration. We needed to combine these two methods because RAKE does not distinguish between semantically meaningful words and ignores the semantic relationship of the words, which can result in generating two similar keywords or keyphrases such as "difficult project" and "challenging assignment". Second, it does not pre-process (*e.g.*, stemming) the words, which can generate two very similar keywords or keyphrases such as "computing science" and "computer science". As a solution, we utilized a Word2Vec model by generating the vector representations of each pre-processed keyword and performed a pairwise cosine similarity of all the candidate keyword vectors to identify the similarities of keywords. Next, we merged the similar keywords and displayed the highest ranked keyword as the final keyword.

7) Video Recording of the Collaborative Activity: Video recording of the activity allows children to see what they had done in a specific part on the activity. We integrated this feature and added buttons to allow users to increase/decrease the size of the video based on their preference.

IV. Methods

In this study, our overarching goal was to investigate how children change their future collaborative dialogue after seeing their own dialogue visualizations. This section describes the study we conducted with middle school students to examine the change (if any) in their dialogues after seeing the dialogue visualizations.

A. Participants

This study was conducted with middle school students in the southeast United States in Spring 2020. Due to the the COVID-19 pandemic, the study was conducted remotely through a video conferencing application (Zoom[3]). To schedule the activity meeting times with the children, we first met with the class teacher and talked about the possible ways to implement

[3]https://zoom.us/

the study. Based on the teacher's suggestion, we used an online meeting scheduling tool[4], which children had previously used to participate in some of their remote class activities. We provided all available times for the next three weeks (from 9am to 6pm). Our goal was to provide as many time slot options as possible since children might not be following their regular daily schedules during the COVID-19 pandemic. The teacher sent an online meeting scheduling link along with the online consent form to the children's parents.

26 children's parents consented for their children to join the study and 20 children participated in all three days of the study. Out of the 20 children, the gender distribution was 12 girls (60%), 7 boys (35%), and 1 unspecified (5%). Race/Ethnicity were Asian (60%), White (20%), Hispanic (5%), and other (15%). The mean age was 12.2 years with ages ranging from 11 to 13 years.

We randomly matched each child with another child based on their selected available times. Next, the teacher sent them a link to a Google doc, which contained information about the matched pairs, their meeting time and the Zoom remote meeting link. Children would simply go to the document and click the link to join the meeting.

Before we recruited the students, we obtained approval from the university's Institutional Review Board (IRB), which examined all the details of this study to protect the rights and welfare of participants. When we met with the students for the study, we verbally explained the key information in the consent form. Before each study, we obtained children's permission for video recordings and explained to them that the video recordings would only be used in scholarly work. If a child reported discomfort with audio/video recordings, we did not record them. There was no penalty for not participating in the studies and children were free to leave the activities anytime.

B. Study Context

This study was a part of a set of computer science learning activities implemented in a science classroom. The purpose of these activities was to help children learn the fundamental CS concepts (including variables, conditionals, loops, and object-oriented programming), and create science models based on lesson topics [54]. All the students in the science class had attended programming activities and developed various science models and simulations on topics such as light waves and evolution (Fig 2) before this study. Thus, they had the fundamental knowledge and skills to participate in more advanced CS+Science activities. However, none of these students had used or seen the dialogue visualization application before.

During the coding activities, the students followed the pair programming paradigm, which has been widely used in programming education [1], [2]. In pair programming, each student has a role: the driver controls the computer to implement the solution, and the navigator provides feedback and helps catch mistakes. During pair programming, students switch roles regularly.

[4]www.signupgenius.com

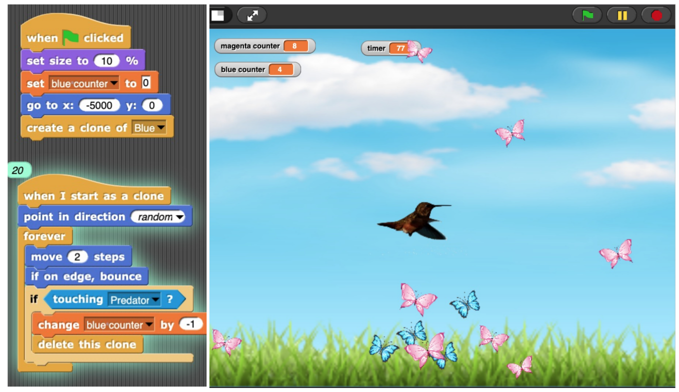

Fig. 2: Sample activity modeling concepts of evolution, created with the Snap*!* block based programming language.

C. Data Collection Procedure

We implemented the study in three sessions: In session 1, our goal was to make sure students were prepared to do the activities remotely and were comfortable with the Zoom video conferencing application. All the students had previously attended some remote class sessions with their class teacher and were familiar with Zoom. Additionally, we also showed them the basic functionalities that they would need to use during the coding activities (*e.g.*, screen sharing, switching driver/navigator roles). As part of the pair programming paradigm, only one of the children was allowed to make changes on the Snap*!* programming interface during a given time (driver) while the other child was providing feedback (navigator). During these activities, the driver would share his/her screen. To switch roles, the driver would save their progress on the project and send the link to the navigator who would open the project and continue from where his/her partner left off. This switching roles process took about one minute to complete. Students also learned a new coding concept in Snap*!*: cloning, similar to object oriented programming and recursion CS concepts, which they needed to use in the next remote coding activities. Finally, we also provided a reference guide summarizing the functionality of some blocks and CS concepts.

In session 2, students completed a CS+Science activity using pair programming, in which they coded a science model with the Snap*!* programming language. During each session, a researcher was available to answer children's questions and address possible technical difficulties. At the beginning of the session, the researcher first provided written instructions for the activity and then reminded the children to follow the pair programming paradigm. After the short introduction, the researcher turned off their camera/microphone but was always available to answer children's questions. The researcher did not involve themselves in the activity unless the children explicitly asked for help. Our goal was to allow children to have a natural interaction and collaborate on the problem together before asking the researcher for help. While some pairs asked for help a few times, some pairs did not ask any questions throughout the entire session. The researcher also reminded children to switch driver/navigator roles every

20 minutes so that both children held each role for roughly the same duration. Children were allowed to use the cheat sheet provided to them in session 1. The activity took about 40 minutes, and the full duration was video/audio recorded through the video conference interface. As the existing automatic speech recognition (ASR) technologies are still not able to provide acceptable results for transcribing collaborative dialogues in noisy classroom environments (our dataset had an average word error rate (WER) of 0.65), the recordings were manually transcribed verbatim after session 2. Each transcript included timestamps, which indicated the beginning of each child speaking, and manually added punctuation such as question marks, which allowed our tool to extract the questions from the conversation. Next, these transcriptions were used to generate dialogue visualization for each child as described in Section III.

Session 3 was comprised of two parts: First, we conducted a think-aloud interview with each child individually to investigate the ways in which they chose to interact with the visualizations, including their perceptions, preferences, and expectations of the dialogue visualizations. The researcher shared their computer screen with the child, and gave the student access to control the researcher's computer remotely, which allowed the student to freely navigate through the dialogue visualization application as a user. Each child navigated through their own dialogue visualizations and responded to the researcher's questions about what they were thinking while exploring the visualizations. The researcher also asked each child questions such as whether they would want to use these charts in future, which chart would be most helpful for them, and their suggestions for design improvements. Each think-aloud interview lasted about 15-20 minutes. After both children in a pair participated in the think-aloud interview individually, they met to work on another collaborative coding activity with the same partner as before. After the coding activity, the students completed the peer evaluation survey individually and responded to a set of survey questions about the visualizations, such as whether they found them useful and whether they changed their behavior in the final collaborative activity.

D. Analysis Methods

This study focused on the following research question: "*In what ways does children's collaborative dialogue change after interacting with the dialogue visualizations?*" To answer the research question, we investigated two hypotheses and quantitatively compared the change in children's dialogue before and after seeing the visualizations. We also examined children's responses to the post-surveys to understand their perceived change after seeing the visualizations.

Hypothesis 1: Less-engaged children will be more engaged (*e.g.*, total number of words, total number of questions, and total number of open questions) in the next collaborative problem-solving activity.

To test this hypothesis, we first identified the less talkative child in each pair based on their total number of words in the first collaborative CS+Science activity. After identifying the less engaged child in each pair, we normalized their total number of words, total number of questions and total number of open questions based on the duration of each activity and put the score on a scale of 100 for further statistical analysis. Due to the small sample size, we conducted a Wilcoxon signed-rank test to compare how their dialogue changed before and after seeing the dialogue visualizations.

Hypothesis 2: Pairs of children will have a more balanced dialogue in the next collaborative problem-solving activity.

To test this hypothesis, we examined the entire duration of the dialogue via time series visualizations rather than a cumulative approach. For each pair, we first computed the average difference in word counts between children in the first activity. Given that each pair is different and some children are naturally more talkative than others, we used this information as the control/default value for what to expect when these two children work together again in the future activity. Thus, we used the average difference in word counts between children in the first activity as a threshold value for detecting imbalanced periods of time during the dialogue. For example, Figure 3 shows the interaction between two children in an imbalanced dialogue over time. As Figure 3 - Left illustrates, it is fairly easy to detect that one student was dominating the dialogue from almost beginning to the end of the activity. However, to make a more accurate interpretation and quantify the difference between children, we created another visualization (Figure 3 - Right) that depicts the difference of spoken words over time between the same children for the same activity, which makes it easier to examine how much each student dominated the conversation and how much the number of spoken words between the same children changes over time.

In a balanced dialogue, the difference in participation between partners would be close to "0" with slight spikes in negative and positive directions. However, Figure 3 - Right shows that the difference between this dyad was almost always in the positive direction, mostly between +50 and +150, indicating that one student was dominating the conversation during those times. To be able detect whether the children changed their collaborative dialogue, there needed to be a threshold for quantifying the domination behavior. However, it could be misleading to set a predetermined threshold (*e.g.*, $|100|$) to detect the areas of dialogue that a student was dominating because some individuals might be more talkative in nature or have more knowledge to share for problem solving; therefore, the threshold should be customized for each pair. As a solution, for each pair of children we computed the average of the difference their word counts and used that average as the threshold for comparison in their subsequent interaction.

For example, the average difference of spoken words between children for this presented pair was 78 words. The yellow box on Figure 3 - Right shows the parts that were below of above that threshold by taking the upper (+78) and lower limit (-78) of the word counts. By applying the same threshold (*i.e.*, $|78|$) to the second interaction (Figure 4), we now can assess how much of the time children were above or below the threshold to detect the domination behaviors. When the amount of time a child was above the threshold

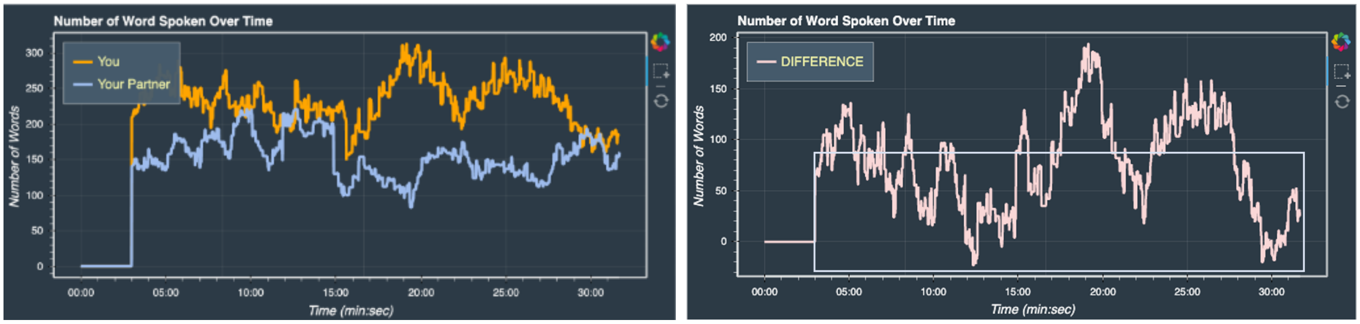

Fig. 3: **Before interacting with the dialogue visualization tool.** *Left:* This visualization illustrates the word count between two children with imbalanced dialogue in the first activity (before seeing the dialogue visualizations). *Right:* This visualization shows the difference in spoken words over time between the same children for the same activity.

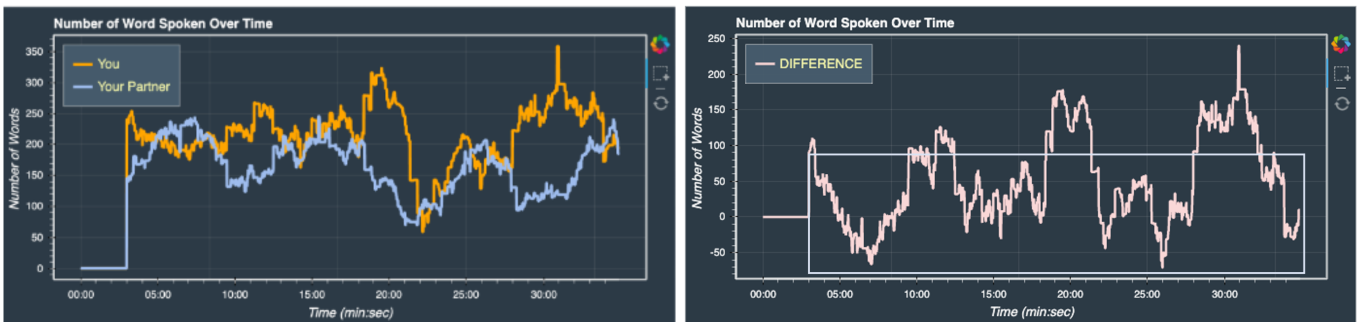

Fig. 4: **After interacting with the dialogue visualization tool.** *Left:* This visualization illustrates the visualization between the same two children with a more balanced dialogue in the second activity (after seeing the dialogue visualizations). *Right:* This visualization shows the difference of spoken words over time between the same children for the same activity.

is computed, we can quantitatively compare the children's domination behaviors over time. As can be seen on Figure 4 - Right, most instances of difference in this dialogue were between +50 and -50 words, indicating a more balanced dialogue.

Based on this analysis, we measured the amount of time that each pair was above the threshold for before and after pair interactions. We utilized a Wilcoxon signed-rank test to compare the amount of imbalanced time for each dyad before and after seeing visualizations. We hypothesized that children's dialogues would be significantly more balanced in the second activity after seeing the dialogue visualizations.

V. Results

The analysis of the children's dialogue indicated significant differences between their dialogues before and after seeing the visualizations. In this section, we present the results for each hypothesis.

A. Hypothesis 1: Less-engaged children will be more engaged (e.g., total number of words, total question numbers, and the total number of open questions) in the next collaborative problem-solving activity.

To examine this hypothesis, we compared the total number of words, total number of questions, and the total number of open questions for the less engaged children separately (Figure 6). The results showed that the less-engaged children's total number of words significantly increased after they used the dialogue visualization application (z = -2.803, p < 0.01) with a large effect size (r = -0.63). The median word count score increased from the first CS+Science activity (Md = 36.8) to the second CS+Science activity (Md = 53.26). Next, we examined whether less engaged children's total number of questions changed after interacting with the dialogue visualizations. The results showed that their total number of questions significantly increased after seeing the dialogue visualizations (z = -2.09, p = 0.037) with a medium effect size (r = -0.47). The median question count score increased from the first CS+Science activity (Md = 1.5) to the second CS+Science activity (Md = 1.9). Finally, we examined whether children changed their number of open questions asked after vieweing the visualizations. The results showed that the less-engaged children's total number of open questions did not change after they used the dialogue visualization application (z = -0.663, p = 0.51). The median open question count score for the first CS+Science activity was 0.34 and the median open question count score for the second CS+Science activity was 0.33.

The analysis so far did not consider how the proportion of less engaged student's dialogue out of the total pair's dialogue changed compared to their partners. Thus, we additionally examined whether the less-engaged children increased their word counts, question counts, and open question counts compared to their partners (Figure 6) before and after seeing the dialogue visualizations. For each pair, we divided the total word counts, question counts, and open question counts of the

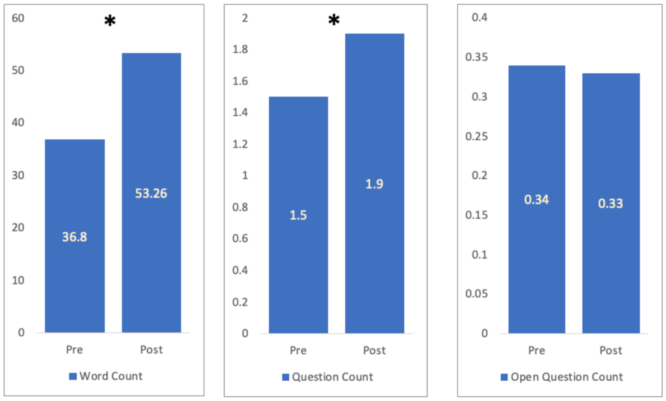

Fig. 5: Less-engaged children's total word counts, total question counts and total open question counts before and after seeing the dialogue visualizations.

less-engaged child (C1) by the total word counts, total question counts, and total open question counts of the pair (C1+C2). After computing the scores for both collaborative activities, we compared these scores with a Wilcoxon signed-rank test.

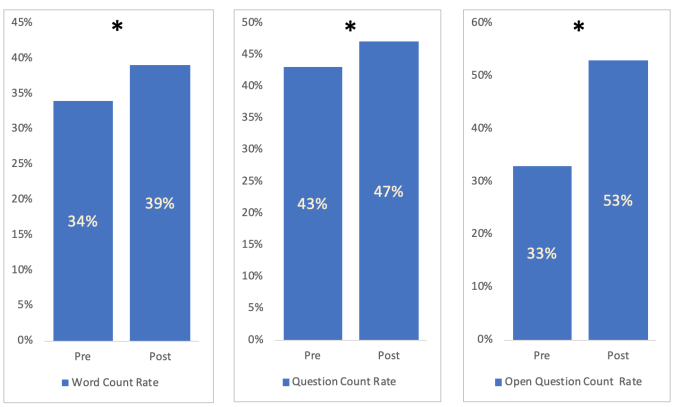

Fig. 6: Proportion of less engaged student's dialogue out of the total pair's dialogue before and after seeing the dialogue visualizations.

The results showed that the less-engaged children's total number of words significantly increased compared to their partners after they used the dialogue visualization application (z = -2.19, p = 0.028) with a medium effect size (r = -0.49). The median score increased from the first CS+Science activity (Md = 0.34) to the second CS+Science activity (Md = 0.39). We applied the same approach to the total question counts and the results showed that the less-engaged children's total question count significantly increased after they used the dialogue visualization application compared to their partners (z = -2.19, p = 0.028) with a medium effect size (r = -0.49). The median score increased from the first CS+Science activity (Md = .43) to the second CS+Science activity (Md = .47). Finally, we applied the same approach to the total open question counts and the results showed that the less-engaged children's total open question count significantly increased after they used the dialogue visualization application compared to their partners (z = -2.701, p = 0.007) with a large effect size (r = -0.6). The median score increased from the first CS+Science activity (Md = 0.33) to the second CS+Science activity (Md = 0.53).

B. *Hypothesis 2: Pairs of children will have a more balanced dialogue in the next collaborative problem-solving activity.*

To test this hypothesis, we examined how the amount of imbalanced dialogue time between both children changed before and after seeing the dialogue visualizations. We conducted a Wilcoxon signed-rank test and the results showed that there was no statistical difference between the imbalanced time in the activities before and after seeing the dialogue visualizations (z = -1.376, p = 0.17) with a medium effect size (r = -0.31). However, aggregating the results for all the pairs can be tricky: when a pair is already balanced in the first activity, there will not be much room for improvement and the defined threshold (based on the first activity) will already be very low. Consequently, even a small imbalance in the second activity will potentially be above the threshold, which is defined as the average word difference for each pair.

Consider the following example on Figure 7 - Left for a highly balanced pair with an average difference of 3 words, which is almost close to perfect balance. When the average difference (3 words for this specific pair) is applied to the pre and post activities as the threshold, the pair's word difference was more than 3 words for 94.73% of the time in the pre activity and for 96.78% of the time in the post activity (Figure 7 - Right). Thus, when the average word difference is already very low, even a very small difference that is more than the threshold (*e.g.*, 5 words) is calculated as an imbalance between partners. As a solution, we applied the k-means clustering method [55] to divide the pairs into three clusters based on each pair's threshold: low balanced, medium balanced and high balanced pairs. As shown on Figure 8, this method divided the pairs into three subgroups. Pairs 1, 2, 3 and 4 are in the high balanced group; Pairs 5, 6, 7 and 8 are in the medium balanced group; and Pairs 9 and 10 are in the low balanced group.

Given that the pairs who are already in the high balanced group will not need to make much changes in their dialogues, our analysis focused on the medium and low balanced groups. We conducted another Wilcoxon signed-rank test with the medium and low balanced pairs (6 pairs in total), which showed that dialogue between these pairs became significantly more balanced after they used the dialogue visualization application (z = -2.201, p = 0.028) with a large effect size (r = -0.64). The median imbalance score decreased from the first CS+Science activity (Md = 48.04) to the second CS+Science activity (Md = 38.618).

C. *Children's perceived change in their dialogue after seeing the dialogue visualizations.*

The responses to the survey questions asked after children completed all the activities indicated that children felt that they had changed their dialogues in the following collaborative coding activity after seeing the visualizations of their interaction with their partner in the previous activity. Out of 20 children, 19 responded to multiple choice and open ended questions on

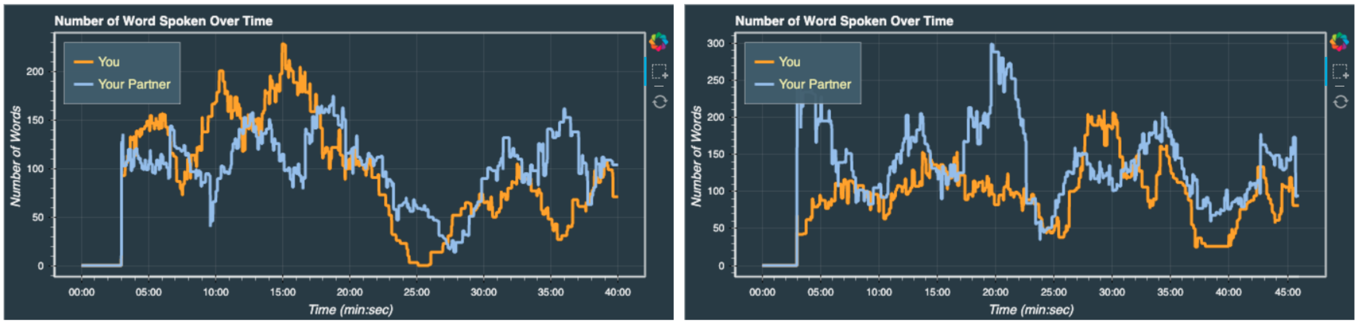

Fig. 7: Comparison of two activities of a pair with an already high balance in the first activity.

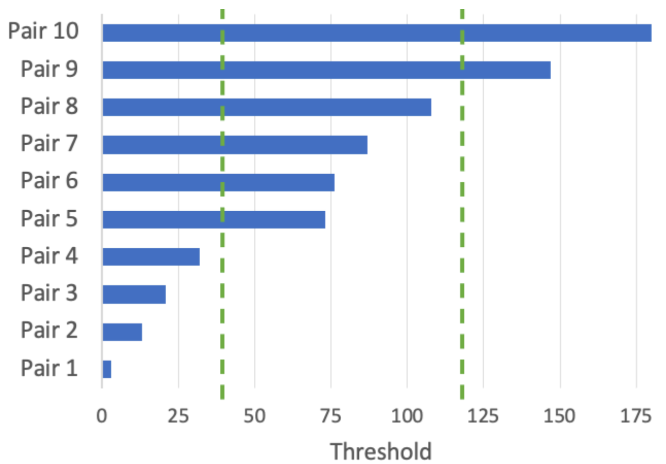

Fig. 8: Data sorted by partners' balance levels (high, medium, low). As the difference in children's word counts decreases, the balance level increases.

the the post-survey. 100% of the children said that seeing the visualizations changed their behavior and they would want to use the visualization application again in the future. We discuss these results with students' detailed responses to the survey questions in the following section.

VI. Discussion

This study investigated how children's collaborative dialogue changed after seeing the dialogue visualizations. The quantitative results indicated that less engaged students significantly increased their total number of words and total number of questions in the next collaborative problem-solving activity after seeing the dialogue visualizations. Moreover, when we compared the difference between less engaged and more engaged students, the results indicated that less engaged students increased the proportion of their total number of words, total number of questions and total number of open question significantly. Next, we examined whether the pairs had more balanced dialogues throughout the activity. The results showed that the pairs who had low or medium balanced dialogues adjusted their dialogues and had more balanced dialogues in the next activity. Finally, the survey results showed that children felt that they changed their dialogue and adjusted it in the final activity based on the information provided by the dialogue visualizations.

Learners bring different levels of motivation and knowledge into collaborative problem solving activities, and it is not uncommon to see a power imbalance between learners during the collaborative activities [12]. When students with different motivation or knowledge levels work on a task, one student may end up taking full control of the task, marginalizing the other student. Likewise, there are cases that a student violates the assumption of individual accountability and relies solely on the other student to solve the problem [56], [57]. The results in this study suggested that children adjusted their dialogue in the second activity. One student who spoke less in the first activity said:

> *[The visualizations] got me to notice that I need to communicate with her more and ask different questions. We did better this time because of it..*

Another child who had spoken less in the first activity said:

> *I think I asked better questions and talked more. I just said what I was thinking more often.*

On the other hand, children who spoke more in the first activity said that they tried to foster more balanced dialogues by allowing their partner speak more after seeing the visualizations. One such child said:

> *The graphs that I was shown changed my interaction with my partner because the graphs had shown that I talked more. So, I tried to let her speak more.*

Previous research has emphasized the importance of providing scaffolding during collaborative learning activities for better outcomes [17], and the results presented here provided further evidence that dialogue visualizations as "mirroring tools" can provide valuable information for children to reflect on their behaviors. Moreover, even after one intervention, children paid close attention to the dialogue visualizations and actively contributed to the conversation: The children who spoke less in the first activity reported that they tried to speak more and make the dialogue more balanced in the second activity.

The children in this study not only had a more balanced dialogue, they also asked more questions. Unlike existing dialogue visualization applications, we utilized the power of computational methods to automatically classify question types and present additional information about the content of the dialogue, in particular the question types. Questions play an important role in productive dialogues [50] and children in this study adjusted their question asking behaviors based on their previous dialogue, and tried to ask more and deeper questions after seeing the dialogue visualizations. For example, one student said:

> *It made me want to ask more questions and talk more to my partner.*

Another said:

> *I think it made me talk more, because in the line graph, I didn't really talk, so this time, I tried talking more and asking more open questions.*

Similar to adult learners, these children found that dialogue visualizations helped them reflect on their dialogues and work to improve them by balancing their talk time and also asking more questions during the collaborative learning activity.

VII. LIMITATIONS

Due to the COVID-19 pandemic, this study was conducted with a small group of students in a remote setting. The findings presented here may not generalize to other populations of learners, and it is important to conduct similar research studies with a larger and more diverse group of children. Another important limitation to note is that this study did not have a control group, and consequently we cannot draw any causal link between the application and the reported changes in children's dialogues. Despite the children's reports of perceived changes in their dialogues, comparing the results with a control group would provide evidence on the effectiveness of the application on students' behaviors while controlling for such factors as repeated interaction together. Finally, this tool is not fully automated and this study relied on manually transcribed dialogues. With advancement in the accuracy of automatic speech recognition tools, it will be possible to fully automate the application from transcribing the dialogues to generating the visualizations.

VIII. CONCLUSION AND FUTURE WORK

Collaborative dialogue provides useful information about the quality of a learning interaction between children. In this study, our overarching goal was to create and evaluate an application that analyzes middle school students' dialogues and presents visualizations to raise their awareness about their collaborative dialogue. We aimed to provide children with information about their dialogue that can help them to observe, regulate, and adjust their collaborative behaviors. The results showed that less engaged students spoke more and asked more questions after seeing the visualizations. They also increased their speaking time, question count and open question counts significantly more compared to their partners. Time series analysis also revealed that students had a more balanced dialogue throughout the activity. These results suggest that children benefit from dialogue visualizations and that they can make a significant difference in subsequent collaborative dialogue.

This study is one of the first of its kind to create dialogue visualizations for young learners and investigate how those learners change their future collaborative dialogue; thus, it revealed many promising research areas for the future. This tool has the potential to help the students balance their dialogue and become better partners during collabortive learning activities. In this study, we attempted to go beyond existing dialogue visualization approaches and present more information than simple word counts. Dialogue between children during collaborative problem solving activities provides a rich source of data, and future research should extend this work by exploring what new informative features can be extracted from the dialogue.

One promising direction is the utilization of time-series visualizations, which have been shown to be effective for depicting the dialogue over time in this study. There are many potential ways to improve them, such as putting useful information on top of the line charts. For example, when there is a spike in the line chart, the application can highlight additional information for that specific part of the chart. There are also some open questions about the time series visualizations that future work can further investigate. In this research, we investigated the dialogue visualizations as a whole, rather than each individual visualization (e.g., the question count pie chart); thus, it might still be beneficial to examine each type of visualization and their impacts on the outcomes.

Another open question is regarding when balanced dialogue is best and under what circumstances an unbalanced dialogue may work equally well or better. Children with different characteristics behave differently during collaborative problem solving, and there is a potential risk of undervaluing a child who is engaged but speaks less. Thus, there might be reasons why balanced word count would not necessarily mean balanced contribution to the activity or result in utmost engagement. Utilization of additional multi-modal indicators (*e.g.*, gaze behaviors and posture) can provide additional insight on how children engage during the paired activities.

Finally, this study focused on a specific collaboration approach, pair programming, and it would be interesting to investigate the impact of dialogue visualizations with a group of more than two children given that team dynamics would be quite different. When two children are in a dialogue, they are both aware that they are the only interlocutor (or collocutor) and thus, they feel obligated to participate in the dialogue. However, this situation is different when there are 3-4 people in a group, as some children can be completely silent and not feel the need to participate.

IX. ACKNOWLEDGEMENTS

This research was supported by the National Science Foundation through grant DRL-1640141. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the authors, and do not necessarily represent the official views, opinions, or policy of the National Science Foundation.

REFERENCES

[1] S. Campe, J. Denner, E. Green, and D. Torres, "Pair programming in middle school: variations in interactions and behaviors," *Computer Science Education*, vol. 30, no. 1, pp. 22–46, 2020.

[2] J. Denner, L. Werner, S. Campe, and E. Ortiz, "Pair programming: Under what conditions is it advantageous for middle school students?" *Journal of Research on Technology in Education*, vol. 46, no. 3, pp. 277–296, 2014.

[3] N. Mercer, S. Hennessy, and P. Warwick, "Dialogue, thinking together and digital technology in the classroom: Some educational implications of a continuing line of inquiry," *International Journal of Educational Research*, vol. 97, pp. 187–199, 2019.

[4] L. Major, P. Warwick, I. Rasmussen, S. Ludvigsen, and V. Cook, "Classroom dialogue and digital technologies: A scoping review," *Education and Information Technologies*, vol. 23, no. 5, pp. 1995–2028, 2018.

[5] J. C. Turner, A. Christensen, H. Z. Kackar-Cam, M. Trucano, and S. M. Fulmer, "Enhancing Students' Engagement: Report of a 3-Year Intervention With Middle School Teachers," *American Educational Research Journal*, vol. 51, no. 6, pp. 1195–1226, 2014.

[6] E. Deitrick, R. B. Shapiro, and B. Gravel, "How Do We Assess Equity in Programming Pairs?" in *Proceedings of the 12th International Conference on Computer Supported Collaborative Learning (CSCL)*. International Society of the Learning Sciences, (ISLS), 2016, pp. 370–377.

[7] M. T. Chi and R. Wylie, "The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes," *Educational Psychologist*, vol. 49, no. 4, pp. 219–243, 2014.

[8] D. J. Ferreira, A. C. Goncalves, and L. de Oliveira Berreta, "The importance of productive dialog in computer science students' creative thinking," in *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*, 2014, pp. 1–8.

[9] R. Wegerif, B. M. McLaren, M. Chamrada, O. Scheuer, N. Mansour, J. Mikšátko, and M. Williams, "Exploring creative thinking in graphically mediated synchronous dialogues," *Computers & Education*, vol. 54, no. 3, pp. 613–621, 2010.

[10] M. M. M. M. Bakhtin and M. Holquist, *The dialogic imagination : four essays*. University of Texas Press, 1981.

[11] R. Gennari, A. Melonio, and M. Rizvi, "Turn taking with turn-talk in group: Actions and reflections with children and teachers," *Multimedia Tools and Applications*, vol. 78, no. 10, pp. 13 461–13 487, 5 2019.

[12] C. M. Lewis and N. Shah, "How Equity and Inequity Can Emerge in Pair Programming," in *Proceedings of the 11th Annual International Conference on International Computing Education Research (ICER)*. ACM, 2015, pp. 41–50.

[13] J. L. Schultz, J. R. Wilson, and K. C. Hess, "Team-based classroom pedagogy reframed: The student perspective," *American Journal of Business Education*, vol. 3, no. 7, pp. 17–24, 2010.

[14] J. M. DiMicco, A. Pandolfo, and W. Bender, "Influencing group participation with a shared display," in *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW)*. New York, New York, USA: ACM Press, 2004, pp. 614–623.

[15] K. Bachour, F. Kaplan, and P. Dillenbourg, "An interactive table for supporting participation balance in face-to-face collaborative learning," *IEEE Transactions on Learning Technologies*, vol. 3, no. 3, pp. 203–213, 2010.

[16] M. Schittekatte and A. Van Hiel, "Effects of partially shared information and awareness of unshared information on information sampling," *Small Group Research*, vol. 27, no. 3, pp. 431–449, 8 1996.

[17] J. P. Hourcade, "Interaction Design and Children," *Foundations and Trends in Human-Computer Interaction*, vol. 1, no. 4, pp. 277–392, 4 2008.

[18] A. Soller, A. Martinez, P. Jermann, and M. Muehlenbrock, "From Mirroring to Guiding: A Review of State of the Art Technology for Supporting Collaborative Learning," *International Journal of Artificial Intelligence in Education*, vol. 15, no. 4, pp. 261–290, 2005.

[19] C. Richter, E. Simonenko, T. Sugibuchi, N. Spyratos, F. Babic, J. Wagner, J. Paralic, M. Racek, C. Damşa, and V. Christophides, *Mirroring Tools for Collaborative Analysis and Reflection*. Brill, 1 2012.

[20] W. Matcha, N. A. Uzir, D. Gasevic, and A. Pardo, "A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective," *IEEE Transactions on Learning Technologies*, vol. 13, no. 2, pp. 226–245, 4 2020.

[21] R. Bodily and K. Verbert, "Review of research on student-facing learning analytics dashboards and educational recommender systems," *IEEE Transactions on Learning Technologies*, vol. 10, no. 4, pp. 405–418, 10 2017.

[22] S. Charleer, J. Klerkx, E. Duval, T. D. Laet, and K. Verbert, "Towards balanced discussions in the classroom using ambient information visualisations," *International Journal of Technology Enhanced Learning*, vol. 9, no. 2/3, p. 227, 2017.

[23] M. Celepkolu, J. B. Wiggins, A. C. Galdo, and K. E. Boyer, "Designing a visualization tool for children to reflect on their collaborative dialogue," *International Journal of Child-Computer Interaction*, vol. 27, p. 100232, 3 2021.

[24] R. Wegerif, "Towards a dialogic theory of how children learn to think," *Thinking Skills and Creativity*, vol. 6, no. 3, pp. 179–190, 2011.

[25] K. Littleton and N. Mercer, *Interthinking: Putting Talk to Work*. Routledge, 2013.

[26] S. Knight and K. Littleton, *Dialogue and the development of children's thinking: a sociocultural approach*. Routledge, 2007.

[27] S. Michaels, C. O'Connor, and L. B. Resnick, "Deliberative Discourse Idealized and Realized: Accountable Talk in the Classroom and in Civic Life," *Studies in Philosophy and Education*, vol. 27, no. 4, pp. 283–297, 2008.

[28] B. D. Anderson, R. Swenson, and N. D. Gilkerson, "Understanding dialogue and engagement through communication experts' use of interactive writing to build relationships," *International Journal of Communication*, vol. 10, pp. 4095–4118, 2016.

[29] M. L. Kent and M. Taylor, "Toward a dialogic theory of public relations," *Public Relations Review*, vol. 28, no. 1, pp. 21–37, 2002.

[30] K. Kreijns, P. A. Kirschner, and W. Jochems, "Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research," *Computers in Human Behavior*, vol. 19, no. 3, pp. 335–353, 2003.

[31] A. Melonio and M. Rizvi, "The design of TurnTalk for the scaffolding of balanced conversations in groups of children," in *In Proc. of the 1st International Symposium on Emerging Technologies for Education*. Springer Verlag, 2017, pp. 278–287.

[32] J. Y. Kim, R. A. Calvo, K. Yacef, and N. J. Enfield, "A Review on Dyadic Conversation Visualizations - Purposes, Data, Lens of Analysis," *arXiv*, 2019.

[33] I. Amarasinghe, D. Hernandez-Leo, K. Michos, and M. Vujovic, "An Actionable Orchestration Dashboard to Enhance Collaboration in the Classroom," *IEEE Transactions on Learning Technologies*, vol. 13, no. 4, pp. 662–675, 10 2020.

[34] E. L. Starr, J. Reilly, and B. Schneider, "Toward Using Multi-Modal Learning Analytics to Support and Measure Collaboration in Co-Located Dyads," in *Proceedings of 13th International Conference of the Learning Sciences (ICLS)*, 2018, pp. 448–455.

[35] S. Charleer, A. V. Moere, J. Klerkx, K. Verbert, and T. De Laet, "Learning Analytics Dashboards to Support Adviser-Student Dialogue," *IEEE Transactions on Learning Technologies*, vol. 11, no. 3, pp. 389–399, 7 2018.

[36] M. Ez-zaouia, A. Tabard, and E. Lavoué, "EMODASH: A dashboard supporting retrospective awareness of emotions in online learning," *International Journal of Human Computer Studies*, vol. 139, p. 102411, 7 2020.

[37] P. Jermann and P. Dillenbourg, "Group mirrors to support interaction regulation in collaborative problem solving," *Computers & Education*, vol. 51, no. 1, pp. 279–296, 8 2008.

[38] J. Janssen and D. Bodemer, "Coordinated Computer-Supported Collaborative Learning: Awareness and Awareness Tools," *https://doi.org/10.1080/00461520.2012.749153*, vol. 48, no. 1, pp. 40–55, 1 2013. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/00461520.2012.749153

[39] T. Kim, P. Hinds, and A. Pentland, "Awareness as an antidote to distance: making distributed groups cooperative and consistent," in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW)*. ACM, 2012, pp. 1237–1246.

[40] F. Bishop, J. Zagermann, U. Pfeil, G. Sanderson, H. Reiterer, and U. Hinrichs, "Construct-A-Vis: Exploring the free-form visualization processes of children," *IEEE Transactions on Visualization and Computer Graphics*, vol. 26, no. 1, pp. 451–460, 2020.

[41] P. Davis, M. Horn, F. Block, B. Phillips, E. M. Evans, J. Diamond, and C. Shen, ""Whoa! We're going deep in the trees!": Patterns of collaboration around an interactive information visualization exhibit," *International Journal of Computer-Supported Collaborative Learning*, vol. 10, no. 1, pp. 53–76, 2015.

[42] M. Vásquez-Bermúdez, C. Sanz, M. A. Zangara, and J. Hidalgo, "Mirroring Tools and Interaction Indicators in Collaborative Work: A Systematic Review," in *Communications in Computer and Information Science*, vol. 1309. Springer Science and Business Media Deutschland GmbH, 11 2020, pp. 179–192.

[43] B. Chen and M. Resendes, "Uncovering what matters: analyzing transitional relations among contribution types in knowledge-building discourse," in *Proceedins of the Fourth International Conference on Learning Analytics And Knowledge*, 2014, pp. 226–230.
[44] J. Park, S. H. Han, H. K. Kim, S. Oh, and H. Moon, "Modeling user experience: A case study on a mobile device," *International Journal of Industrial Ergonomics*, vol. 43, no. 2, pp. 187–196, 3 2013.
[45] S. Bresciani, "Visual Design Thinking: A Collaborative Dimensions framework to profile visualisations," *Design Studies*, vol. 63, pp. 92–124, 7 2019.
[46] S. Knight and K. Littleton, "Dialogue as Data in Learning Analytics for Productive Educational Dialogue." *Journal of Learning Analytics*, vol. 2, no. 3, pp. 111–143, 2015.
[47] M. Kapur, "Temporality matters: Advancing a method for analyzing problem-solving processes in a computer-supported collaborative environment," *International Journal of Computer-Supported Collaborative Learning*, vol. 6, no. 1, pp. 39–56, 2011.
[48] M. Datar, A. Gionis, P. Indyk, and R. Motwani, "Maintaining Stream Statistics over Sliding Windows," *SIAM Journal on Computing*, vol. 31, no. 6, pp. 1794–1813, 2002.
[49] C. Chin and J. Osborne, "Students' questions: a potential resource for teaching and learning science," *Studies in Science Education*, vol. 44, no. 1, pp. 1–39, 2008.
[50] S. Hennessy, S. Rojas-Drummond, R. Higham, A. M. Márquez, F. Maine, R. M. Ríos, R. García-Carrión, O. Torreblanca, and M. J. Barrera, "Developing a coding scheme for analysing classroom dialogue across educational contexts," *Learning, Culture and Social Interaction*, vol. 9, pp. 16–44, 2016.
[51] M. Aziza, "An analysis of a teacher's questioning related to students' responses and mathematical creativity in an elementary school in the UK," *International Electronic Journal of Elementary Education*, vol. 10, no. 4, pp. 475–487, 2018.
[52] S. Rose, D. Engel, N. Cramer, and W. Cowley, "Automatic Keyword Extraction from Individual Documents," in *Text Mining: Applications and Theory*. John Wiley and Sons, 2010, pp. 3–20.
[53] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," *arXiv*, 1 2013.
[54] M. Celepkolu, D. A. Fussell, A. C. Galdo, K. E. Boyer, E. N. Wiebe, B. W. Mott, and J. C. Lester, "Exploring middle school students' reflections on the infusion of cs into science classrooms," in *Proceedings of the 51st ACM Technical Symposium on Computer Science Education - SIGCSE '20*. Association for Computing Machinery (ACM), 2020, pp. 671–677.
[55] S. P. Lloyd, "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 1982.
[56] J. W. Strijbos, "Assessment of (Computer-Supported) Collaborative Learning," *IEEE Transactions on Learning Technologies*, vol. 4, no. 1, pp. 59–73, 2011.
[57] N. L. Kerr and S. E. Bruun, "Dispensability of member effort and group motivation losses: Free-rider effects," *Journal of Personality and Social Psychology*, vol. 44, no. 1, pp. 78–94, 1983.

Mehmet Celepkolu is a Research Scientist at the Computer and Information Science and Engineering Department at the University of Florida. He received his Ph.D. in Computer Science and his M.S. in Computer Engineering from University of Florida. His research investigates artificial intelligence and human-computer interaction with an emphasis on natural language processing applications in education.

Aisha Chung Galdo is a combined degree student in the Warrington College of Business at the University of Florida receiving her B.S in Accounting and M.S in Information Systems. Her research focuses on middle school computer science education.

Kristy Elizabeth Boyer is an Associate Professor of Computer and Information Science and Engineering at the University of Florida. Her research focuses on dialogue for teaching and learning, including tutorial dialogue and computer-supported collaborative learning. Her work in computer science education spans K-12 and post-secondary contexts through game-based learning and intelligent learning.