International Journal of Child-Computer Interaction 27 (2021) 100232

Contents lists available at ScienceDirect

International Journal of Child-Computer Interaction

journal homepage: www.elsevier.com/locate/ijcci

Designing a visualization tool for children to reflect on their collaborative dialogue

Mehmet Celepkolu *, Joseph B. Wiggins, Aisha Chung Galdo, Kristy Elizabeth Boyer

University of Florida, Gainesville, FL, USA

ARTICLE INFO

Article history:
Received 2 September 2020
Received in revised form 4 November 2020
Accepted 5 November 2020
Available online 29 November 2020

Keywords:
Collaborative dialogue
Data visualization
Reflection

ABSTRACT

Collaborative learning is an essential part of children's development, positively impacting academic achievement and fostering higher levels of reasoning. However, young learners often face challenges with taking turns in conversation, openly listening to ideas, and respecting different viewpoints. One way to foster collaborative skills may be to raise children's awareness of their own collaborative dialogue. In this paper, we present a new interactive visualization application that supports children in reflecting on their collaborative dialogue from a recent prior interaction. The tool analyzes children's completed dialogue and then presents temporal information about their interaction with their partner. We implemented two studies with 36 seventh grade children who collaboratively completed computing activities. We conducted think-aloud sessions to investigate children's perceptions, preferences, and expectations of the collaborative dialogue visualizations. The results showed that the dialogue visualizations hold promise for helping children increase their awareness of collaborative dialogue and set their own goals regarding ways they would like to improve.

1. Introduction

Collaboration is a critical skill for children to master, and today's generation of learners are likely to be collaborating with individuals across the globe via a constantly evolving digital landscape. Collaboration will be a tremendous asset in their life-long learning, with numerous studies indicating the benefits of collaborative learning. Not only does it positively impact learners' academic and social abilities (Turner, Christensen, Kackar-Cam, Trucano, & Fulmer, 2014), but it also helps them develop higher levels of reasoning and critical thinking, and exposes them to diverse viewpoints (Major, Warwick, Rasmussen, Ludvigsen, & Cook, 2018; Mercer, Hennessy, & Warwick, 2019). The interaction between individuals presents numerous opportunities for students to develop, defend, and evaluate their perspectives through constructive argumentation (Mercer et al., 2019).

On the other hand, not all group interactions around learning are successful (Kreijns, Kirschner, & Jochems, 2003). Previous literature reported pitfalls such as conflicts between partners (Celepkolu & Boyer, 2018b; Weinberger, Stegmann, & Fischer, 2010) or inequity arising within the discussion during collaborative problem-solving activities (Engle, Langer-Osuna, & McKinney de Royston, 2014; Lewis & Shah, 2015). These challenges may be even more prominent amongst young learners who lack collaboration skills such as taking turns in conversation (Deitrick, Shapiro, & Gravel, 2016) or openly listening to and respecting different ideas (Tsan, Lynch, & Elizabeth Boyer, 2018). Moreover, in a traditional classroom setting, students may not explain their thought process or ask high-level questions unless they are explicitly required (Chinn, O'donnell, & Jinks, 2000). These challenges, if unaddressed, can lead to problems such as an imbalance in conversation (Lewis & Shah, 2015), and negative dispositions toward collaboration in the future (Schultz, Wilson, & Hess, 2010).

Various studies and theories have provided insights into how to design intelligent systems to overcome these challenges and positively influence collaborative learning. Soller, Martinez, Jermann, and Muehlenbrock (2005) suggest that managing collaborative interaction means informing and guiding students about their metacognitive activities through analyzing their interaction and providing dynamic feedback. Providing knowledge-related group awareness information consisting of data on how the collaboration is going can help learners observe, regulate, and adjust their collaborative behaviors based on the needs of the team (Bodemer & Dehler, 2011; Fransen, Kirschner, & Erkens, 2011; Soller et al., 2005). Several studies have suggested that providing students with visual analytics that mine their collaborative dialogue and display metadata can increase students' awareness of their participation in collaborative problem solving (Charleer, Klerkx, Duval, Laet, & Verbert, 2017; DiMicco, Pandolfo, & Bender, 2004). These visualizations can lead to more cooperative

* Corresponding author.
E-mail address: mckolu@ufl.edu (M. Celepkolu).

https://doi.org/10.1016/j.ijcci.2020.100232
2212-8689/

groups (Kim, Hinds, & Pentland, 2012), and also foster more balanced participation through increasing students' conversational awareness of matters such as letting others speak (Samrose, Zhao, White, Li, Nova, Lu, et al., 2018).

While visualization systems have shown promise in supporting collaborative learning with respect to encouraging more equitable participation, previous research has often overlooked two important points: First, prior work has designed visualizations of children's interactions without involving those children in the development process. Relatedly, visualization tools are often designed to serve educators or adult users rather than children. However, it is crucial to identify children's understandings and expectations in light of their social-emotional development by investigating how they interpret visualizations (Bodemer & Dehler, 2011), and exploring whether the children perceive the visualizations as useful (Janssen, Erkens, & Kirschner, 2011; Nova, Wehrle, Goslin, Bourquin, & Dillenbourg, 2007). The second shortcoming of most existing applications is that they display only summative visualizations of the total number of contributions within a dialogue, disregarding temporal dynamics that are key facets of the unfolding dialogue process over time (Knight & Littleton, 2015; Wiltshire, Butner, & Fiore, 2018). While cumulative visualization approaches can be useful in uncovering some patterns related to the productivity of collaborative learning, they inevitably lose important sequential information of the interactions (Chen, Yang, Zhao, Cai, & He, 2018; Kapur, 2011).

These two shortcomings leave an important research gap that we aim to fill. Our goal is to design and develop a visualization tool that children can use to reflect on their own dialogues and set goals for themselves. We have co-created an interactive visualization tool tailored to middle school children's developmental level and expectations. This process has involved investigating how children interpret dialogue visualizations, how they reflect on the implications of the visualizations, and how they form goals for their future collaborations.

Based on these research goals, this paper focuses on the following open research questions:

1. *How do middle school children perceive visualizations illustrating their collaborative dialogue?*
2. *How might middle school children employ dialogue visualizations to reflect on their collaborative dialogue?*
3. *What design implications emerge from children's feedback on the dialogue visualizations?*

To investigate these research questions, we iteratively prototyped a dialogue visualization tool (Fig. 1) that automatically mines children's dialogues from their previously-recorded paired collaborative learning activities within middle school classrooms.

The initial version of this tool was informed by theory, as we detail in the next section. This initial version included interactive visualizations showing total word count, total question count, and number of words spoken over time alongside the dialogue transcript. We conducted a set of individual think-aloud sessions with 18 children, which led us to revise some of the visualizations and include additional features in the second version, such as video recordings of the activity screen and children's group interactions. We conducted a second set of individual think-aloud sessions with a different group of 18 children who used the revised system. In that study, we investigated how children used the application to reflect on their collaboration, and collected feedback on further design expectations. The results showed that most children had positive feelings toward the visualizations and perceived them as useful. They self-reported benefits in subsequent collaborations after having viewed the visualizations. The children also offered many suggestions for improving the dialogue visualization, and these suggestions provide design guidelines and insight into the goals children may set when empowered to reflect on their own dialogues through visualizations.

The novel contributions of this work are threefold. First, previous research often focused on visualization studies in higher education or informal learning settings; in contrast, this study reflects middle grade children's understandings and expectations. Second, we present a novel collaborative dialogue visualization tool that preserves the temporal dynamics in dialogue flow in the context of pair programming. Finally, we discuss implications for the design of collaborative dialogue visualization tools to support learners in this age group.

2. Background

This work builds upon a body of prior research on supporting children during collaborative problem solving. In this section, we present the theoretical framing of this study and then discuss the productive dialogue patterns along with empirical results from previous research, which motivated the studies presented in this paper. Next, we discuss the motivation behind dialogue visualizations, such as supporting student's reflection and increasing their awareness of their own behavior. Finally, we critically discuss how collaborative dialogue has been visualized in previous research by providing insight into how these visualizations can be improved based on children's understanding and expectations.

2.1. Collaboration among children

Dialogue is one of the primary communication channels during many collaborative learning interactions (Wegerif, 2011). During dialogue, students discuss and challenge ideas, explore different points of view, discover new information, and co-construct knowledge (Major et al., 2018). However, in collaborative learning, simply placing students in groups does not ensure productive dialogues (Kreijns et al., 2003). Some interactions can result in a shallow discussion rather than a thought-provoking process through which students give and counter arguments. Thus, facilitating strong collaborative dialogue is essential for generating new ideas, enhancing cognitive understanding (Kuhn & Crowell, 2011), provoking reasoning (Mercer, 2008), and improving critical thinking (Kuhn, 2018).

A wide body of literature has investigated how the type of dialogue during interaction influences the effectiveness of problem-solving processes. Bakhtin's dialogic theory (Bakhtin & Holquist, 1981) emphasizes the importance of productive talk and defines it as an active double-voiced discourse, where sides do not "dominate the other's thought". According to this theory, creativity emerges from the tension between different ideas. Similarly, Wegerif et al. (2010) define this dialogic interaction process "as a dance of voices and perspectives", and suggest that confrontation of ideas stimulates new idea generation. The "Thinking Together" model Mercer (2013), Mercer et al. (2019) suggests that learning to reason with others helps students be independent thinkers. Similarly, the "Exploratory Talk" model Knight and Littleton (2007), Littleton and Mercer (2013) posits that in a productive dialogue, every member of the group contributes to the conversation with relevant information; every member is critical but constructive of each other's ideas; members make sure they reach consensus at each task before proceeding to the next task; they treat each idea as worthy of consideration; and they are open to questions.

In a traditional classroom setting, students may not explain their thought process or ask high-level questions (Nystrand, Wu, Gamoran, Zeiser, & Long, 2003) unless explicitly required to

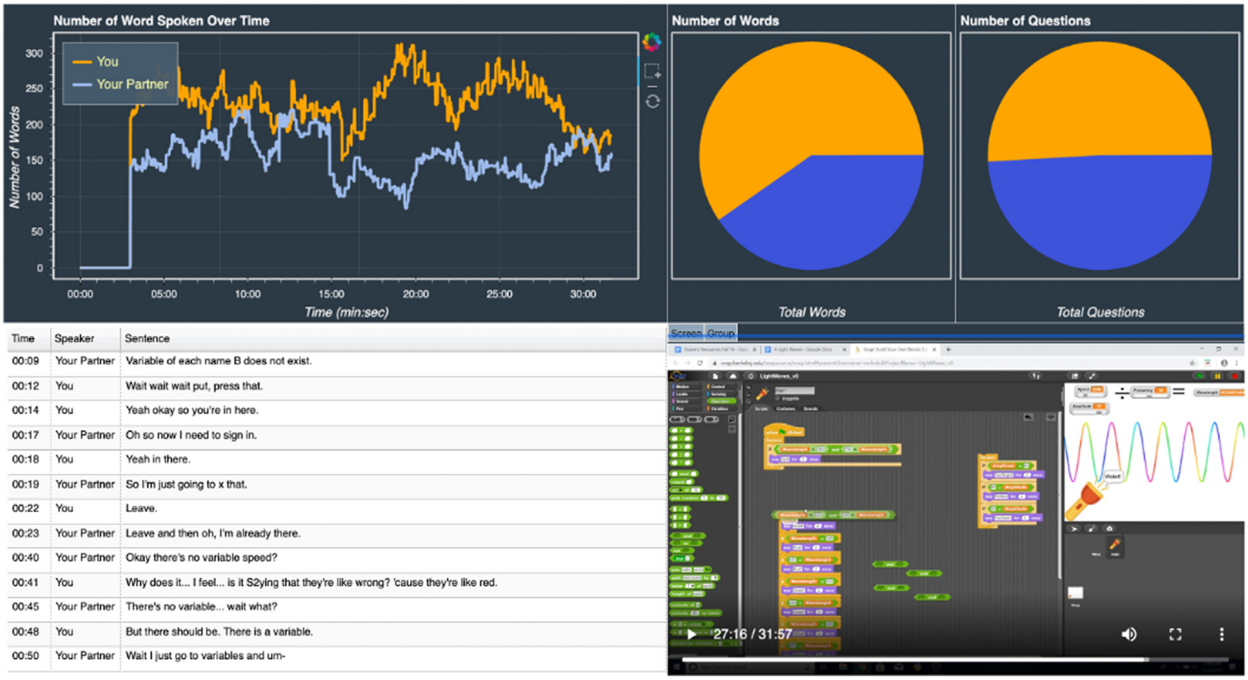

Fig. 1. Final version of the visualization tool.

elaborate on their reasoning for their conclusion (Chinn et al., 2000). Galton and Williamson (1992) suggested that one reason could be that students often do not have a clear idea of the purposes of the group activities and the expectations from them. These aforementioned challenges, along with the highly demanding nature of dialogic inquiry, led us to explore new ways of supporting students in collaboration. With the presented visualization tool we aim to increase students' awareness of their dialogue and help them regulate their interactions for better collaborative work.

2.2. Supporting children's reflection and participation

The focus of our work is to develop visualizations as a tool for children to reflect on their dialogue. To the best of the authors' knowledge, based on an extensive literature search, there are no published studies of visualization tools for children's dialogue. The only prior visualization studies we have found regarding children reflecting on their own data involve their health data (Wang, Zhu, Nacenta, & Dai, 2017), which demonstrated that, for children ages 11–18, reflecting on visualizations of health behavior can positively influence their health behaviors. In that study, the children were asked to play a game on a mobile app by placing the food on their plate that they would want to eat. The children could share the food they picked with their classmates and could also see their classmates' food choices. Next, the application visualized whether their food choices were healthy, and children could compare their results with their friends. The paper suggests that this process is motivating for children to explore different food choices and may help foster positive behavior change.

Prior research has investigated the ways in which middle school children understand or generate visualizations, but in those studies the children were not visualizing data related to themselves, but rather to scientific processes (Davis et al., 2015) or for the purposes of learning how to construct visualizations in general (Bishop et al., 2020). Different than these studies, we suggest using visualizations as a tool for children to reflect on their dialogue and set goals regarding their collaborative behavior. We anticipate that children will benefit from such a tool in part because visualization of dialogue can provide a new way of augmenting human cognition by facilitating an easier, structural way to examine the information (Kim, Calvo, Yacef, & Enfield, 2019).

Our tool rests on the premise that children can benefit from reflecting on their own behavior. In a recent study, the researchers provided children with "Scratch memories", a video montage of their coding in the Scratch online community, and found that it promoted children's reflections on their personal trajectories, and increased children's motivation to participate (Dhariwal, 2018). In another study, Rosenbaum (2009) used a program in which children took notes and recorded updates via *Jots*, allowing children to write about their experiences, frustrations, and challenges with the Scratch block-based programming language. The goal of the application was to support children in reflective thinking and learning on four different facets: cognitive, emotional, social, and temporal. The results based on case studies with three middle school students showed that seeing their previous jots helped children see their previous mistakes and how they developed strategies to overcome challenges. This reflection process helped them become more aware of their previous as well as their current activities. Our study contributes to this body of research on supporting children to reflect on their behaviors by providing them interactive visualizations illustrating their own collaborative dialogue.

2.3. Dialogue visualizations for adults

Despite the lack of dialogue visualization applications for children, there have been various studies with adults. Previous work suggests that using visual analytics that mine dialogue and visualize metadata (such as participation) can increase adult students' awareness of their participation in collaborative problem solving (Charleer et al., 2017; DiMicco et al., 2004). One of the most common factors visualized by these applications is participation, a basic willingness to interact and share information (Care, Scoular, & Griffin, 2016). Participation is a minimum requirement for collaboration (Dowell, Nixon, & Graesser, 2018). While overparticipation can lead other students to "free ride" or to feel marginalized within the learning process, under-participation can negatively impact the other students' motivation and lead to imbalanced workload.

A large body of research on intelligent systems has investigated technologies that raise individuals' awareness of balanced teamwork. Measuring participation solely on word count or turn taking is a simplistic approach; however, a wide body of literature on automatic analysis of conversation focuses on these phenomena (DiMicco et al., 2004). For example, Charleer et al. (2017) utilized ambient information visualizations to provide feedback about over- and under-participation and investigated student perceptions of visual participation feedback in two graduate-level courses. During the activities, a teaching assistant manually captured when a group started and stopped talking, and a live visualization system displayed the changes on the board. Similarly, DiMicco et al. (2004) developed the Second Messenger tool, which interpreted the sound level over a certain threshold as an approximation of word count, and used it as the participation measure. The results from 100 adults (mean age = 25) showed that providing live feedback caused over-participators to limit their comments, but it did not help under-participators change their participation levels. In another study, Kim et al. (2012) developed the Meeting Mediator, which measured each individual's participation in the group activity based on the length and speed of talking, the number of turn taken, the average length of the turn, and the variation in volume level, and then provided real-time feedback to the participants. They conducted a study of 180 adults (mean age = 29.4) and found that providing feedback based on those indicators made the participants more cooperative in the teamwork and increased their performance. Aligned with these studies, our goal in this paper is to investigate whether dialogue visualizations can be useful for children and how they reflect on their dialogue to improve their collaboration skills, such as through more-balanced participation. It is also crucial to ensure that these visualizations are based on the users' design expectations, and that they perceive these visualizations as useful (Janssen et al., 2011; Nova et al., 2007).

In addition to participation visualizations, some studies have provided information about some other factors such as engagement, humor, emotion, and agreement/disagreement. For example, Samrose et al. (2018) presented CoCo, which automatically extracted information such as engagement, attention, speech overlap, and turn-taking from 39 college (age: 19–23) students' conversations and investigated whether providing feedback could change the participants' behaviors. They found that the feedback led to more balanced participation and significant improvements in students' self-evaluations of conversational awareness, such as letting partners speak. Yamada, Kaneko, and Goda (2016) developed a social presence visualization function, which automatically categorizes students' posts into one of the 17 predefined categories such as use of humor, expressing emotion, and expressing agreement/disagreement in an LMS system. In another study, Yamada and Goda (2018) investigated the effects of this system on 160 second-year college students' perceived social presence, cognitive learning, and contribution to the project. The results showed that the visualization directly helped with improvement in social relationships and indirectly helped with perceived cognitive learning and contribution to project work.

Our study extends this body of research by investigating additional features such as visualization of participation over time, presenting the dialogue content, displaying the number of questions, and including video recordings of the activity and pair interactions. Our visualization tool provides insight into the temporal dynamics of the dialogue in addition to the common summative approaches to visualize collaborative activity as described in previous literature (Fig. 2).

Fig. 2. Sample summative graphs for illustrating participation and other factors. Top Left: DiMicco et al. (2004); Top Right: Charleer et al. (2017); Bottom Left and Bottom Right: Samrose et al. (2018).

Moreover, we investigate how young learners understand these visualizations, identify their challenges, and explore their expectations from dialogue visualizations. Given that there is no previous study in the context of dialogue visualizations for middle school students, this study receives its motivation from the presented benefits of these applications for supporting adult learners' reflection processes for better collaboration. We believe these applications have the potential to support younger learners, and thus, we used an iterative design process to develop a dialogue visualization application from scratch and made updates based on children's feedback.

3. Methods

We followed an iterative design-based research approach (Baumgartner et al., 2003; McKenney & Reeves, 2014), in which we implemented two cycles of design, implementation, analysis, and evaluation phases. Design-based research provides "a systematic but flexible methodology based on collaboration among researchers and practitioners" and is grounded in both theory and practical applications (Wang & Hannafin, 2005). The first iteration of this study aimed to understand children's perceptions of two different types of visualizations: summative visualizations versus time-series visualizations. In the second iteration, we redesigned some of these visualizations, added new features based on children's feedback, and conducted another user study. The outcomes of this iterative design-based research process lead to implications for theory, design principles for better user experience and an application. The following subsections describe these iterative studies and then discuss the results.

3.1. ITERATION 1: Generating the dialogue visualizations

The visualizations in the first iteration centered around charts illustrating participation in the form of turn-taking behavior and word counts, along with the quality of the interaction through number of questions, an important consideration as evidenced by prior research described above. Children's interactions were video and audio recorded, which was necessary to create the visualizations and to investigate children's interactions with each other and with the application. Next, we manually transcribed the dialogues verbatim, including information such as filler words, false starts, and grammatical errors. Each transcript also included timestamps, which indicated the beginning of each child speaking. The transcriptions included manually added punctuation such as question marks, which later allowed our tool to extract the questions from the dialogue.

We developed the web-based interactive visualization tool with the Python Bokeh visualization library *Bokeh 1.1.0 documentation* (2020), which allows for generating interactive visualizations by uploading transcription files. We first created three summative visualizations in pie charts (Fig. 3): (1) total number of times a child spoke to their partner, (2) total number of words they spoke to their partner, and (3) total number of questions they asked during the activity. Pie charts are powerful for illustrating relative proportions of several groups of data in a simple way. Although studies clearly point out the limitations of pie charts for displaying patterns (Bertin, 1981), a recent experimental study (Siirtola, 2014) showed that 75% of users consider pie charts as the most or second-most pleasing to use, and 44% of users prefer pie charts over bar charts and doughnut charts. The pie charts in our visualization tool had interactive features, and children could see the total number of dialogue turns taken, word spoken, or questions asked when they hovered on the charts (Fig. 3). Also, instead of using their names on the charts, the charts would refer to them as "You" and "Your Partner".

We also generated a time-series line chart (Fig. 4) using a sliding-window approach (Datar, Gionis, Indyk, & Motwani, 2002), showing the number of words spoken over time. The purpose of this visualization was to preserve the temporal dynamics in dialogue flow, which has been shown as a key facet of data representation for dialogue (Knight & Littleton, 2015). While summative approaches such as pie charts can be useful in uncovering some patterns related to the productivity of collaborative learning, these approaches lose sequential information (Chen & Resendes, 2014; Kapur, 2011). We created the time-series based line charts based on the following algorithm: (1) count the total number of words spoken by a child within a predefined time window interval, (2) slide the window by one second (drop the earliest and add a new one) and recalculate the total number of words for the new time window, and (3) repeat the same process until the window slides over the entire dialogue. Fig. 4(a) illustrates a simple example of how the sliding window method works. In the example, the window size is four and it calculates the total number of words for each second. The number of words after each iteration is calculated as 5, 2, 7, and 17, respectively. After identifying the values for each time interval, the chart can be created with the X axis showing the time and the Y axis showing the value (total count of the words in an interval), as shown in Fig. 4(b). We also integrated a dialogue exploration functionality, in which children could select certain areas of the chart and examine the dialogue (Fig. 4(c)).

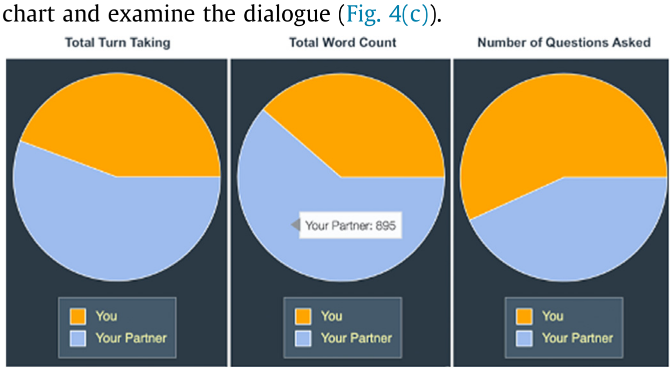

Fig. 3. Pie charts illustrating the total turn-taking count, total word count and total question count.

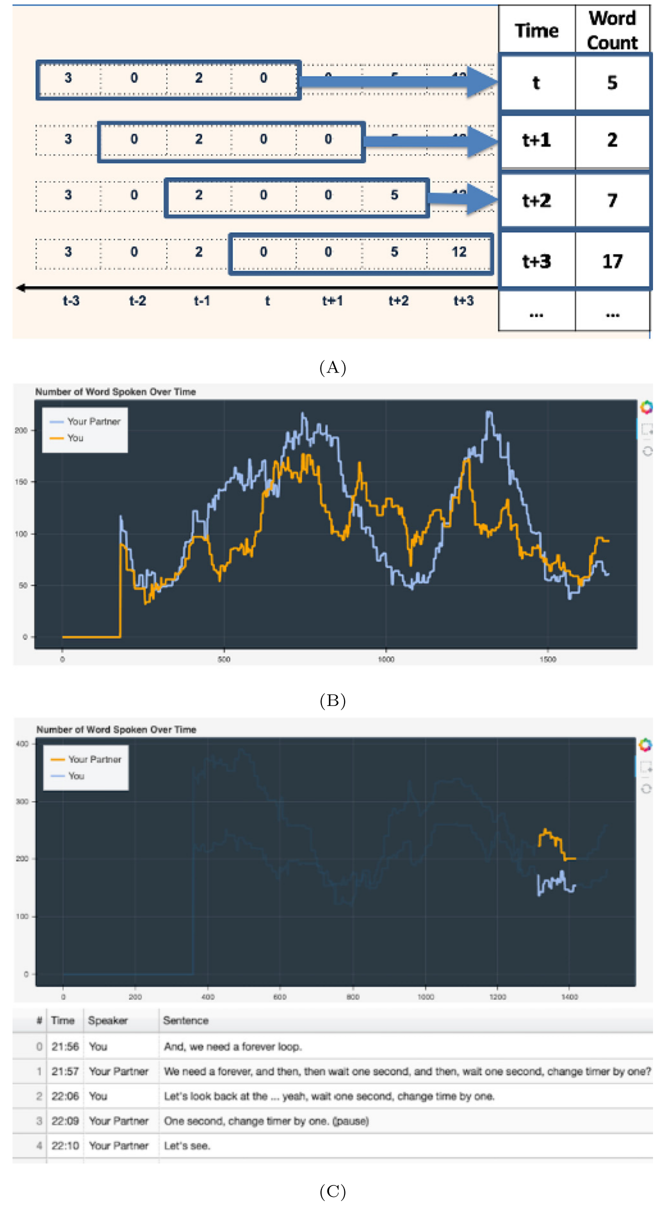

Fig. 4. Interactive line chart. (A) Sliding window example with the window size of four. The window slides after taking the sum of the word counts in the selected time window. (B) Interactive line chart illustrating the number of words spoken over time. (C) Interactive line chart with a subsection of dialogue selected, along with transcript exploration feature.

3.2. STUDY 1: Comparisons of pie charts and time-series visualizations

Previous dialogue visualization approaches have primarily relied upon summative pie charts or bar charts to illustrate the number of dialogue contributions across an entire conversation. We believe this approach misses important information about the temporal nature of dialogue; thus we hypothesized that time-series line charts may be more suitable for understanding the flow of collaborative dialogue. However, time-series charts are under-investigated in dialogue visualization research, and are even further under-investigated with children whose stage of mathematical understanding is an important consideration. Based on the State Standards where our work was conducted, children in seventh grade should be exposed to graphs and statistical reasoning. In first grade, children focus on organizing, representing, and interpreting data. In second grade they move on to drawing graphs (i.e. bar graphs) to represent data. As they move into upper elementary school, they start focusing on the coordinate system. By the seventh grade, they focus on "Strategic Thinking & Complex Reasoning", focusing on interpreting graphs and reasoning about their real-world implications. Therefore, it is important to examine children's understanding of the time-series line chart compared to commonly used summative pie charts. The second goal of this study was to place the children at the center of the design process, by receiving their feedback and investigating their expectations of dialogue visualizations. The following sections describe the first study.

3.2.1. STUDY 1: Participants and context

We conducted the first study in a middle school science classroom in the southeastern United States during Spring 2019. Before beginning the studies, we received approval from our university's Institutional Review Boards (IRB), which examined all the details of this study to protect the rights and welfare of participants. We visited the classrooms, described the study to the children, and verbally explained the key information in the consent form. Children were asked to review the consent forms and take them to their parents for signing if they were willing to participate in the study. Before each study session, 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 and excluded them from the data analysis. There was no penalty for not participating in the studies and all the children (including children who were not part of the study) did the same class activities.

Out of the 97 children in five different class sessions, 75 children's parents consented to data collection. Of those consenting children, we randomly selected 18 children to be recorded and participate in the think-aloud sessions, limited by the number of available on-site researchers. Of the 18 children (ages 12–13), there were 11 female and 7 male children who described themselves as White (7), Asian (6), Hispanic (1), and Multiracial (4). Eleven of these students reported having had some prior coding experience (e.g., attending a girls' camp, taking an elective robotics class) at the beginning of the semester. The focus of this research study is not to evaluate how children's collaboration skills change based on their existing knowledge, but rather to design and develop a visualization tool based on their understanding and expectations. Thus, their pre-knowledge of computer science is beyond the scope of this study.

The class met five times per week and children had been coding in *Snap!* (*Snap! Build your own blocks*, 2020) with a partner (either a randomly assigned or self-selected partner depending on the lesson) since the beginning of the academic year. Children learned various computer science (CS) concepts such as variables, conditionals, loops, and object-oriented programming, and created computationally rich science programs based on the lesson topics (e.g., the food web, evolution). The goal of these activities was to allow children to practice computer science concepts alongside life science concepts to learn how to meaningfully bridge concepts across these different disciplines (Celepkolu et al., 2020). Fig. 6 shows the *Evolution and Natural Selection* CS+Science activity, in which children modeled how the physical size of two different butterfly species can impact their survival and population-size changes over time. For each activity, the researchers presented a CS+Science topic and provided children with a printed copy of the instructions as well as a reference sheet of useful code blocks.

During the coding activities, children followed the pair programming methodology, which has been shown to be an effective collaboration approach for improving productivity (Celepkolu & Boyer, 2018a; Nagappan et al., 2003) and promoting good programming practices (Umapathy & Ritzhaupt, 2017). In pair programming, two students work on the same code with two different roles: the driver is responsible for writing the code and implementing the solution, while the navigator is responsible for assisting with catching mistakes and providing immediate feedback, as shown in Fig. 5. It is expected that students switch roles regularly. Before we implemented the visualization studies, students had the opportunity to work with several partners on coding activities and were comfortable with the pair programming paradigm.

Fig. 5. Two children working during a pair programming activity.

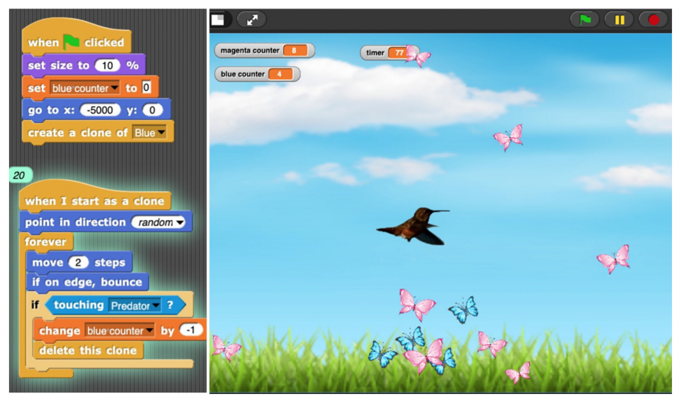

Fig. 6. Sample *Evolution* activity created with Snap*!* block-based programming language.

3.2.2. STUDY 1: Procedure and data collection

The data was collected as part of a three-day study. On Day 1, children were randomly assigned partners to complete an *Evolution* Activity (Fig. 6) using pair programming, which most groups finished in about 30 min.

On Day 2, the interactive visualizations of their previous day's collaboration were presented to the children following the think-aloud protocol (Ericsson & Simon, 1993). Our goal was to gain an understanding of the ways children choose to interact with the visualizations, and investigate their perceptions, preferences, and expectations of the dialogue visualizations. We conducted the think-aloud sessions in three phases. First, we showed learners the pie charts on a page and asked them about their first impression and understanding of the charts. We also asked them what they were thinking as they were looking/clicking and what the numbers on the charts meant to them. Second, we showed them the interactive line chart on another page following the same procedure. Lastly, we asked them questions such as whether they would want to use these types of charts in the future, which chart would be most helpful for them, and whether they had any design suggestions.

On Day 3, pairs completed another *Evolution* activity in *Snap!* and responded to several questions about the visualizations that were shown to them in the previous activity. Some sample questions were, "How much did the graphs change your interaction with your partner?", and "Would you like to use these graphs for your future group work?".

3.2.3. STUDY 1: Analysis

To analyze the children's open-ended responses to the questions during the think-aloud sessions, we applied a qualitative content analysis approach. Three researchers independently labeled the children's responses for each research question to avoid potential subjectivity bias derived from qualitative data analysis. In this section, we present the description of the steps we followed for analyzing each research question.

RQ1: Each visualization in the application had a purpose to convey information about the dialogue to the children, and our goal was to understand whether children interpreted the visualizations (three pie charts and the line chart) as intended. We conducted think-aloud interviews (Ericsson & Simon, 1993), which have been widely used in user studies with children to investigate their thinking process while performing a task (Lee, Mauriello, Ahn, & Bederson, 2014; Tsvyatkova & Storni, 2019; Van Kesteren, Bekker, Vermeeren, & Lloyd, 2003). During the think-aloud sessions, we asked children to interpret each graph one by one, as well as explain the components of each graph (*e.g.*, *x*-axis of the line chart; legends of a pie chart) with their own words. When a visualization was not interpreted as intended, we investigated the reasons behind the differing interpretation and sought suggestions for further improvements from the child. We audio/video recorded both the participants and the computer screen while the children were using the application during the think-aloud sessions.

After collecting the data, we manually transcribed the recorded data and examined children's responses to the visualizations immediately after they were introduced to the visualizations. Our goal was to identify children's responses as soon as they saw the visualizations and capture their first impression and understanding. The researchers independently rated each students' response to the three pie charts (total turn taking, total word counts, and total question counts) and the line chart visualizations as "1" (indicating complete correctness), or "0" (partial or no correctness). Our goal was to be very critical in the design of the application and thus, even a partial correct answer was rated as "0" as it indicates there is still some need for further improvement. After the researchers rated each children's responses to each chart, we computed the inter-rater reliability (Kappa (Cohen, 1960)) scores, a value ranging from −1 (perfect disagreement) to 1 (perfect agreement) showing the consensus between judges for each chart.

RQ 2: The dialogue visualization application serves as a mirroring tool that illustrates some components of children's previous dialogue with each other. During the think-aloud sessions, children reflected on their interaction with their partner and evaluated what could be improved for better group dialogue in the future. While children were interacting with the dialogue application during the think-aloud sessions, we asked questions such as "What are you thinking as you look/click/read?", "What made you select that part of the graph?" and "Are you noticing anything interesting? Why?."

After collecting the data, three researchers open-coded children's responses: they created initial semantic codes representing small amounts of data and entered them on an Excel sheet. Next, they met to discuss each code created for each response and refined the initial codes for the corresponding data to ensure the codes were unique. Next, they adopted an iterative-inductive approach (Elo & Kyngäs, 2008) to group the codes and identify bigger categories indicating how children would use these visualizations to reflect on their collaborative dialogue.

During the think-aloud sessions, we also investigated whether the children would be interested in using these visualizations in the future. To analyze the responses, we again followed the same steps of independently coding the responses and creating higher-level categories as described in this section.

Finally, we examined the children's responses to the survey questions on whether they felt that they had changed their behavior in the subsequent collaborative activity (Day 3) after seeing the visualizations on Day 2.

RQ 3: We investigated the children's feedback on the new integrated features as described above, and we also explored their new design suggestions for future implications. During the think-aloud sessions, we asked questions such as "Can you think of anything else helpful we should have graphed for you?" and qualitatively explored their reasoning for those suggestions and expectations from future designs. Similar to the analysis method for RQ1 and RQ2, three researchers open-coded children's responses, applied an iterative-inductive approach (Elo & Kyngäs, 2008) to group the codes, and identified broader categories reflecting the data for this research question.

Table 1
Count of children who interpreted each chart as intended in Study 1. Interpretations were manually coded by researchers.

Chart type	*Interpreted as intended*	*Kappa score*
Turn taking count	7 / 18	.77
Word count	17 / 18	.93
Question count	18 / 18	.85
Line chart	18 / 18	1

3.2.4. STUDY 1: Results

Think-aloud sessions revealed various important points about children's perceptions of the visualizations (RQ1), how they might employ dialogue visualizations to reflect on their collaborative dialogue (RQ2), and their design suggestions for the dialogue visualizations (RQ3). We discuss each of these points by presenting illustrative excerpts extracted from the children's think-aloud sessions and written responses to the open-ended questions after the CS+Science activities.

RQ1: How do children perceive visualizations illustrating their collaborative dialogue?

We first investigated whether these charts were understandable (interpreted as intended) by the children, what design issues may be present in the application, and whether they find these visualizations useful and beneficial. Table 1 shows the number of children who interpreted the charts as intended and the inter-rater reliability (Kappa (Cohen, 1960)) scores for each chart.

The children were able to interpret most charts as intended except the total turn-taking pie chart. Out of 18 children, 11 struggled to understand this chart, and of those, 10 thought the pie chart showed the amount of time used for controlling the computer. Given that the children were taking turns depending on their roles as driver or navigator during the activities, the label "total turn taking" appeared to refer to the driving turn, not dialogue turns. The remaining child misinterpreted the turn-taking chart as the word-count chart, and vice versa. All children interpreted the question-count pie chart and the line chart as intended.

The children often found these visualizations beneficial for helping them become better teammates and balance the dialogue with their partners. For example, some children reported not being aware of talking more/less during the interaction and they would talk more/less to make it balanced:

"I think this is good, 'cause it shows who talked when and what they said; so it helps them, like, -Oh, I need to talk more, I need to talk less, compared to what my partner did."

"Well, that, I obviously, could've talked more, because the graphs were so different and he obviously, clearly, talked way more and I need to ask more questions, and maybe contribute more."

The children talked about the line chart much more than the pie charts and emphasized the importance of seeing the dialogue over time:

"I personally like the line graph more because you can drag and it says exactly what we said, whereas here it just says how many like words we spoke and stuff."

"... because you can see like, everything... Like, it's easier to see over time."

RQ2: How might children employ dialogue visualizations to reflect on their collaborative dialogue?

The children often reflected on their dialogue while exploring the charts. While some children highlighted the interval parts related to their own dialogue, some focused on the times in which the group talked a lot/equal/a little compared to other parts of the chart. They also mentioned that splitting the talk time shows that their conversation went well.

"...the conversation with my partner went very well because we even split the time talking and stuff and asking each other, and help each other learn what to do and stuff."

The children also often made comments reflecting their thoughts about the activity:

"...this is probably when we were making the cloned area right here. He was talking a lot here. Yeah we were creating the clones, and during this time-frame, he was saying a lot more than I was because he was saying things like, 'Oh this should be right here."

Some children reflected that the content of their dialogues was different as most of his/her talk was about providing task-related instruction to his/her partner, whereas his/her partner just talked about random things:

"I ended up talking more about telling her how to do the actual thing. Then she ended up talking more about random things."

While the children read over the dialogue between themselves and their partner, some children explained that while in the driver role, they were busy with implementing the coding solution, which led them to talk less.

"Interesting in the sense that maybe during when he was working, I would talk more. Maybe when I was working he would talk more. And this part here ... oh yeah we were doing the majority of our work in these two sections right here..."

Next, we examined whether children would be interested in using these visualizations in the future. Out of 18 children, only two children said they would not be interested in using these visualizations. Out of 16 children who gave positive responses for using the charts in the future, nine children preferred the line chart, four children preferred the pie chart and two children preferred both types of charts. The most common reasons to use the line chart were that it presents more details, shows when they talked the most/least, shows the dialogue content, and is useful for record keeping for future use (remembering how they solved the problem). For example, one child said:

"It doesn't just tell you how many words you spoke and things. It tells you exactly when you spoke and how much you spoke at that time, so it's more detailed."

Finally, we examined whether children felt that they had changed their behavior in the subsequent collaborative activity (Day 3) after seeing the visualizations on Day 2. Out of 18 children who participated in the visualization study day, 11 of them reported that seeing the visualizations of their previous conversation changed the way they talked with their partner. They reported that they took their partner more seriously, let their partner speak when he/she had something to say, tried harder not to interrupt their partner, and paid more attention to thinking deeply about their partner's questions. On the other hand, six children reported that the visualizations did not change the way they talked with their partner and there was one child with missing data due to a technical error.

RQ3: What design implications emerge from children's feedback on the dialogue visualizations?

The findings from the think-aloud sessions indicated that the interactive time-series chart is promising for helping children explore the evolving nature of the dialogue during pair programming activities, and may support productive reflections on an individual child's dialogue behaviors as well as the group dynamic. Most of the children preferred the line charts; yet, some children indicated the benefits of the pie charts and expressed a desire to use both charts. This suggestion led us to consider merging these two approaches on one page in the next iteration so that children can benefit from both visualizations. Also, some children wanted to see the video of their interactions with their partners as well as the screen recordings of their activity, to make it easier to know what they were talking about while exploring the dialogue. We also integrated these suggestions into the next version. The other suggestions were to show the distribution of on/off-topic conversations, question types (e.g., relevance to coding), partner's openness to ideas, amount of time taken to figure out a problem, and number of mistakes they made during the activity. These potential enhancements are left to future work.

3.3. ITERATION 2: Merging the charts and additional visuals

In this iteration, we made the following updates to create the second version of the tool (Fig. 1) based on the feedback received from the children in Study 1: (1) We put all the visualizations on the same page. (2) We combined the line chart and the pie charts. When a child highlights a specific sub-region of the line chart, the pie chart is updated to reflect counts from the highlighted region in a dynamic way. This allows children to benefit from both types of charts and gives them the freedom of exploring the same data on their preferred chart. (3) We integrated the screen recording of the children's work as well as the video recordings of their group interactions. The children could switch between these videos by clicking on the tabs.

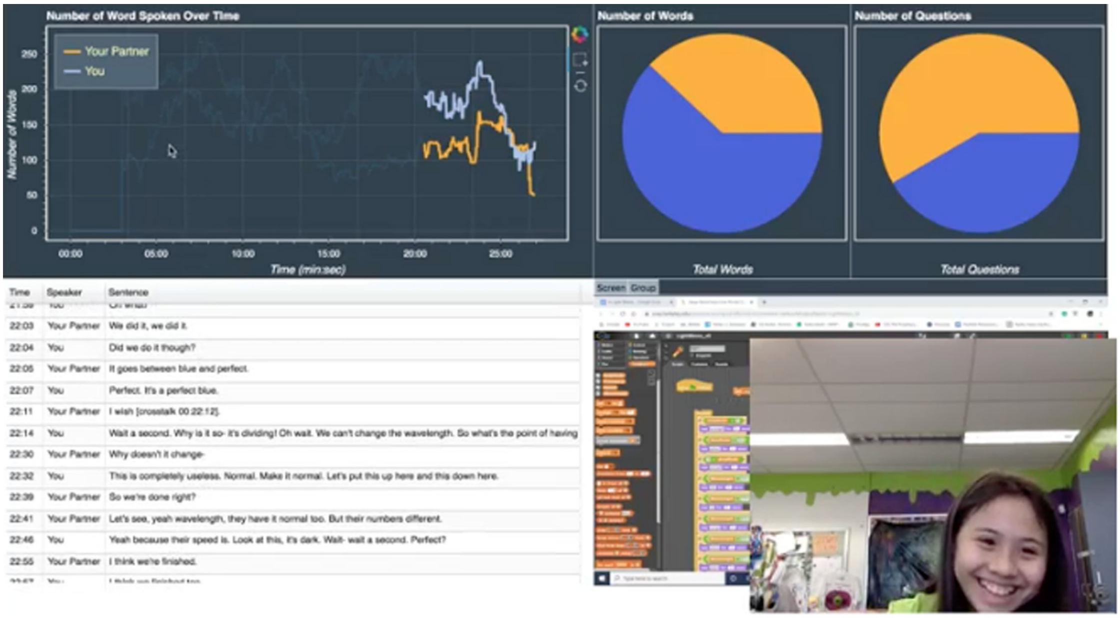

Fig. 7. Think-aloud Session.

3.4. STUDY 2: Investigating children's feedback and reflections on the second version

The overarching goal of the second study was to elicit the children's feedback on the new version, and examine the reflections from the dialogue visualization application. The following sections describe the participants, context, data collection procedure and the results of the second study.

3.4.1. STUDY 2: Participants and context

This study was conducted with different children in the same teacher's science class in Fall 2019. The class again met five times per week and the children had been coding with a partner (either a randomly assigned or self-selected partner depending on the lesson) since the beginning of the academic year. Out of 107 children in the class, 70 consented to data collection, and nine groups (18 children) were randomly selected for this study. Of the 18 children, there were 11 female and seven male children. Children described themselves as White (8), Asian (4), Hispanic (1), and Multiracial (5). These children had also previously learned the *Snap!* programming language, practiced various CS concepts, and followed the pair programming paradigm during coding, similar to the participant population in Study 1.

3.4.2. STUDY 2: Procedure and data collection

The second study was conducted over a three-day period similar to the first study. On Day 1, the children were randomly assigned partners to create a *Light Waves* computer program that would classify the color and the brightness of the light based on the amplitude and the wavelength of the light. On Day 2, the children interacted with the second version of the interactive visualization tool in think-aloud sessions (Fig. 7). In contrast to the first study, we did not ask the children to make any comparisons; rather, we investigated their feedback and reflections on the new version of the visualization tool as a whole (Fig. 1). On Day 3, pairs programmed a *Homeostasis* computer program that would determine the homeostasis status of a human model based on several different variables such as temperature and physical condition. To analyze the data, two researchers followed the same steps as the first study, described in Section 3.2.3.

Table 2
Count of children who interpreted each chart as intended in Study 2. Interpretations were manually coded by researchers.

Chart type	*Interpreted as intended*	*Kappa score*
Line chart	15 / 18	82.4
Word count	18 / 18	1
Question count	17 / 18	89.5
Screen video	18 / 18	1
Group video	18 / 18	1

3.4.3. STUDY 2: Results

In this section, we present children's interpretations of the visualizations, their feedback on the benefit of these visualizations and whether they would want to use the visualizations in the future, their feedback on the graphs and reflections on their group work, and their design suggestions for future iterations.

RQ1: How do children perceive visualizations illustrating their collaborative dialogue?

Again, we investigated whether these charts were understandable (interpreted as intended) by the children, what design issues may be present in the application, and whether they find these visualizations useful and beneficial, following the same methodology as previously described. Table 2 shows the number of children who interpreted the charts as intended and the inter-rater reliability scores.

Out of 18 children, two children thought the line chart was showing the sound waves or the sound level of their dialogue because the line charts were similar to sound waves. One child thought the line chart was showing the number of words they typed while working on the computers. It is important to note that the rating reported in the table is based on the children's responses immediately after the charts were introduced to them. All these children corrected themselves after they noticed the labels on the line chart (in several seconds). One child thought the question-count pie chart was actually showing the number of questions he asked and how many questions his partner was able to answer.

The children also provided feedback on how they would benefit from these visualizations to become better teammates and balance the dialogue with their partners:

"I really... Maybe it would teach me to be a better partner. If I'm talking too much and not letting my partner talk as much or share his opinion. I think I could definitely do that, and stay more silent."

"...we could just see how we interacted so that next time we could interact better."

Some children mentioned that these visualizations may be useful for increasing their awareness during group work:

"I know some people, some people are really quiet [...] they might not realize that their partner needs help or that they're not being helpful. So if they see this they might be like, 'Oh I should talk more. I should ask more questions'."

Another point was that these visualizations could increase their interaction with their partner:

"... if you see that you're not interacting enough, not helping, you'll be encouraged to help more..."

Finally, some children reported positive feelings toward the interactivity features:

"I'm just curious to look at like what's going on."

"I think it's cool that shows like what we said and the questions we asked and stuff. And it shows like what we did in this time period."

RQ2: How might children employ dialogue visualizations to reflect on their collaborative dialogue?

During the think-aloud sessions, most children highlighted various sections on the line charts to examine the dialogue during that time and often reflected about the details of their dialogues with their partners while interpreting the charts. For example, one child interpreted the well aligned lines as an indicator of excitement or talking about the same thing:

"...we were both maybe they were talking about the same thing because it's aligned."

"...we just were both talking a lot because we were excited or something."

Another child pointed out that the similarity of the graphs would be an indicator of having similar knowledge:

"So I would say we probably know about the same because the graphs are very similar."

"Like in some section she talks a lot more than I do, but like, in others I talk more and that's cool."

Some children described their understanding of how a good collaboration looks and how some of their previous collaborative learning sessions were frustrating. For example, one child reported that one of his teammates was not sharing the workload (free-riding), which was unfair to him:

"Sometimes I get in a group with my friends and then recently there was one kid that came over but he didn't really do anything and he still got credit for the project, which I felt like that was a bit unfair."

Some children described the good collaboration as an interaction in which both sides ask questions and contribute:

"I guess it's kind of good because we both equaled in the amount of things we did. We both spoke the... So that means we both input the same amount and we asked questions that are important for us."

"That I would be respectful to them, and always use their ideas, and let them speak and do what they want to do. Instead of doing what you always want to do."

Next, we again examined whether children would be interested in using these visualizations in the future. Out of 18 children, 17 children gave positive responses for using the charts and only one child said she would not be interested in using these visualizations in the future because she already remembered her group work and did not need these visualizations to remember:

"... I feel like it's cool and like it shows you a lot of cool things, but I have it in my own memory of how we work together and what happened, how we worked. I remember what happened and either way I didn't feel like there's anything I would've changed."

Finally, we examined whether children felt that they had changed their behavior in the subsequent collaborative activity (Day 3) after seeing the visualizations of their interaction with their partner in the previous activity on Day 2. Out of 18 children who participated on the visualization study day, 16 were present on Day 3 and of these, 10 reported that seeing the visualizations of their previous conversation changed the way they talked with their partner in their short-answer written feedback:

"In the last graphs, I saw that I didn't talk very much, so I tried to talk a little bit more. I also tried to help my partner more."

"I saw that i talked more so i tried to let her talk more. that is why i did more coding today then last time also it was really cool to just see how we acted and what we said."

On the other hand, six children wrote that they didn't think about the dialogue visualizations while working with their partners in the last activity:

"I don't think that the graphs changed my interaction with my partner at all. I didn't really think about them."

RQ3: What design implications emerge from children's feedback on the dialogue visualizations?

The children suggested several ways to improve the tool. For example, they suggested that the videos could be larger, and the font of the dialogue content could also be larger. One child suggested that the graphs could have descriptions about what they represent to make it easier for some children to understand. The children also reported that the screen recording of the activity is more useful than the group recording, but that it is good to have the option to view the group interaction video.

The children also expressed the need for creating visualizations that differentiate between task-related dialogue and off-topic dialogue:

"... a lot of the times us kids we talk about random stuff and this would include it in the number of words and maybe number of questions."

Similarly, one child expressed the desire to categorize the question types. That child believed that most of his questions contributed more to the problem solving, whereas his partner's questions were mostly about receiving help:

"... number of questions should be divided into number of questions based on the concept and number of requests to do something. Because those can both be used as two separate things to gauge the person."

Furthermore, some children asked for a better way to navigate to specific topics in the dialogue. For example, one child suggested using keywords extracted from the dialogue:

"Maybe if there are keywords that you put in and if it hears those words, it'll say number of relevant comments."

Some children asked whether it would be possible to visualize the emotion of the partners toward each other:

"And also perhaps some sort of... I guess emotion would also be important to kind of see the tone of the conversation and how partners are correlating with each other."

A child also suggested comparing the number of words and the questions to all the children in the class to see if they are above or below the average.

4. Discussion and design implications

The overarching goal of the line of investigation reported in this paper was to understand middle school learners' perceptions, preferences, and expectations of dialogue visualizations. We developed two versions of a dialogue visualization tool based on children's feedback and conducted think-aloud sessions in which children could reflect on the dialogue visualizations and help us design more effective systems. The first study revealed that children derive value from both types of charts: while line charts provide more details about the dialogue over time, pie charts give a summary of the work at a glance. After considering the children's feedback, we decided to merge them on the same screen so that children could benefit from both types of charts by focusing on the one that they prefer or that meets their goals at a given time. They also found the screen and group videos useful for understanding their coding activity that had been happening concurrently with their dialogues. Compared to previous research, which often used only one or two of these features, our tool provided more detail about the collaborative dialogue process with different types of visualizations. For example, Kim et al. (2012), Charleer et al. (2017), and Samrose et al. (2018) only used summative visualizations such as pie charts and bar charts to show participation. We added the line chart, which provided even more details by showing how the dialogue process evolved over time. We also added the dialogue content feature, in which children could see textual transcripts of their dialogue with their partner. In addition, similar to the previous studies which showed that children benefit from seeing activity recordings to reflect on their creative trajectories (Dhariwal, 2018), our tool also displays recordings of how the children solved the problems. Additionally, our tool displays the recordings of their interaction with their partner to help them observe their team dynamics.

Dialogue visualization can be a powerful tool in classrooms, allowing children to reflect on their interactions with their partners and critically reason about their contributions. In our studies, children excitedly explored their dialogues and many children said that they considered their findings from the visualizations in their subsequent collaborations.

Children's Capability to Interpret Graphs. One of our questions when we began this work was whether children at this age would be capable of interpreting graphs of dialogue, given their level of mathematical development. We discovered that the seventh grade collaborators from this work had years of increasingly rich experience interpreting graphs, and were able to reason about the data being provided; they even asked insightful questions about how data was being represented. An important limitation to note is that the findings presented here may not generalize to other populations of learners, and the interpretability of dialogue graphs for each intended audience would need to be carefully investigated.

Selecting What to Represent. A central challenge in this work lies in the design of a dialogue visualization interface and determining which visualizations to include. Collaborative dialogue is a rich data source, and the process of informing the interface based on theory as well as children's feedback guided us to the current prototype. It would be easy to overwhelm a user with too much information, and we feel that our current prototype may be at the borderline of how much different information children can usefully engage with in a single view. Further study is needed on this question to ensure an appropriate balance of rich information with simplicity and clarity.

Interactivity to Promote Exploration. One component that children responded well to was being able to interact with the graphs. After first looking over the visualizations on the screen, they intuitively began to attempt to interact with them via clicking, dragging, and hovering. Children expressed interest in filtering to focus on milestone in their dialogues. Many participants would compare their initial dialogue to the end of their session, dragging over sections of each to check how their number of questions changed and what exactly was said by themselves and their partner.

Additional Features to Investigate. The children in our studies wanted more detail about the content of their collaborative dialogue in addition to frequency-related information. The most common design suggestion was for a percentage of on/off-topic utterances, saying it would allow them to see who was contributing the most relevant content to the conversation, not just utterances. Other children requested that the dialogue be clustered by topic, so that they could see different segments of their dialogue unfolding. Finally, they wanted the visualizations to include performance measurements of the code artifacts that they were working on, which would allow for them to see the relationship between the completion/improvement of their code and the unfolding collaborative dialogue.

Another set of requests was the addition of social and emotional indicators. The children were interested in seeing not only information about their sessions, but how that compared to the rest of their class. They were also interested in seeing moments of discomfort or excitement that they experienced during learning.

5. Limitations

The limitations include that this study was conducted in an actual middle school class (not a controlled environment), in which additional complications are always present during study implementations. For example, children interacted with other students and their teacher during the problem-solving activity, not just with their assigned partner within the dialogues we later analyzed. We opted to omit those outside conversational interactions, but they may have influenced the dialogues in important ways.

Second, we conducted these two iterative studies with a relatively small subset of students who were available during the study days. We do not claim generalizability of these findings, and it is important to run similar studies with a larger and more diverse group of students.

Third, the visualization process was fully automated except for transcription of spoken dialogue into textual transcripts. This process took about four days. It is possible that in those days, the children may not have remembered all the details of their interactions with their partner in the previous activity for which they were viewing visualizations.

Finally, and perhaps the most important limitation of all, is that this study relies entirely on children's self-reported data from think-aloud interviews and surveys. We know children may lean toward providing socially acceptable answers in these contexts, and this limitation must be considered when interpreting their favorability toward the tool. Crucially, to ascertain the impact of the tool on future behavior, additional studies are needed that expand beyond self-reports to examine behavioral indicators of change within collaboration.

Despite the limitations of self-reported data, we believe it is imperative to first create an application based on children's understanding and expectations, and only after careful iterative refinement, to proceed with evaluation. Due to the lack of applications aimed for this younger age group, our goal was to follow a user-centered approach that puts the potential users of the application in the center of the design and development process, and create the application based on their understanding and expectations. The studies presented in this manuscript demonstrate that the iteratively refined visualization tool holds great potential for supporting children to reflect on their dialogue and form goals regarding their own behavior.

6. Conclusion and future work

In this paper, we have discussed a series of iterative versions and participatory design sessions with middle school children who explored visualizations of their dialogue with their partner. These studies show that the 7th graders are not only accepting of dialogue visualizations, but had meaningful reflections for their future collaborations. Almost all the children reported that they would want to use these visualization in their future group work, and the majority of the children reported that they felt the change in their behavior in the final collaborative activity. The children reported that they tried to be more balanced in their dialogue and more considerate of their partners during the problem-solving activity after seeing the visualizations.

Moving forward, it is important to develop dialogue visualizations that adapt to children's needs, such as understanding their own contributions, on/off-topic shifts in dialogue, and building a deeper understanding of productive dialogue patterns. A component that we did not have in our designs, which should be heavily considered in future designs, is social information. The children wanted to be able to compare their sessions with the classroom and their peers and determine if their discussions were out of the norm in some ways.

We also acknowledge the importance of the need for evaluating the impact of these dialogue visualizations on children's future collaborative tasks. In future work, we aim to analyze how children's dialogues change (e.g., more balanced talking time, better question-asking behaviors) after seeing the visualizations and which visualizations specifically impact each (potential) change. For example, despite the children's favorable reflections toward time-series visualizations compared to pie charts, it is important to investigate how much each type of visualization impacts outcomes. Future work should examine whether different types of visualizations lead to comparable results (*e.g.*, pie charts versus time-series visualizations).

Another important direction for future work may be to add task-related information such as how much time children spend on specific code tasks or topics. There is also a need for developing fully automated applications. Our application already automates everything except the transcription of the dialogue, and with advancements in automated speech recognition, this visualization will be feasible to implement in real time. This approach will open other important research questions including how to mitigate potential speech recognition errors and exploring methods to extract useful information from dialogues even in noisy environments such as classrooms. Finally, teachers can also provide very useful information about the design of these applications. Future work should investigate involving teachers in the design process to receive feedback in design and evaluation of the potential benefits of these tools from a pedagogical perspective. Overall, this line of investigation holds great potential to empower children to reflect on their collaboration and set their own goals for how they engage in collaborative learning.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by the National Science Foundation, USA 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

Bakhtin, M. M. M. M., & Holquist, M. (1981). *The dialogic imagination : four essays*. University of Texas Press.

Baumgartner, E., Bell, P., Brophy, S., Hoadley, C., Hsi, S., Joseph, D., et al. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, *32*(1), 5–8. http://dx.doi.org/10.3102/0013189X032001005.

Bertin, J. (1981). *Graphics and graphic information processing*. Walter de Gruyter.

Bishop, F., Zagermann, J., Pfeil, U., Sanderson, G., Reiterer, H., & Hinrichs, U. (2020). Construct-A-Vis: Exploring the free-form visualization processes of children. *IEEE Transactions on Visualization and Computer Graphics*, *26*(1), 451–460. http://dx.doi.org/10.1109/TVCG.2019.2934804.

Bodemer, D., & Dehler, J. (2011). Group awareness in CSCL environments. *Computers in Human Behavior*, *27*(3), 1043–1045. http://dx.doi.org/10.1016/j.chb.2010.07.014.

Bokeh 1.1.0 documentation. (2020). URL https://bokeh.pydata.org/en/latest/.

Care, E., Scoular, C., & Griffin, P. (2016). Assessment of collaborative problem solving in education environments. *Applied Measurement in Education*, *29*(4), 250–264. http://dx.doi.org/10.1080/08957347.2016.1209204.

Celepkolu, M., & Boyer, K. E. (2018a). The importance of producing shared code through pair programming. In *Proceedings of the 49th ACM technical symposium on computer science education* (pp. 765–770). ACM, http://dx.doi.org/10.1145/3159450.3159506.

Celepkolu, M., & Boyer, K. E. (2018b). Thematic analysis of students' reflections on pair programming in CS1. In *Proceedings of the 49th ACM technical symposium on computer science education* (pp. 771–776). ACM, http://dx.doi.org/10.1145/3159450.3159516.

Celepkolu, M., Fussell, D. A., Galdo, A. C., Boyer, K. E., Wiebe, E. N., Mott, B. W., et al. (2020). 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* (pp. 671–677). ACM, http://dx.doi.org/10.1145/3328778.3366871.

Charleer, S., Klerkx, J., Duval, E., Laet, T. D., & Verbert, K. (2017). Towards balanced discussions in the classroom using ambient information visualisations. *International Journal of Technology Enhanced Learning*, *9*(2/3), 227. http://dx.doi.org/10.1504/IJTEL.2017.084501.

Chen, B., & Resendes, M. (2014). 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* (pp. 226–230). http://dx.doi.org/10.1145/2567574.2567606.

Chen, Z., Yang, R., Zhao, Z., Cai, D., & He, X. (2018). Dialogue act recognition via CRF-attentive structured network. In *The 41st international ACM SIGIR conference on research & development in information retrieval* (pp. 225–234). ACM, http://dx.doi.org/10.1145/3209978.3209997.

Chinn, C. A., O'donnell, A. M., & Jinks, T. S. (2000). The structure of discourse in collaborative learning. *The Journal of Experimental Education*, *69*(1), 77–97. http://dx.doi.org/10.1080/00220970009600650.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*(1), 37–46.

Datar, M., Gionis, A., Indyk, P., & Motwani, R. (2002). Maintaining stream statistics over sliding windows. *SIAM Journal on Computing*, *31*(6), 1794–1813. http://dx.doi.org/10.1137/S0097539701398363.

Davis, P., Horn, M., Block, F., Phillips, B., Evans, E. M., Diamond, J., et al. (2015). "Whoa! we're going deep in the trees!": Patterns of collaboration around an interactive information visualization exhibit. *International Journal of Computer-Supported Collaborative Learning*, *10*(1), 53–76. http://dx.doi.org/10.1007/s11412-015-9209-z.

Deitrick, E., Shapiro, R. B., & Gravel, B. (2016). How do we assess equity in programming pairs?. In *Proceedings of the 12th international conference on computer supported collaborative learning (CSCL)* (pp. 370–377). International Society of the Learning Sciences, (ISLS).

Dhariwal, S. (2018). Scratch memories: A visualization tool for children to celebrate and reflect on their creative trajectories. In *IDC 2018 - Proceedings of the 2018 ACM conference on interaction design and children* (pp. 449–455). ACM, http://dx.doi.org/10.1145/3202185.3202770.

DiMicco, J. M., Pandolfo, A., & Bender, W. (2004). Influencing group participation with a shared display. In *Proceedings of the 2004 ACM conference on computer supported cooperative work (CSCW)* (pp. 614–623). New York, New York, USA: ACM Press, http://dx.doi.org/10.1145/1031607.1031713.

Dowell, N. M. M., Nixon, T. M., & Graesser, A. C. (2018). Group communication analysis: A computational linguistics approach for detecting sociocognitive roles in multiparty interactions. *Behavior Research Methods*, 1–35. http://dx.doi.org/10.3758/s13428-018-1102-z.

Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing*, *62*(1), 107–115. http://dx.doi.org/10.1111/j.1365-2648.2007.04569.x.

Engle, R. A., Langer-Osuna, J. M., & McKinney de Royston, M. (2014). Toward a model of influence in persuasive discussions: Negotiating quality, authority, privilege, and access within a student-led argument. *Journal of the Learning Sciences*, *23*(2), 245–268. http://dx.doi.org/10.1080/10508406.2014.883979.

Ericsson, A., & Simon, H. A. H. A. (1993). *Protocol analysis : verbal reports as data* (p. 443).

Fransen, J., Kirschner, P. A., & Erkens, G. (2011). Mediating team effectiveness in the context of collaborative learning: The importance of team and task awareness. *Computers in Human Behavior*, *27*(3), 1103–1113. http://dx.doi.org/10.1016/J.CHB.2010.05.017.

Galton, M., & Williamson, J. (1992). *Group work in the primary classroom*. Routledge.

Janssen, J., Erkens, G., & Kirschner, P. A. (2011). Group awareness tools: It's what you do with it that matters. *Computers in Human Behavior*, *27*(3), 1046–1058. http://dx.doi.org/10.1016/J.CHB.2010.06.002.

Kapur, M. (2011). Temporality matters: Advancing a method for analyzing problem-solving processes in a computer-supported collaborative environment. *International Journal of Computer-Supported Collaborative Learning*, *6*(1), 39–56. http://dx.doi.org/10.1007/s11412-011-9109-9.

Kim, J. Y., Calvo, R. A., Yacef, K., & Enfield, N. J. (2019). A review on dyadic conversation visualizations - Purposes, data, lens of analysis. arXiv.

Kim, T., Hinds, P., & Pentland, A. (2012). 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)* (pp. 1237–1246). ACM, http://dx.doi.org/10.1145/2145204.2145391.

Knight, S., & Littleton, K. (2007). *Dialogue and the development of children's thinking: a sociocultural approach*. Routledge.

Knight, S., & Littleton, K. (2015). Dialogue as data in learning analytics for productive educational dialogue.. *Journal of Learning Analytics*, *2*(3), 111–143.

Kreijns, K., Kirschner, P. A., & Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Computers in Human Behavior*, *19*(3), 335–353. http://dx.doi.org/10.1016/S0747-5632(02)00057-2.

Kuhn, D. (2018). A role for reasoning in a dialogic approach to critical thinking. *Topoi*, *37*(1), 121–128. http://dx.doi.org/10.1007/s11245-016-9373-4.

Kuhn, D., & Crowell, A. (2011). Dialogic argumentation as a vehicle for developing young adolescents' thinking. *Psychological Science*, *22*(4), 545–552. http://dx.doi.org/10.1177/0956797611402512.

Lee, T. Y., Mauriello, M. L., Ahn, J., & Bederson, B. B. (2014). CTArcade: Computational thinking with games in school age children. *International Journal of Child-Computer Interaction*, *2*(1), 26–33. http://dx.doi.org/10.1016/j.ijcci.2014.06.003.

Lewis, C. M., & Shah, N. (2015). How equity and inequity can emerge in pair programming. In *Proceedings of the 11th annual international conference on international computing education research (ICER)* (pp. 41–50). ACM, http://dx.doi.org/10.1145/2787622.2787716.

Littleton, K., & Mercer, N. (2013). *Interthinking: Putting talk to work*. Routledge.

Major, L., Warwick, P., Rasmussen, I., Ludvigsen, S., & Cook, V. (2018). Classroom dialogue and digital technologies: A scoping review. *Education and Information Technologies*, *23*(5), 1995–2028.

McKenney, S., & Reeves, T. C. (2014). Educational design research. In *Handbook of research on educational communications and technology: Fourth edition* (pp. 131–140). Springer, http://dx.doi.org/10.1007/978-1-4614-3185-5_11.

Mercer, N. (2008). Talk and the development of reasoning and understanding. *Human Development*, *51*(1), 90–100. http://dx.doi.org/10.1159/000113158.

Mercer, N. (2013). The social brain, language, and goal-directed collective thinking: A social conception of cognition and its implications for understanding how we think, teach, and learn. *Educational Psychologist*, *48*(3), 148–168. http://dx.doi.org/10.1080/00461520.2013.804394.

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

Nagappan, N., Williams, L., Ferzli, M., Wiebe, E., Yang, K., Miller, C., et al. (2003). Improving the CS1 experience with pair programming. *ACM SIGCSE Bulletin*, *35*(1), 359. http://dx.doi.org/10.1145/792548.612006.

Nova, N., Wehrle, T., Goslin, J., Bourquin, Y., & Dillenbourg, P. (2007). Collaboration in a multi-user game: impacts of an awareness tool on mutual modeling. *Multimedia Tools and Applications*, *32*(2), 161–183. http://dx.doi.org/10.1007/s11042-006-0065-8.

Nystrand, M., Wu, L. L., Gamoran, A., Zeiser, S., & Long, D. A. (2003). Questions in time: Investigating the structure and dynamics of unfolding classroom discourse. *Discourse Processes*, *35*(2), 135–198. http://dx.doi.org/10.1207/S15326950DP3502_3.

Rosenbaum, E. (2009). *Jots: Cultivating reflective learning in scratch* (Ph.D. thesis), Massachusetts Institute of Technology.

Samrose, S., Zhao, R., White, J., Li, V., Nova, L., Lu, Y., et al. (2018). CoCo: Collaboration coach for understanding team dynamics during video conferencing. In *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies, Vol. 1* (pp. 1–24). ACM, http://dx.doi.org/10.1145/3161186.

Schultz, J. L., Wilson, J. R., & Hess, K. C. (2010). Team-based classroom pedagogy reframed: The student perspective. *American Journal of Business Education*, *3*(7), 17–24.

Siirtola, H. (2014). Bars, pies, doughnuts & tables – Visualization of proportions. In *Proceedings of the 28th international BCS human computer interaction conference (HCI)* (pp. 240–245). BCS Learning & Development, http://dx.doi.org/10.14236/ewic/hci2014.38.

Snap! build your own blocks. (2020). URL https://snap.berkeley.edu/.

Soller, A., Martinez, A., Jermann, P., & Muehlenbrock, M. (2005). From mirroring to guiding: A review of state of the art technology for supporting collaborative learning. *International Journal of Artificial Intelligence in Education*, *15*(4), 261–290.

Tsan, J., Lynch, C. F., & Elizabeth Boyer, K. (2018). "Alright, what do we need?": A study of young coders' collaborative dialogue. *International Journal of Child-Computer Interaction*, *17*, 61–71. http://dx.doi.org/10.1016/j.ijcci.2018.03.001.

Tsvyatkova, D., & Storni, C. (2019). A review of selected methods, techniques and tools in Child–Computer Interaction (CCI) developed/adapted to support children's involvement in technology development. *International Journal of Child-Computer Interaction*, *22*, http://dx.doi.org/10.1016/j.ijcci.2019.100148.

Turner, J. C., Christensen, A., Kackar-Cam, H. Z., Trucano, M., & Fulmer, S. M. (2014). Enhancing students' engagement: Report of a 3-year intervention with middle school teachers. *American Educational Research Journal*, *51*(6), 1195–1226. http://dx.doi.org/10.3102/0002831214532515.

Umapathy, K., & Ritzhaupt, A. D. (2017). A meta-analysis of pair-programming in computer programming courses: Implications for educational practice. *ACM Transactions on Computing Education*, *17*(4), http://dx.doi.org/10.1145/2996201.

Van Kesteren, U. E., Bekker, M. M., Vermeeren, A. P., & Lloyd, P. A. (2003). Assessing usability evaluation methods on their effectiveness to elicit verbal comments from children subjects. In *Proceedings of the 2003 conference on interaction design and children (IDC)* (pp. 41–49). ACM, http://dx.doi.org/10.1145/953536.953544.

Wang, F., & Hannafin, M. J. (2005). Design-based research and technology-enhanced learning environments. *Educational Technology Research and Development*, *53*(4), 5–23. http://dx.doi.org/10.1007/BF02504682.

Wang, R., Zhu, X., Nacenta, M., & Dai, Q. (2017). Visualization of health behavior data for children and young adults. In *Proceedings of computing conference 2017* (pp. 1207–1216). IEEE, http://dx.doi.org/10.1109/SAI.2017.8252244.

Wegerif, R. (2011). Towards a dialogic theory of how children learn to think. *Thinking Skills and Creativity*, *6*(3), 179–190. http://dx.doi.org/10.1016/J.TSC.2011.08.002.

Wegerif, R., McLaren, B. M., Chamrada, M., Scheuer, O., Mansour, N., Mikšátko, J., et al. (2010). Exploring creative thinking in graphically mediated synchronous dialogues. *Computers & Education*, *54*(3), 613–621. http://dx.doi.org/10.1016/J.COMPEDU.2009.10.015.

Weinberger, A., Stegmann, K., & Fischer, F. (2010). Learning to argue online: Scripted groups surpass individuals (unscripted groups do not). *Computers in Human Behavior*, *26*(4), 506–515. http://dx.doi.org/10.1016/J.CHB.2009.08.007.

Wiltshire, T. J., Butner, J. E., & Fiore, S. M. (2018). Problem-solving phase transitions during team collaboration. *Cognitive Science*, *42*(1), 129–167. http://dx.doi.org/10.1111/cogs.12482.

Yamada, M., & Goda, Y. (2018). The effects of social presence visualization based on a community of inquiry .... *2018*, In *Proceedings of society for information technology & teacher education international conference* (1), (pp. 1014–1019). Association for the Advancement of Computing in Education (AACE).

Yamada, M., Kaneko, K., & Goda, Y. (2016). Social presence visualizer: Development of the collaboration facilitation module on CSCL. In *Proceedings of the international conference on collaboration technologies* (pp. 174–189). Springer, http://dx.doi.org/10.1007/978-981-10-2618-8_14.