

Research Paper

The impact of near-peer virtual agents on computer science attitudes and collaborative dialogue



Toni V. Earle-Randell ^{a,*}, Joseph B. Wiggins ^a, Yingbo Ma ^a, Mehmet Celepkolu ^a, Dolly Bounajim ^b, Zhikai Gao ^b, Julianna Martinez Ruiz ^a, Kristy Elizabeth Boyer ^a, Maya Israel ^a, Collin F. Lynch ^b, Eric Wiebe ^b

^a University of Florida, Department of Computer & Information Science & Engineering, Gainesville, FL, USA

^b North Carolina State University, Department of Computer Science, Raleigh, NC, USA

ARTICLE INFO

Keywords:

Virtual learning companions
Elementary school
Computer science learning
Collaborative learning

ABSTRACT

Virtual learning companions, or pedagogical agents situated as “near peers”, have shown great promise for supporting learning, but little is known about their potential to scaffold other practices, such as collaboration. We report on the development and evaluation of a first-of-their-kind pair of virtual learning companions, designed to model good collaborative practices for dyads of elementary school learners, that are integrated within a block-based coding environment. Results from a study with fifteen dyads of children indicate that the learning companions fostered more higher-order questions and promoted significantly higher computer science attitude scores than a control condition. Qualitative analyses revealed that most children perceived the virtual learning companions as helpful, felt that the companions changed their interaction with their partners, and wanted to have the companions in their future work. These results highlight the potential for virtual learning companions to scaffold collaboration between young learners and provide direction for future investigation on the role that near-peer agents play in collaborative and task support.

1. Introduction

Children need to develop collaborative learning skills early in their lives in order to learn diquote problem-solving perspectives and integrate multiple viewpoints to create a joint solution to a problem (Wegerif, Mercer, & Major, 2019). However, many school experiences do not scaffold or support young learners’ collaborative practices (Mercer & Howe, 2012), and in a traditional learning setting, young learners may not perform productive collaboration practices such as asking higher-level questions and sharing ideas (Nystrand, Wu, Gamoran, Zeiser, & Long, 2003). Previous research emphasizes the importance of creating engaging opportunities for learners to practice these essential collaborative skills (Deiglmayr & Spada, 2010; Mercer, Wegerif, & Dawes, 1999; Alaimi, Law, Pantasdo, Oudeyer, & Sauzeon, 2020; Fagerlund, Vesisenaho, & Häkkinen, 2022).

Virtual agents may be one way of scaffolding children’s collaborative skills. They provide engaging and fun learning experiences for children (Ceha & Law, 2022; Paranjape, Ge, Bai, Hammer, & Cassell, 2018) that promote learning and social skills in a variety of domains for children such as algebra education (Lopes et al., 2019), AI education (Yu et al., 2019), language learning (Kim, 2019; Joaquim, Bitten-court, de Amorim Silva, Espinheira, & Reis, 2022; Baker et al., 2021)

and storytelling (Pires, Alves-Oliveira, Arriaga, & Martinho, 2017). In these learning environments, virtual agents hold many different roles, sometimes as authority characters (tutors or mentors) (Graesser, Forsyth, & Lehman, 2017; Han et al., 2021), but a particular type of virtual agent, referred to as *virtual learning companions*, are designed as near-peers to the learners and have been shown to provide particularly relatable social support (Chou, Chan, & Lin, 2003; Chen, Liao, Chien, & Chan, 2011; Kumar & Rosé, 2014; Schroeder & Gotch, 2015). While there is clear potential for virtual learning companions to model productive collaborative talk for children, whether these agents can achieve that potential has been an open question, in large part because research with older learners demonstrated that sometimes the agents’ behavior would “backfire”, resulting in the opposite of the desired behaviors or abuse toward the agents (Garner, Brown, Sanders, & Menke, 1992; Schroeder & Gotch, 2015).

This paper reports on the results of a multi-year design, development, and iterative refinement effort in which virtual learning companions were integrated into a block-based coding environment for upper elementary school learners (4th and 5th grades, typically ages 9–11). These virtual learning companions model productive collaborative talk, specifically *exploratory talk*, in which learners critically and

* Corresponding author.

E-mail address: tearlerandell@ufl.edu (T.V. Earle-Randell).

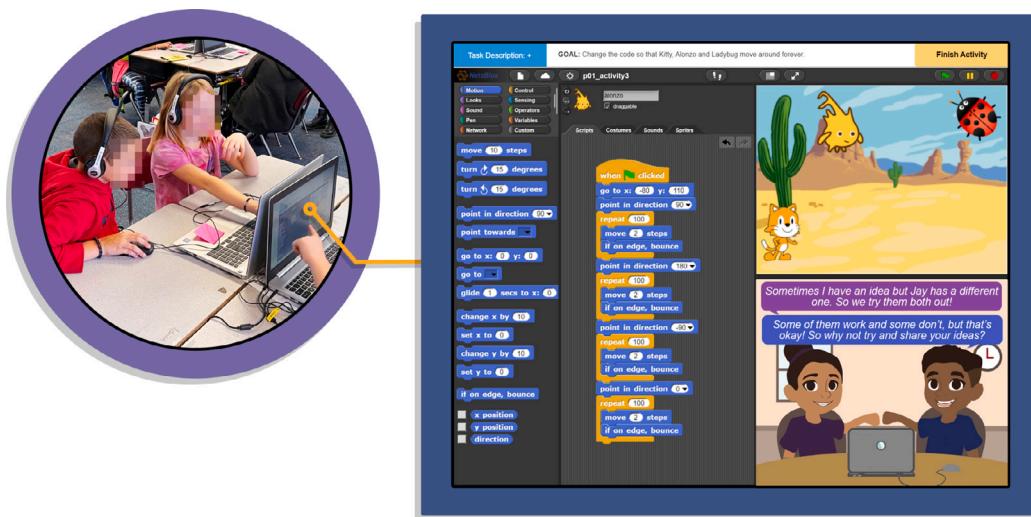


Fig. 1. Virtual learning companions integrated within a block-based coding environment, where the agents model productive collaborative talk to dyads of human learners.

constructively engage with each other's ideas to form a shared understanding (Mercer, 2002). The agents model this behavior by presenting their own experience working on the same task and explaining they solved the problem by sharing their thoughts, listening to their partner, and constructively challenging their partner's ideas. We examine the extent to which these virtual learning companions influence learners' productive collaborative talk, and in particular, we investigate the following research questions:

RQ1: In what ways do virtual learning companions who model exploratory talk influence the collaborative dialogue of upper elementary learners as they collaboratively code?

RQ2: What is the impact of virtual learning companions that model exploratory talk on learners' knowledge gain and attitudes toward computer science?

RQ3: What design implications emerge from learners' feedback on the learning companions?

To investigate these questions we conducted a two-condition study over six 45-minute classroom periods in elementary school classrooms in which children engaged in pair programming within a block-based programming environment, as shown in Fig. 1. Pair programming, which will be explained more extensively in Section 2.3, consists of two students working together on one computer to complete a coding task. The treatment condition included near-peer learning companions who modeled productive collaborative behavior, while the control condition did not include learning companions. Results of dialogue analysis indicate that children in the treatment condition asked each other more higher-order questions, an important component of exploratory talk. Furthermore, analysis of the post-surveys indicated that the virtual learning companions fostered more positive attitudes toward computer science compared to the control condition. This paper details those findings and examines excerpts from children's dialogue that illustrate the impact of the virtual learning companions on children's interactions. Moreover, we analyze the children's posthoc open-ended feedback on the agents and discuss design implications arising from that feedback when taken in context with the study's findings.

This paper makes the following novel contributions. Foremost, we report on the development of the FLECKS virtual learning companions that were designed and iteratively refined over the course of multiple classroom studies that included participatory design and children-in-the-loop feedback. Then, we present a deep analysis of children's dialogues during collaborative coding, using a novel dialogue annotation scheme derived from collaborative learning theory. Finally, this work provides novel results establishing virtual learning companions' potential for scaffolding children's collaborative practices and provides directions for much-needed future investigation on virtual learning companions' interactions with young learners.

2. Background and related work

This section begins by presenting the exploratory talk framework, which has been long studied in collaborative learning research and forms the theoretical foundation for the learning companions' collaborative talk. We then move on to describe related work on virtual learning companions for elementary school children, and finally, we describe the collaborative framework of pair programming, which we utilized in the studies reported in Sections 3 and 4.

2.1. Theoretical foundation: Exploratory talk

This work builds upon the theoretical foundation of exploratory talk, which is focused on the sharing of learners' partial understandings, and on thoughtful listening to the thoughts of their partner (Mercer, 2002). Communicating in this way requires a certain level of rapport, as each learner is opening their incomplete understanding to criticism from their partner (Barnes, 2008). Productive criticism can come in the form of higher-order questions; in fact, a hallmark of productive exploratory talk is learners engaging in higher-order questioning (T'Sas, 2018). Exploratory talk can expand the joint Zone of Proximal Development (Vygotsky & Cole, 1978) of a dyad by enabling partners to achieve a better mutual understanding of the problem (Fernández, Wegerif, Mercer, & Rojas-Drummond, 2015), in part through questioning and substantive responses, because through effective communication strategies and exploratory talk, learners can mutually support each other's progress and scaffold each others learning (Sylvia Rojas-Drummond, 2003). Dialogue processes within exploratory talk have been shown to support learning, such as the process of justification, which has proved beneficial for both children's learning outcomes (Wegerif & Dawes, 1998) and improving problem-solving capacity (Rojas-Drummond & Zapata, 2004). Exploratory talk needs to be explicitly modeled, as most children do not naturally engage in productive collaborative dialogue (Boyd & Kong, 2017), and it has been shown that collaborative dialogue is more productive than children learning individually (Tenenbaum, Winstone, Leman, & Avery, 2020).

2.2. Virtual learning companions for elementary school learners

A virtual learning companion, first introduced as a new class of intelligent tutoring system in 1988 (Chan & Baskin, 1988), is a system in which a computer acts not only as an educational agent but also plays the role of companion, grounded in the theory that individual cognitive development is influenced by social interactions or in collaboration

with peers (Vygotsky, 1978). Over decades of development in this line of research, learning companions have shown significant potential for supporting learners, demonstrating impacts such as enhancing learners' engagement and motivation (Hsu, Chou, Chen, Wang, & Chan, 2007) and promoting learners' empathy (Kim, Baylor, & Shen, 2007).

Virtual learning companions designed for children have demonstrated that they can facilitate learning. Ryokai, Vaucelle, and Cassell (2003) developed a virtual learning companion to help young children to learn how to read and write. The virtual companion played the role of a learning partner in the context of telling stories with children. The virtual learning companion would tell her own stories and use non-verbal gestures to engage the children (e.g., occasionally looking up at the viewer to interest the children in the story). The results with 31 early elementary school-aged girls demonstrated that the virtual companion could effectively improve children's literacy learning skills, as children told stories that more closely resembled the virtual companion's linguistically advanced stories. In another study, Pires et al. (2017) developed a virtual learning companion for developing young children's creativity. The virtual learning companion simulated idea generation during a storytelling activity. The results with 20 children (ages 7–9) showed that the system contributed to the generation of more ideas by children when they performed learning tasks collaboratively with the agent; in addition, the social component of the system helped children to better express and represent their knowledge and opinions.

Although these and other learning companions have shown promise in developing young learners' literacy and social skills (Han et al., 2021; Kim, 2019; Joaquim et al., 2022; Baker et al., 2021), there has been very little work on developing virtual learning companions for helping young learners develop collaborative practices, such as interpersonal communication skills, knowledge sharing strategies, and active participation (Hesse, Care, Buder, Sassenberg, & Griffin, 2015). Our study aims to fill this gap by developing virtual learning companions to promote young learners' collaborative practices by modeling productive collaborative talk.

2.3. Collaborative framework: Pair programming

In the present work, we conduct elementary school studies in the context of pair programming, a popular framework for collaborative learning in computer science (Williams, Wiebe, Yang, Ferzli, & Miller, 2002). In pair programming, two learners work on the programming task with two different roles: the *navigator* guides the coding process and provides feedback while the *driver* implements the solutions. During pair programming, learners switch roles regularly. Pair programming with elementary school children has been demonstrated to have positive effects on problem-solving skills, computer science knowledge, and children's interest in learning about technology (Calder, 2010; Lai & Yang, 2011). There is a growing body of research on pair programming with elementary school learners (Zhong, Wang, & Chen, 2016), but much work remains in order to understand the collaborative behaviors that emerge during pair programming, and how these collaborative experiences could be enhanced. Research has also suggested the need to provide elementary students with guidance and support during pair programming, as issues of conflict (Tsan et al., 2021) and equity (Lewis, 2011; Shah, Lewis, & Caires, 2014) may arise during their collaborative coding tasks. The novel use of virtual learning companions to support learners in pair programming (i.e., two agents supporting two children) could enhance collaborative outcomes and make computer science learning more accessible.

3. FLECKS: Collaborative CS learning with intelligent virtual companions

To build a platform to investigate virtual learning companions to support collaborative coding, we began with the Netsblox programming environment,¹ which itself was originally built on the Snap! block-based coding framework.² During the initial development of the FLECKS system, we re-designed the layout of Netsblox and repositioned components (e.g., sprite buttons) to create enough space on the screen to integrate the virtual agents into the environment (Fig. 2). FLECKS allows users to share projects with other users and to work on the same code synchronously.

The virtual learning companions are the central novel feature in the FLECKS learning environment. FLECKS features two virtual learning companions, Viviana and Jeremy, who model exploratory talk by talking to each other. The virtual learning companions are shown modeling collaborative talk in the form of vignettes: brief dialogues between the agents, sometimes directly addressing the dyad of learners. These vignettes model intended learner behavior using narrative devices, dialogue, and body language. The agents switch between static poses to add interest and emphasis to their dialogue, open their mouths to speak, and look at children or at each other, depending on who they are addressing. Their spoken dialogue was professionally recorded by voice actors, and text captions are provided in color-coded speech bubbles. The dialogue was collaboratively designed with elementary school children, as will be discussed later in this section. The FLECKS environment includes 46 different agent vignettes revolving around Viviana and Jeremy learning about the importance of asking "why" (higher-order) questions, sharing their ideas with each other, and listening to each other's suggestions (Fig. 3). In the studies reported here, the timing of agent vignettes was controlled through a Wizard of Oz approach, which is detailed in Section 4.1.

3.1. Iterative design process

To design the agents' appearance, dialogue, and behavior, we worked with multiple groups of children using an iterative process. This section briefly reviews the series of interview and participatory design studies we conducted over a three-year period (Fig. 4).

3.1.1. Design phase 1: Co-designing graphical representations of agents with children

We conducted three iterative co-design studies with 4th and 5th-grade learners to develop and refine the initial visual design of the virtual agents (Fig. 5). We have previously reported on Design Phase 1, which began with children's open sketches and moved toward a structured design (Wiggins, Wilkinson, Baigorria, Huang, Boyer, Lynch, & Wiebe, 2019). Design Phase 1 took place over the course of Spring 2019.

User study 1. In the first study, we asked 22 children ages 9–11 to draw what "good collaboration" looks like on a blank piece of paper (Fig. 5a), and then invited them to label the drawings and explain their reasoning (Wiggins et al., 2019). We performed a thematic analysis of these artifacts and used our findings to create a drawing scaffold (Fig. 5b), similar to a coloring book page, that had two agents at a circular table.

User study 2. We used this drawing scaffold in a new study with a new set of elementary learners (38 total, from two classrooms of students in grades 4 and 5). These learners were asked to draw their own collaboration "buddies" on the drawing scaffold (Fig. 5c). Researchers used thematic analysis to identify themes around agent appearance, agent dialogue, and agent surroundings. We then applied these insights to design the first digital versions of the virtual learning companions (e.g., Fig. 5d).

¹ <https://netsblox.org/>

² <https://snap.berkeley.edu/>

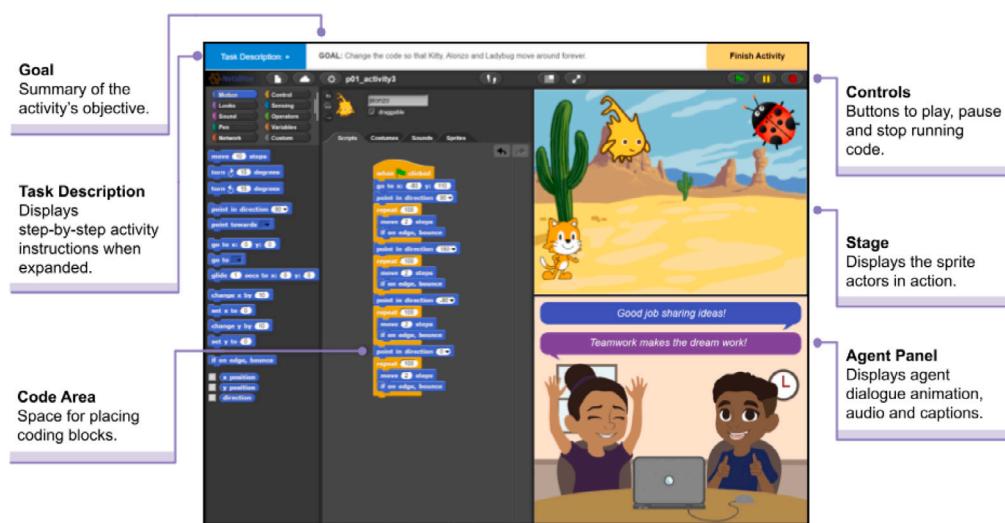


Fig. 2. The FLECKS learning environment, which embeds virtual learning companions (lower right) in the block-based coding environment.

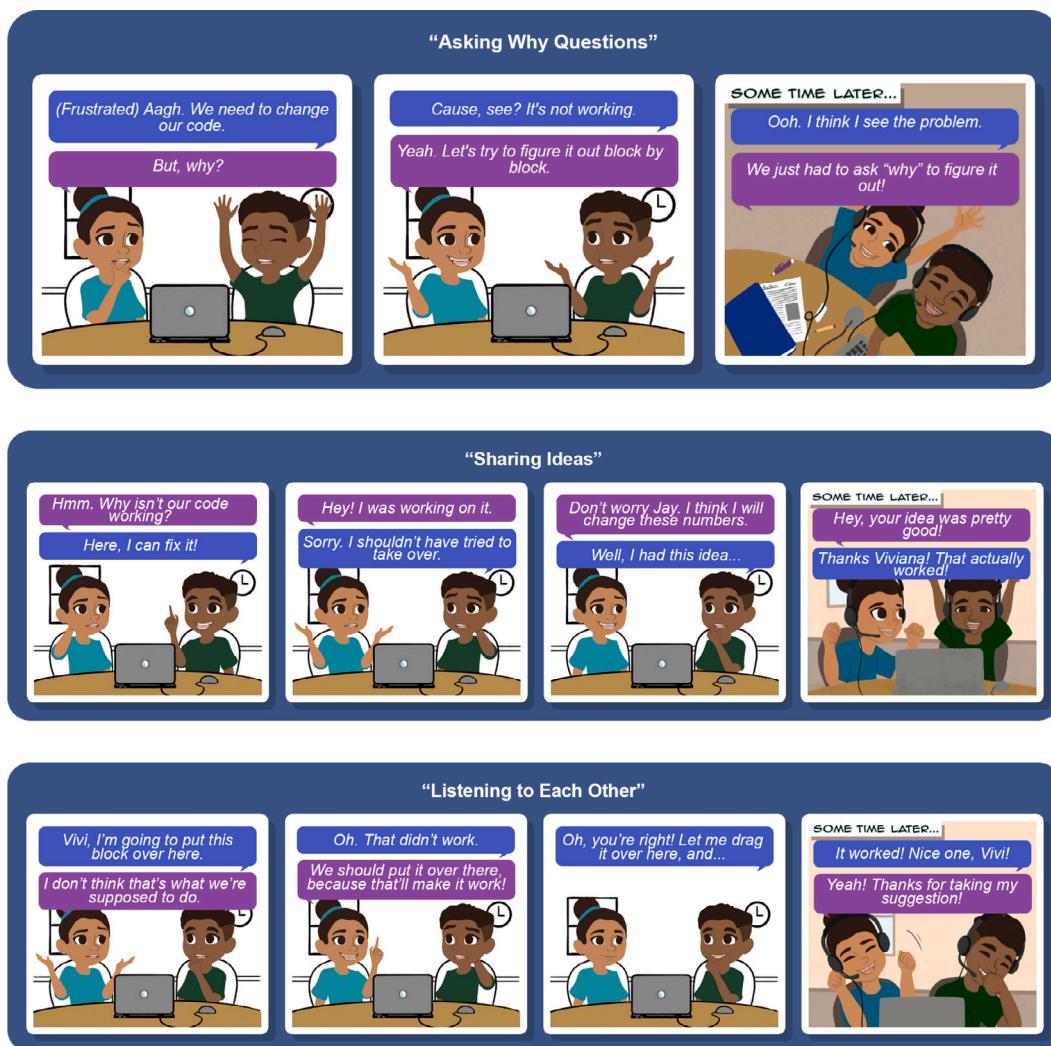


Fig. 3. Virtual learning companion scenes that show Viviana and Jeremy learning the importance of different collaborative practices.

User study 3. The final user study returned to the same learners from User Study 2, inviting them to evaluate the agents we had designed based on their drawings. Three pairs of agents were presented, and

learners evaluated them along three dimensions: relatability, appeal, and believability. We asked learners questions, such as “Can you imagine being friends with these kids?”, “Would you want to talk with

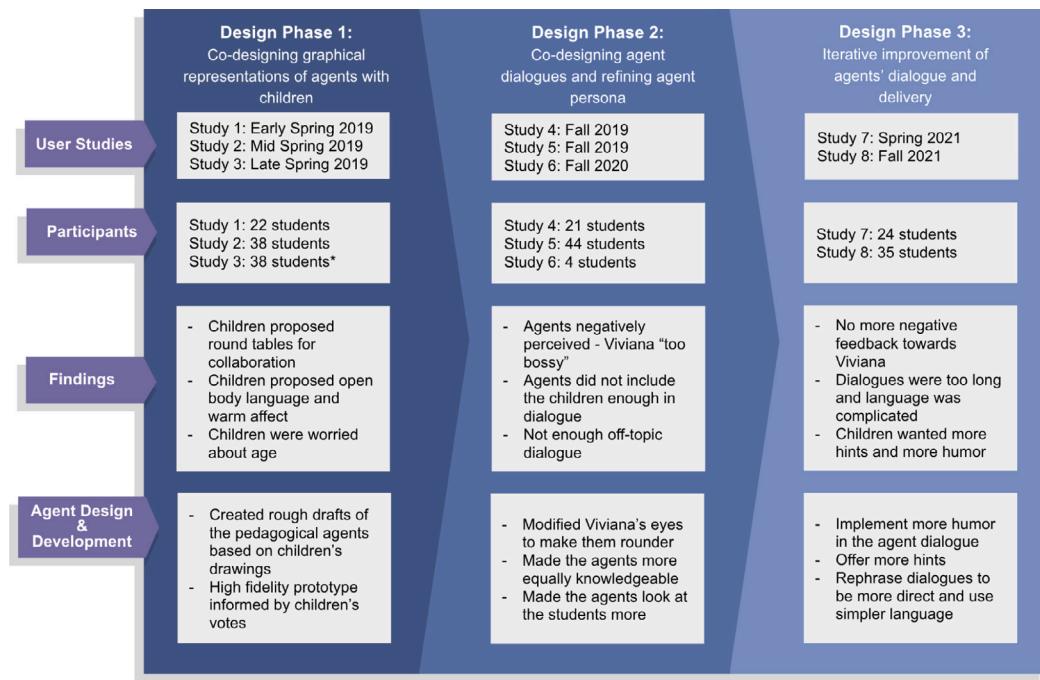


Fig. 4. Phases of virtual learning companion co-design.

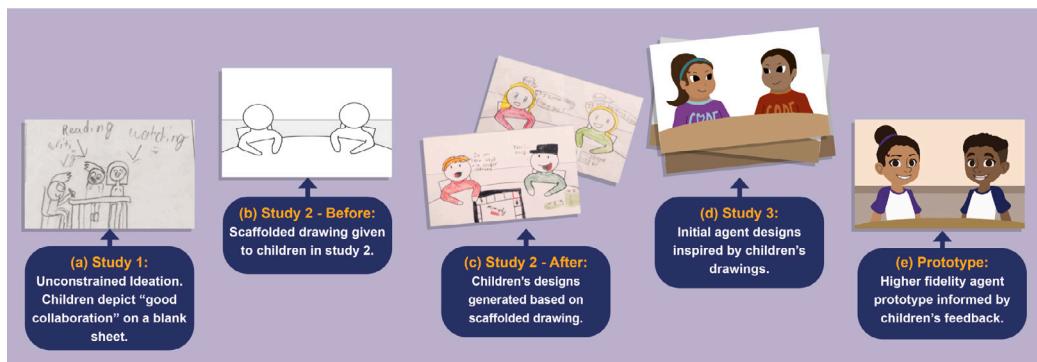


Fig. 5. Design Phase 1: Iterative design process of agents' graphical representation.

these kids?", and "Do these kids look like they could be in your class?" Results from this study were synthesized to develop one higher-fidelity pair of virtual agent designs, shown in Fig. 5e.

3.1.2. Design phase 2: Co-designing agent dialogues and refining agent personas

In this phase, we worked with children to refine the agents' movement and co-author the agents' dialogue. To do this, we conducted two classroom studies in Fall 2019 followed by a remote study in Fall 2020. These design activities have not been reported elsewhere.

User study 4. In this study, we presented 21 4th and 5th-grade learners with preliminary designs of the virtual agents' animations (Fig. 6) and asked them what they thought each animation was trying to convey. We used this feedback to change or repurpose the animations where necessary so that they could be accurately interpreted by learners. For example, the "hand behind neck" animation (Fig. 6, Left, #11) was originally intended as a joyful animation, but children interpreted it as expressing embarrassment, so we repurposed the animation accordingly.

User study 5. In this focus group study, we presented a total of 44 learners (all in fourth or fifth grade) with a set of animated vignettes in which the prototype learning companion agents aimed to model collaborative talk behaviors. A prominent theme was that learners found the female agent (Viviana) less approachable compared to the male agent (Jeremy). To address this, we softened Viviana's facial expression (Fig. 7) and more equally distributed the advice-giving between the two characters' dialogue. Feedback also guided a shift away from a diegetic "flashback" format for the vignettes; instead, the agents' interactions were set in the present, and their gaze was directed at the human viewers.

User study 6. Due to the COVID-19 pandemic, this pilot study was conducted remotely with just two dyads of 5th-grade learners. Children were taught the pair programming approach (see 2.3) and used this practice to complete a coding activity in the FLECKS learning environment, now featuring the revised virtual agents as shown in Fig. 2. The learners were shown three agent vignettes while they worked: one in which the agents introduce themselves, and two modeling a kind of collaborative talk. Overall, learners reported that they liked the agents and would want to work with them again. While we acknowledged the limitations of the size of this study, we still used observations from this

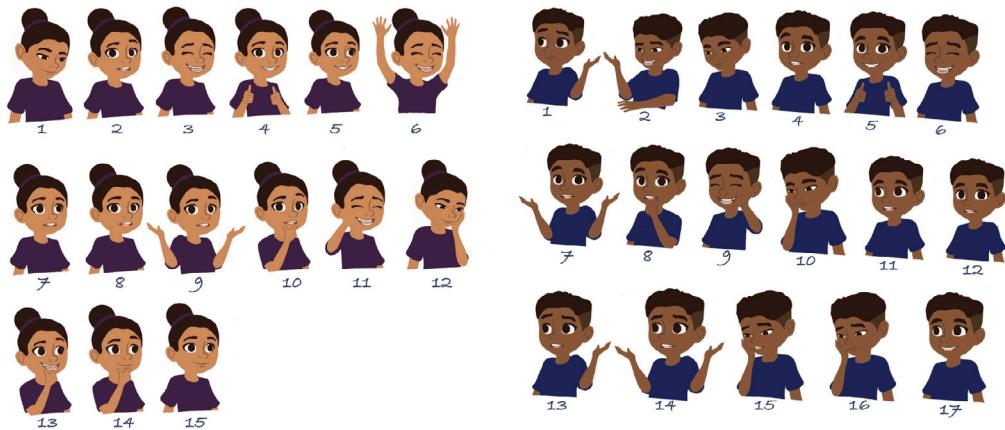


Fig. 6. Keyframes of virtual learning companion gestures and expressions evaluated by the learners in User Study 4.



Fig. 7. From left to right, iterative refinement of Viviana's eyes, making her brows rounder and her pupils larger in response to children's feedback.

study to guide refinements to the user interface behavior surrounding the agents, to develop two new learning activities for the environment, and to create additional agent dialogues around exploratory talk, including the ‘switch roles’ mechanic that appears on the screen after 20 min. Caution should be used in interpreting design outcomes with this small study.

3.1.3. Design phase 3: Refining agent personas and expanding agent dialogues

Phases 1 and 2 led to the design of two agents with distinct personas: Viviana, an energetic risk-taker who loves to try new things, and Jeremy, a careful planner who loves solving puzzles. In the third design phase, over the course of 2021, we expanded the agents' dialogue repertoire to model exploratory talk in a variety of ways.

User study 7. We involved a total of 24 fourth and fifth-grade learners in collaboratively revising the agents' dialogue in this six-session study. First, students used the FLECKS learning environment with embedded virtual agents, featuring the changes and refinements made after Study 6. Then, we led a class-wide discussion eliciting feedback on the agents, and learners were asked to read the agents' dialogue aloud, taking the role of Jeremy or Viviana. Learners were also asked for feedback on each dialogue using an online survey, including an item that asked them to rewrite the lines themselves, with the goal of making the dialogue sound more natural for their age group. Fig. 8 shows an example of the revisions we made to the agents' dialogue following the study. Some learners suggested including more humor in the dialogues, and others requested functionality beyond what we intended to build with these learning companions, such as providing help with coding if learners found themselves stuck. Negative feedback about Viviana, which had been prevalent in previous studies, was not observed in these results. The analysis also suggested that we had successfully positioned Jeremy and Viviana as knowledgeable coders.

User study 8. In preparation for our final evaluation study, we piloted a **Wizard-of-Oz** study design (Riek, 2012), where trained researchers (wizards) remotely observed the children's collaboration, selected appropriate vignettes, and delivered them at appropriate moments. Wizards were shown a live video feed of the learners themselves from the learners' built-in laptop camera and a live stream of their work on the laptop's screen. This was piloted in two 5th-grade classrooms. We also piloted a control condition (with identical learning activities and environment, but no virtual agents) in another two classrooms. In Study 8, data showed that learners followed the wizards' suggestions approximately 30% of the time and that the likelihood of learners following the suggestions varied greatly depending on the context (e.g., they were less likely to mimic the collaborative practices presented if the students were arguing with each other when the message was sent) (Wiggins, Earle-Randell, Bounajim, Ma, Ruiz, Liu, Celepkolu, Israel, Wiebe, Lynch, & Boyer, 2022). Wizards observed that children seemed to imitate exploratory talk behaviors more often when dialogues included a direct “call to action”, so we introduced more of these in the final prototype. In keeping with the agents' positionality as near-peers, not authority figures, these calls to action were phrased as friendly suggestions, such as: “So why not try and share your ideas?”, and: “Don't be afraid to ask ‘why’ to your partner! It'll help you understand things!”.

3.2. Learning companions in final study

By the end of this three-year design process, we had created a pair of animated virtual learning companion agents with a total of 27 dialogue vignettes designed to support the learners' collaborative practices. To avoid repetition, multiple vignettes were written for many goals, and for some goals, such as “Hint”, unique vignettes had to be written for each learning task (Fig. 9).

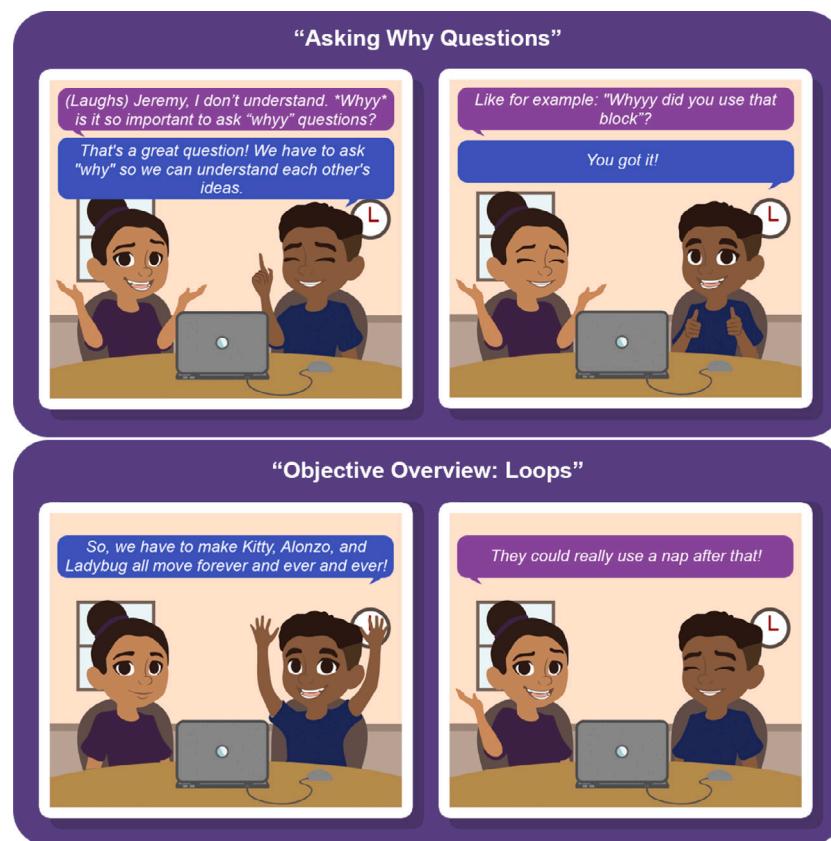


Fig. 8. Example of humorous agent dialogue vignettes.

4. Quasi-experimental study design and resulting data

After designing and refining the agents as described above in Studies 1 through 8, we moved on to conduct a quasi-experimental study to investigate the impacts of the virtual learning companions on the collaboration of elementary school learners. We designed this study to answer the following research questions:

RQ1: In what ways do virtual learning companions who model exploratory talk influence the collaborative dialogue of upper elementary learners as they collaboratively code?

RQ2: What is the impact of virtual learning companions that model exploratory talk on learners' knowledge gain and attitudes toward computer science?

RQ3: What design implications emerge from learners' feedback on the learning companions?

4.1. Wizard of Oz protocol

Using the interface shown in Fig. 10, six researchers acted as wizards (three of which are co-authors) by navigating a decision tree based on the behavior they observed the children make. The use of multiple wizards was necessary in order to simultaneously support multiple dyads. Every 3–6 min, they chose an appropriate vignette for the virtual learning companions to present. Wizards were trained on video data of elementary computer science learner dyads collected in prior studies and through mock learner dialogue, where they watched clips and collectively discussed what wizarded moves should be taken. Following this, they met to collectively agree upon the set of child behavioral and dialogue cues that would guide the selection of each “class” of vignette in order to reduce variability between the wizards, which was included on the interface in the “what to look for” section. (Wizards would select a dialogue goal, and FLECKS would automatically send one of the associated vignettes, as described in Section 3.2).

The team then developed a decision tree to guide the selection of a vignette class based on these behavioral and dialogue cues, and this tree formed the structure of the wizard interface. During the study, each wizard connected via Zoom to one pair of children in each of the two classes (morning and afternoon). As they supported the learners, wizards were in contact with one another and with onsite facilitators via instant messaging. After each session, wizards were prompted to answer discussion questions about their experience: (1) *How did your session go?* (2) *What are some things you wish you could have said but could not?* (3) *Were there any undesired collaborative behaviors that you wished you could have commented on?*

4.2. Participants

The study was conducted with 4th-grade learners in three different classrooms in a rural elementary school in the Spring of 2022 in the southeastern United States. The school's student body was approximately 72% White/Caucasian, 15% Hispanic/Latinx, 9% Black/African American, 4% multiracial, and 1% other. The school served a large percentage of economically disadvantaged learners, with 74% of the student body eligible for free or reduced meals.

Before beginning the study, we obtained Institutional Review Board (IRB) approval to implement the study with children, and parents were given consent forms with details of the study procedures. We also verbally explained the study's purpose to the children and asked for verbal assent before collecting their data for research purposes. All the children (consenting or non-consenting) participated in the same learning activities as a part of their class activities, but we collected data only from the children with written parental consent and learner assent. Out of the 80 children in three different classes, 59 learners provided assent and parental consent. Of those 59, 33 were female, 19 were male, and 7 preferred not to report their gender. The mean age

Goals to Model	Amount of Vignettes	Example Dialogue
Confusion about objective	3	Jeremy: "So, we have to make Kitty, Alonzo, and Ladybug all move forever and ever and ever!" Viviana: "They could really use a nap after that!"
Not looking at instructions	1	Jeremy: "Whenever I get stuck, I look at the instructions again." Viviana: "Yeah, the instructions are in "Task Description". Let's try clicking that!"
Blocks are causing problem	1	Viviana: "Hmm. It might be useful to figure out which block's causing the problem." Jeremy: "You got this!"
Number error inside block	1	Viviana: "Let's see. Maybe the numbers are causing the problem?" Jeremy: "You might wanna keep changing them!"
Hints	3	Jeremy: "When we did this we forgot to put blocks inside the "Forever" block." Viviana: "Yeah, and then our sprites kinda *never* moved..." Jeremy: (Embarrassed) "Yeeeah. So you should try putting blocks inside the "Forever block"!"
Inactive Coding	3	(Agents encourage children to read the "Goal", run the code, or review the instructions)
Inactive Talking	3	Jeremy: "Hey, maybe we should talk about our ideas! We could come up with an answer together."
Distraction	3	Viviana: "Hey friends! Are you still coding? Where were we?"
Asking "Why" Questions	3	Jeremy: "Nice question! I would've never thought of that!"
Sharing Ideas	3	Viviana: "Good job sharing ideas! Teamwork makes the dream work!"
Listening to Each Other	3	Jeremy: "Great job listening to your partner's ideas! Go team!"

Fig. 9. List of goals that shows the number of dialogue vignettes per goal, and examples of the agent dialogues.

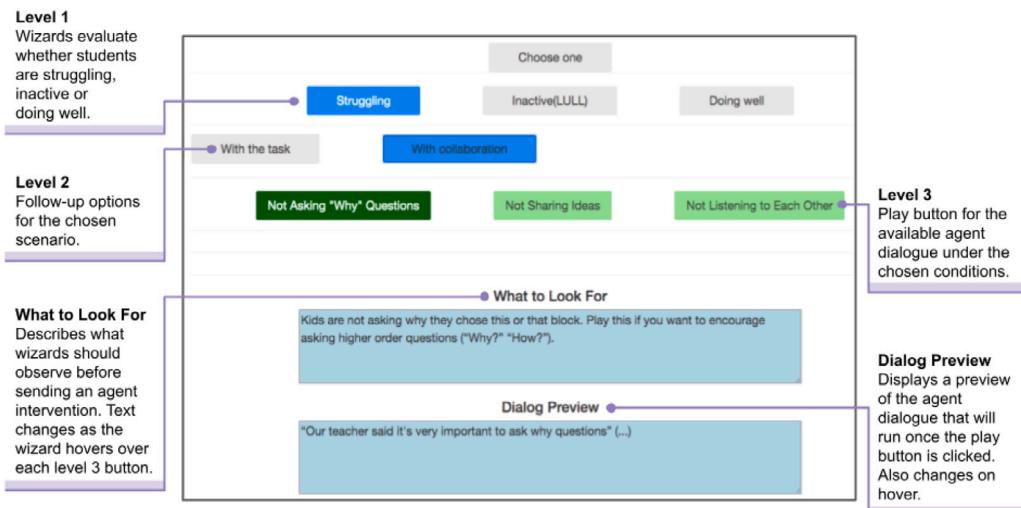


Fig. 10. Interface used by wizards to adaptively select agent vignettes.

was 9.29, with ages ranging from 8 to 12. In response to the question, "Have you ever done any computer science, programming, or coding?", 18.6% of learners reported prior coding experience (e.g., code.org, Scratch) before the first day of the study.

4.3. Elementary classroom study procedure

The quasi-experimental study was designed with a treatment condition (with virtual learning companions) with 20 dyads and a control condition (identical coding interface without virtual learning companions) with 10 dyads. In the control condition, children worked in pairs to program after being instructed on the pair programming paradigm, switching roles between driver and navigator after 20 min. In the treatment condition, dyads of children were given the same instruction as the control condition, and they were joined by the virtual learning companions. We implemented the study over six 45-minute class periods:

- Day 1 consisted of a pre-data collection with survey items covering demographics, CS attitudes (where learners stated their level

of agreement with statements like "I would like to use coding to make something new" and "I believe I can be successful in coding") (Vandenbergh, Rachmatullah, Lynch, Boyer, & Wiebe, 2021), and questions from the CS Concepts Assessment (Vandenbergh, Rachmatullah, Lynch, Boyer, & Wiebe, 2021) followed by an introduction to block-based programming using Blockly games.³

- On Day 2, children were introduced to the FLECKS block-based programming interface and learned fundamental block-based programming and debugging skills.
- On Day 3, learners were paired randomly and would work in these pairs for Days 3, 4, and 5. Non-consenting students were paired with one another and no data was collected from these pairs; a subset of learners with special needs were paired in advance by the teacher, still matched by consent status. Pairs of learners used the pair programming paradigm to work on the first coding task (Fig. 11). Coding tasks were the same in

³ <https://blockly.games/>



Fig. 11. Elementary learners working together using the pair programming paradigm during the classroom study.

both conditions across all days of the study. On this day the task drew on coding concepts such as program flow and function parameters. In the treatment condition, Jeremy and Viviana appeared and introduced themselves for the first time on this day. Next, a vignette about the benefits of **asking higher-level questions, such as “why” questions**, was played automatically at the start of the activity. Then, students began working on the coding task, while the wizards, based on their observations of students' behavior, periodically chose dialogues for the learning companions to deliver. These dialogues were iteratively refined with learners, as detailed in Section 3.1. Wizards provided this support on Days 3, 4, and 5.

- The **Day 4** coding activity employed coding concepts such as program flow, function parameters, and loops. The treatment condition's session began with a learning companion vignette modeling the importance of **sharing ideas with each other** and was interspersed with wizard-selected dialogues.
- The **Day 5** activity introduced the coding concept of conditionals and continued to draw on program flow and loops. The treatment condition's session began with a learning companion vignette modeling the importance of **listening to one another** and was interspersed with wizard-selected dialogues.
- On **Day 6**, learners completed identical CS Attitudes and CS Concepts Assessment survey items from Day 1, as well as a post-survey further addressing computer science attitudes and their learning experience.

4.4. Dataset of children's collaborative dialogue

In this paper, we analyze data from Day 4 of the six-day study. Day 4 was the second day of project work, and in the treatment condition, Day 4 was the second of the three days (Days 3–5) when the virtual learning companions were present. We selected this day because the largest number of dyads were present and had video recordings available on Day 4 (see Table 1, and the learners had also settled into their dyads and were familiar with the agents by this point.

Table 1

Consenting dyads present per day with audio and video recordings.

Condition	Dyads on Day 3	Dyads on Day 4	Dyads on Day 5
Treatment	9	10	9
Control	4	5	4

Out of the 25 consenting dyads, 15 were present on Day 4 and had video and audio recordings of their collaboration. The video and audio recordings were manually transcribed and prepared for tagging. The wizards delivered 105 total agent vignettes on Day 4, with a mean of 5.83 ($sd = 4.22$) vignettes per dyad. On average, each dyad had 269 turns of collaborative talk with each other.

4.5. Dialogue act tagging

After collecting the corpus of data, we conducted a *dialogue act analysis*, which emerged from the field of sociolinguistics (Austin, 1975) and has been deeply investigated by dialogue researchers and computational linguists (Core & Allen, 1997; Stolcke et al., 2000; Raheja & Tetreault, 2019). In dialogue act analysis, researchers label a dialogue corpus with a dialogue act taxonomy, typically one constructed for the specific corpus and the research goals based on theory and prior relevant taxonomies. This methodology is used to label dialogue utterances with the intent or pragmatic function underlying them.

Our dialogue act taxonomy draws upon the exploratory talk framework (Mercer, 2002) and upon a closely related dialogue act taxonomy by Zakaria et al. (2022) that was designed for elementary school learners' classroom dialogues. We modified Zakaria et al.'s dialogue act scheme to isolate the exploratory talk moves. This meant collapsing some of the dialogue act tags (e.g., combining "Self-Explanation" with "Justification", and "Suggestion" with "Alternative Idea"). It also required adding a tag to capture utterances directed at the agents, separating them from utterances meant for the human partner. The final dialogue act scheme is shown in Table 2.

To establish reliability of the dialogue act labeling, two annotators first engaged in a training phase where they collaboratively applied the dialogue act taxonomy and discussed any disagreements. Once training was complete, they independently tagged an overlapping 20% of the data, reaching a Cohen's kappa score of 0.816, a strong agreement (Landis & Koch, 1977). They then proceeded to divide and tag the remaining data independently. This effort resulted in 4039 total tagged dialogue acts.

5. Results

In this section we examine differences that appeared across conditions in learner dialogue, then differences in learning outcomes (i.e., learning gain and computer science attitude improvements), and finally we present a thematic analysis of learner feedback on the virtual learning companions.

5.1. Virtual learning companions' influence on learner exploratory talk

RQ1 asks, *In what ways do virtual learning companions who model exploratory talk influence the collaborative dialogue of upper elementary learners as they collaboratively code?* We investigate this question through *dialogue act analysis*, which has been successfully used to encode a diquite range of dialogues related to learning, with very recent work capturing learners' metacognitive moves (Bosch et al., 2021), and how learners exchange ideas with peers as they co-creatively write code for computational music remixing (Katuka et al., 2022). We use dialogue act analysis to annotate exploratory talk moves that occurred while learners were pair programming.

Once the dialogue acts were annotated, we compared the counts of each dialogue act between the treatment and control condition

Table 2
Dialogue act schema.

Tag	Description	Example	Combined %	Treatment %	Control %
Question - Higher-order	Asks a why question or a question that challenges a partner's idea.	"Why is he moving like that?" "What happens if you keep it that way?"	1.2%	1.5%	.004%
Question - Other	Asks anything other than a why question	"What does that block do?"	11.3%	10.7%	12.4%
Agreement / Acknowledgement	Agreement on any opinion/edit	"Looks good" "Good job" "ok"	11.1%	10.6%	12.1%
Disagreement / Negative Feedback	Disagreement on any opinion/edit	"No no no" "That is wrong"	6.5%	6.5%	6.4%
Self-Explanation / Justification	Explain the steps they are taking, or thoughts	"Now we can put this here to make it happen twice" "I did that because of how slow it is"	12.6%	13.7%	10.3%
Disagreement with Justification	Disagrees but provides reasoning	"No, that won't work because it needs to go in a square."	0.38%	0.3%	0.5%
Directive	Telling partner to do something	"Give me the keyboard" "Click that one"	8.0%	7.7%	8.8%
Suggestion /Alternative Idea	Any suggestions when directly talking to partner. (Leaving wiggle room)	"Maybe we should put two of those" "How about doubling that?"	10.7%	11.3%	9.5%
Seeking Help	Learner directly or indirectly seeking help from partner	"I'm confused" "IDK"	3.3%	3.2%	3.5%
Antagonistic Action	Actions or interactions that cause tension including harmful comments, instigating fights, prodding, putting down partner contributions, and showing annoyance with partner	"You are being ridiculous" "Stupid" "You don't know anything"	1.6%	1.7%	1.3%
Social	Social dialogue	"Did you hear about James?" "I love your pencil" [Quoting meme] "Thank you"	13.3%	13.1%	13.5%
Directed at Agents	It was said to the agent, not the partner	"Thanks, we know we are great"	1.0%	1.5%	0%
Other	Something not covered by any of the other tags	[Reading instructions] "Turn your volume up.", "Are my headphones working?" [Discussion about headphones, sound, volume] "Wow, look at it go.", "Well, that was quick." [Observations and comments on the activity]	19.2%	18.1%	21.3%

Table 3
Differences in the treatment and control exploratory talk moves.

	Treatment (mean, sd)	Control (mean, sd)	p-value	Benjamini- Hochberg critical value
Question - Higher-order	4.9 (2.33)	1.4 (0.89)	0.0114*	Rank 1: (1/4)*.05 = .0125
Suggestion	37.0 (14.28)	30.8 (10.03)	0.2957	Rank 2: (2/4)*.05 = .025
Self- Explanation/ Justification	43.6 (14.63)	33.8 (14.72)	0.3272	Rank 3: (3/4)*.05 = .0375
Total Dialogue	335.6 (123.98)	332.4 (95.97)	0.8065	Rank 4: (4/4)*.05 = .05

using Wilcoxon ranked-sums tests. Table 3 summarizes the differences in counts of exploratory talk-related dialogue acts between the two conditions. Due to several hypotheses being tested, we applied the Benjamini-Hochberg correction to reduce the false discovery rate (Thissen, Steinberg, & Kuang, 2002). While the control condition asked more questions overall, most of these questions were categorized as other and were not relevant to the task the dyad was working on. This analysis was conducted to investigate the exploratory talk moves of the learners, and the results reveal one statistically significant difference: *children asked significantly more higher-order questions in the treatment condition than in the control condition*.

5.2. Learner knowledge gain and attitudes toward computer science

We next examined the outcomes of the sessions with RQ2: *What is the impact of virtual learning companions that model exploratory talk on learners' knowledge gain and attitudes toward computer science?* To test for incoming differences, we conducted the Wilcoxon ranked-sums test on the *pretest* (CS Concepts Assessment) and *presurvey* (CS Attitude) scores, which measured incoming knowledge and attitudes, respectively. There was no statistical difference in the *pretest* score (Treatment: 3.76 ($sd = 1.84$); Control: 4.92 ($sd = 2.18$); $p = 0.1222$), but children in the treatment condition had significantly higher *presurvey* scores (Treatment: 44.82 ($sd = 7.08$); Control: 41.31 ($sd = 4.03$); $p = \textbf{0.0296}^*$). Since this

Table 4

Pre/post changes in learners' CS knowledge and attitudes by experimental condition.

	Learning companions		No learning companions		Between- Subjects Repeated Measures p-Value
	Pre	Post	Pre	Post	
Knowledge Assessment	3.76 (1.84)	5.00 (1.75)	4.92 (2.18)	4.58 (1.73)	0.676
Attitude Survey	44.82 (7.08)	47.17 (6.05)	41.31 (4.03)	41.17 (6.91)	0.042*

finding does indicate a difference between the treatment and control groups that were not accounted for in the two-condition study design, we control for it using a repeated-measures analysis, specifically the SPSS toolkit's repeated-measures, between-subjects design. This uses the pre-survey scores to account for the extent to which any changes observed might be attributable to pre-existing differences between groups. The analysis thereby isolates the extent to which observable change can be attributed to the independent variable of the condition (treatment vs. control). **Table 4** displays the significant effect that the learning companions had on CS attitude change.

5.3. Analyzing feedback on virtual learning companion design

Finally, we investigated RQ3: *What design implications emerge from learners' feedback on the learning companions?* To address this question, two researchers collaboratively coded learners' open-ended responses to survey items after their interaction with FLECKS.

We analyzed learner responses to three open-ended questions administered at the end of the study: (1) "How much did Jeremy and Viviana change your interaction with your partner? In what ways?"; (2) "What do you expect them to do? How can they be more useful?"; and (3) "Would you like to have Jeremy and Viviana for your future group work? Why?" Two researchers collaboratively examined learners' responses and performed thematic coding (Kiger & Varpio, 2020). Five major themes emerged from the feedback:

- *Helpful/Friendly.* The most consistent remark from learners was that the agents were helpful to them during their collaborative learning and generally friendly. 19 out of 32 learners' feedback fell into this category with comments such as "They are helpful", "They are really thoughtful", "They help some and there cool.", or "I want to be a scientist and they would help me fix some problems".
- *Creepy/Annoying.* Not all students saw the agents as an aid to their experiences, with 5 out of 32 learners commenting such things as "...gives me so much anxiety" or "they are a little creepy". During the interaction with the agents, these learners were also likely to be vocal about not wanting the agents to join in the conversation (e.g., sighing when the agents started talking).
- *Desire for More Hints.* Many learners (20 out of 32) remarked that the agents should have provided more task-specific support in the form of hints: "[the virtual learning companions should] talk about code. Help with code". We deliberately included very little code-related scaffolding in this iteration of the agents because of the focus on modeling collaborative skills.
- *Supporting Exploratory Talk.* Learners also frequently reported that the agents promoted their team's use of exploratory talk. 11 out of 32 reflected, "They made me share my ideas" or "ask why questions".
- *Self-Expression.* Some learners (4 out of 32) felt that agents should have some changes to their design, writing "I want there hair and [their] clothes to be cuter". Suggestions for variations in hair and clothes have been common across all studies, such as changing them from one day of the study to the next or wearing seasonal clothes. Learners also suggested that agents should: "talk to me

when I'm bored". More off-task interactivity where the agents could talk about their lives outside of coding has been a prevalent suggestion from earlier studies.

6. Discussion and implications

This section discusses the differences in the child dyads' dialogue, the computer science attitude differences, areas for improvement when supporting collaboration, and finally the implications for the design of virtual learning companions.

6.1. Differences in learner's dialogue

Learner dyads in the treatment condition asked more higher-order questions than their counterparts in the control condition (**Table 3**). Higher-order questions can take several forms, such as the "How" question demonstrated in **Table 5**, a dialogue between P1 and P2, and the "Why" question in **Table 6**, a dialogue between P3 and P4. Both excerpts are taken from the treatment condition, in which one of the learning companions' first vignettes focuses on asking "Why" questions.

P1 and P2 (**Table 5**) asked each other one "Why" question on their first day of collaboration, then asked three on the second day (after seeing the learning companion vignette on this topic). **Table 5**'s Excerpt 1 demonstrates P1 and P2 utilizing higher-order questions to come to a better group understanding of the error that they were currently facing. P1 and P2 were both male and neither reported prior coding experience. As described in Section 2, all dyads were pair programming, a practice in which one learner is the Driver (i.e., the learner controlling the mouse and keyboard) and the other is the Navigator (i.e., the learner watching and commenting on the Driver's coding). Around three minutes before this excerpt, the dyad received a vignette encouraging them to use more exploratory talk. The two children were struggling with an error in their code, and P1, who was the Driver, was verbalizing what he did not understand about the way the code was working. P2, the Navigator, gestured toward the "Move" block, and P1 was able to explain the cause of the error.

Next, we examine P3 and P4 (**Table 6**), both girls, one with prior coding experience. As they were attempting to fix a bug in their code, they noticed that a sprite was not moving. P3 asked a higher-order question to P4, who was then able to make a suggestion that would eventually lead the pair to a viable solution. These children showed strong collaboration skills by using this question at the start of the session, so their wizard chose to deliver a praise vignette: "Nice work asking why questions to each other!"

Both of these examples show the potential of higher-order questions to help children arrive at a mutual understanding of a problem. The process of asking and attempting to answer higher-order questions has been shown to improve important skills in learners, such as critical thinking skills (Renaud & Murray, 2007; Barnett & Francis, 2012). In exploratory talk, these questions are thought to be a particularly important component because they engage the learner dyad in critical thinking while also highlighting areas of confusion or uncertainty (Mercer, 2002). The agents facilitate higher-order question-asking both by suggesting the use of higher-order questions when learners are reaching moments of uncertainty and by praising learners when they use higher-order questions.

Table 5

Excerpt 1, highlighting learners performing an exploratory talk move recently modeled by the virtual learning companions.

Speaker	Utterance	Dialogue Act
P1 - Driver	<3 min prior, the virtual learning companions model exploratory talk.> <P1 and P2 have recently switched driver and navigator roles.> Yeah. Okay, so that's in.	N/A N/A Agreement / Acknowledgement
P1 - Driver	That's how he wanted to do it. Here's your headphones.	Other
P1 - Driver	Okay. How are they still moving when you deleted it?	Question - Higher order
P2 - Navigator	<gestures toward code>	N/A
P1 - Driver	What?	Question - Other
P1 - Driver	Oh my gosh. I'm so- if and- move 10 steps.	Self-explanation / Justification

Table 6

Excerpt 2, highlighting virtual learning companions responding with praise when learners use exploratory talk moves.

Speaker	Utterance	Dialogue Act
P3 - Driver	<S1 is coding while talking>	N/A
P3 - Driver	So I feel like we should go to control and repeat ever. So ... motion, correct motion, and then move 10 steps.	Self-explanation / Justification
P3 - Driver	<As S1 is moving mouse towards run button> Let's see what Kitty does now. Kitty's [inaudible].	Self-explanation / Justification
P3 - Driver	<Executes code to see if problem was correct>	Question - Higher order
P3 - Driver	Why isn't Kitty moving?	Justification
P3 - Driver	<The virtual learning companions praise the higher order question>	N/A
P4 - Navigator	Weird. I'm on ... I changed it. That's weird. [inaudible] sprites, and do it to all of them.	Other Suggestion / Alternative Idea

6.2. Differences in learner computer science attitudes

The results also suggest that the virtual learning companions made a significant positive impact on self-reported computer science attitudes. Computer science attitudes can be a barrier to participating or persisting in CS for some children, who report low levels of self-efficacy and interest in computer science (Schulte & Knobelsdorff, 2007; Hinckle et al., 2020). Experiences that improve children's attitudes about computer science could promote their involvement with the field in the future. In FLECKS, the virtual learning companions model strong collaborative practices while reflecting on their own computer science challenges and eventual successes. Similar improvements in attitudes and engagement have been seen in studies with other virtual learning companions (Kim, 2005; Pezzullo et al., 2017).

6.3. Implications for design of virtual learning companions

Through the iterative refinement process and the results of the two-condition study, this work suggests implications for the design of virtual learning companions that can effectively support children in productive collaborative talk.

6.3.1. Near-peer agents can effectively model collaborative talk

Children in the learning companion condition asked significantly more higher-order questions, a component of effective collaborative talk, than those in the control condition (Table 3). It is important to note that while our learning companions did provide support to the learners, they did not always do so through direct interactions. While this differentiates our study from virtual agent studies in other domains, our novel finding is consistent with these other studies in which these agents positively influenced users' behaviors (Pires et al., 2017; Olafsson, O'Leary, & Bickmore, 2020). Importantly, the findings suggest that a near-peer agent approach with upper elementary children may not "backfire", a phenomenon that has been observed with older users and which we had recognized as a possible outcome throughout the project (Garner et al., 1992; Schroeder & Gotch, 2015). However,

learners also sometimes noted that the agents could be "annoying" or "creepy". These potentially negative interpretations of the agents' actions should always be considered in order to effectively design agents that aid more than they distract.

6.3.2. Designing near-peer agents requires extensive iterative refinement with input from the target users

Regardless of how experienced a design team is with the target user population, designing agents that "feel like" a peer to a user is a challenging endeavor. Involving children in this project was crucial to understand what they desired from a virtual learning companion. Our iterative design process engaged children in participatory design through a form of cooperative inquiry (Druin, 1999) and provided meaningful insight into the minds of our target user population to help us design agents who felt authentic. With these children playing the roles of an informant and design partner (Druin, 2002), they demonstrated that judgments made about whether an agent is "like them" is based on many factors, including the setting the agent is drawn in, their clothing, hairstyle, tone of voice, dialogue, and even details as specific as eye direction, size, and shape. Furthermore, when designing a dyad of near-peer agents, such as Jeremy and Viviana, it is also critical to workshop the agents' interactions with children: Are the agents interacting like children at the intended age? While it is not possible to design agents that appeal to every user, an iterative process in which users are treated as collaborators shows promise to support the design of agents that children perceive as near peers. A direction for future work could include further personalization of the agents.

6.3.3. Effectively conveying suggested behavior through a non-authority-figure agent is a delicate balancing act

When we initially designed the dialogue of the virtual learning companions, we adhered to a strict principle of non-authority: the agents did not tell the human children what to do, but rather simply shared their experiences. We hoped the children would extrapolate to their own behavior. However, we discovered that the agents had to be more direct; otherwise, the children did not see the connection to their

Previous "Asking Why" Vignette	→	Rewrites
<p>Viviana: When I come up with an idea, Jeremy asks me: "Why did you do that?"</p> <p>Jeremy: That's because we need to ask "why" so we can understand what we're doing.</p>		<p>Viviana: When I come up with an idea, Jeremy asks me: "Why did you do that?"</p> <p>Jeremy: Don't be afraid to ask "why" to your partner! It'll help you understand things!</p>
Previous "Sharing Ideas" Vignette	→	Rewrites
<p>Viviana: Sometimes I have an idea but Jeremy has a different one.</p> <p>Jeremy: It's great when we share different ideas.</p> <p>Viviana: It makes our code better when we *both* share our ideas.</p> <ul style="list-style-type: none"> • Passive, easy to ignore • No direct instructions to act upon 		<p>Viviana: Sometimes I have an idea but Jeremy has a different one. So we try them both out!</p> <p>Jeremy: Some of them work and some don't, but that's okay! So why not try and share your ideas?</p> <ul style="list-style-type: none"> • Introduce active/direct voice • Provide actionable items

Fig. 12. Highlighted in yellow, the previous dialogue iteration shows indirect modeling of sharing ideas. The rewritten iteration provides a more direct call to action while trying to avoid authoritative wording.

own work. For example, consider a previous iteration of a dialogue highlighting the importance of asking "Why" questions:

Viviana: When I come up with an idea, Jeremy asks me: "Why did you do that?"

Jeremy: That's because we need to ask "why" so we can understand what we're doing.

While this was an implied call to action, we later modified the dialogue to make the intention more direct. As shown in Fig. 12, we modified Jeremy's response:

Viviana: When I come up with an idea, Jeremy asks me: "Why did you do that?"

Jeremy: Don't be afraid to ask "why" to your partner! It'll help you understand things!

In another example of a dialogue that we made more direct, initially, Viviana began with, "Sometimes I have an idea but Jeremy has a different one", and Jeremy responded, "It's great when we share different ideas". Viviana then asserted the main idea: "It makes our code better when we *both* share our ideas". The revised dialogue iteration was shortened and made more direct. Viviana initiates with, "Sometimes I have an idea but Jeremy has a different one. So we try them both out!", and Jeremy replies with a soft call to action: "Some of them work and some don't, but that's okay! So why not try and share your ideas?"

7. Limitations

The current work has several important limitations. First, due to the repercussions of the COVID-19 pandemic, the sample size in our classroom study was limited. Null results (such as lack of difference between conditions in several aspects of dialogue or pre/post comparisons) may be attributable to a lack of statistical power. Further investigation is needed to examine these effects and to determine whether the findings reported here generalize to other populations of children. Second, the breadth of the agent dialogue library for the current iteration of FLECKS only supports a few days of project work (after which the vignettes would become repetitive). It is possible that additional effects will be observed with longer-term interactions with the learning companions.

Third, we did not account for children's prior relationships with one another when we randomly assigned partners, and these relationships could have had significant impacts on any given pair's experience. Finally, we conducted this study in the context of computer science learning, and further studies are needed to investigate whether the effects generalize to different domains.

8. Conclusion and future work

Collaboration between children is a powerful tool for learning, but productive collaborative skills need to be scaffolded and promoted. Virtual learning companions provide many opportunities for offering real-time feedback to children as they collaborate. The findings reported in this paper show that virtual learning companions can positively influence children's collaborative talk. The virtual learning companions fostered more higher-order questions than a control condition and supported significantly more positive attitudes toward computer science.

The results described here point to numerous areas for future work. One such direction concerns the length of interactions: future work should investigate the longitudinal effects of learning companions who are built to engage learners over the course of an academic term or year. The results presented here also raise the question of where the role of virtual learning companions should end. It is important to move toward greater support of domain and task knowledge through intelligent technologies, but providing content-related hints and feedback breaks the near-peer positionality of a learning companion agent. It is also important to delve deeper into the dialogue that surrounds exploratory talk and investigate how learning companions can support learners' collaboration using this additional data. Future work should investigate how best to design domain and task support alongside support for collaborative talk. By investigating these and related research directions, we can harness the power of virtual learning companions to support children in learning to collaborate well.

CRediT authorship contribution statement

Toni V. Earle-Randell: Data curation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing, Investigation. **Joseph B. Wiggins:** Conceptualization, Formal analysis, Methodology, Supervision, Writing – original draft. **Yingbo Ma:** Data curation,

Methodology, Writing – original draft. **Mehmet Celepkolu:** Conceptualization, Data curation, Supervision, Writing – original draft, Project administration, Investigation, Resources, Writing – review & editing. **Dolly Bounajim:** Data curation, Formal analysis, Writing – original draft. **Zhikai Gao:** Investigation, Writing – original draft, Formal analysis. **Julianne Martinez Ruiz:** Formal analysis, Visualization, Writing – original draft. **Kristy Elizabeth Boyer:** Conceptualization, Funding acquisition, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Maya Israel:** Conceptualization, Project administration, Supervision, Writing – review & editing, Funding acquisition, Resources. **Collin F. Lynch:** Conceptualization, Project administration, Supervision, Writing – review & editing, Funding acquisition, Resources. **Eric Wiebe:** Conceptualization, Project administration, Supervision, Writing – review & editing, Funding acquisition, Resources.

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.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work is supported by the National Science Foundation through grant DRL-1721160. Any opinions, findings, conclusions, or recommendations expressed in this report are those of the authors and do not necessarily represent the views of the National Science Foundation.

References

- Alaimi, M., Law, E., Pantasdo, K. D., Oudeyer, P.-Y., & Sauzeon, H. (2020). Pedagogical agents for fostering question-asking skills in children. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–13).
- Austin, J. L. (1975). *How to do things with words*. Oxford University Press.
- Baker, D. L., Ma, H., Polanco, P., Conry, J. M., Kamata, A., Al Otaiba, S., et al. (2021). Development and promise of a vocabulary intelligent tutoring system for second-grade Latinx English learners. *Journal of Research on Technology in Education*, 53(2), 223–247.
- Barnes, D. (2008). Exploratory talk for learning. (pp. 1–15). Exploring Talk in School.
- Barnett, J. E., & Francis, A. L. (2012). Using higher order thinking questions to foster critical thinking: A classroom study. *Educational Psychology*, 32(2), 201–211.
- Bosch, N., Zhang, Y., Paquette, L., Baker, R., Ocumphaug, J., & Biswas, G. (2021). Students' verbalized metacognition during computerized learning. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–12).
- Boyd, M., & Kong, Y. (2017). Reasoning words as linguistic features of exploratory talk: Classroom use and what it can tell us. *Discourse Processes*, 54(1), 62–81.
- Calder, N. (2010). Using scratch: An integrated problem-solving approach to mathematical thinking. *Australian Primary Mathematics Classroom*, 15(4), 9–14.
- Ceha, J., & Law, E. (2022). Expressive auditory gestures in a voice-based pedagogical agent. In *CHI conference on human factors in computing systems* (pp. 1–13).
- Chan, T.-W., & Baskin, A. B. (1988). Studying with the prince: The computer as a learning companion. In *Proceedings of the international conference on intelligent tutoring systems: vol. 194200*.
- Chen, Z.-H., Liao, C., Chien, T.-C., & Chan, T.-W. (2011). Animal companions: Fostering children's effort-making by nurturing virtual pets. *British Journal of Educational Technology*, 42(1), 166–180.
- Chou, C.-Y., Chan, T.-W., & Lin, C.-J. (2003). Redefining the learning companion: The past, present, and future of educational agents. *Computers & Education*, 40(3), 255–269.
- Core, M. G., & Allen, J. (1997). Coding dialogs with the DAMSL annotation scheme. In *AAAI fall symposium on communicative action in humans and machines: vol. 56*, (pp. 28–35). Boston, MA.
- Deiglmayr, A., & Spada, H. (2010). Collaborative problem-solving with distributed information: The role of inferences from interdependent information. *Group Processes & Intergroup Relations*, 13(3), 361–378.
- Druin, A. (1999). Cooperative inquiry: Developing new technologies for children with children. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 592–599). New York, NY, USA: Association for Computing Machinery, ISBN: 0201485591.
- Druin, A. (2002). The role of children in the design of new technology. *Behaviour and information technology*, 21(1), 1–25.
- Fagerlund, J., Vesinenaho, M., & Häkkinen, P. (2022). Fourth grade students' computational thinking in pair programming with scratch: A holistic case analysis. *International Journal of Child-Computer Interaction*, 33, Article 100511.
- Fernández, M., Wegerif, R., Mercer, N., & Rojas-Drummond, S. (2015). Reconceptualizing "scaffolding" and the zone of proximal development in the context of symmetrical collaborative learning. *Journal of Classroom Interaction*, 54–72.
- Garner, R., Brown, R., Sanders, S., & Menke, D. J. (1992). "Seductive details" and learning from text. In *The role of interest in learning and development* (pp. 239–254). Lawrence Erlbaum Associates, Inc.
- Graesser, A. C., Forsyth, C. M., & Lehman, B. A. (2017). Two heads may be better than one: Learning from computer agents in conversational dialogues. *Teachers College Record*, 119(3), 1–20.
- Han, J.-H., Shubeck, K., Shi, G.-H., Hu, X.-E., Yang, L., Wang, L.-J., et al. (2021). Teachable agent improves affect regulation. *Educational Technology & Society*, 24(3), 194–209.
- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In *Assessment and teaching of 21st century skills* (pp. 37–56). Springer.
- Hinkle, M., Rachmatullah, A., Mott, B., Boyer, K. E., Lester, J., & Wiebe, E. (2020). The relationship of gender, experiential, and psychological factors to achievement in computer science. In *Proceedings of the 2020 ACM conference on innovation and technology in computer science education* (pp. 225–231).
- Hsu, S.-H., Chou, C.-Y., Chen, F.-C., Wang, Y.-K., & Chan, T.-W. (2007). An investigation of the differences between robot and virtual learning companions' influences on students' engagement. In *2007 first IEEE international workshop on digital game and intelligent toy enhanced learning* (pp. 41–48). IEEE.
- Joaquim, S., Bittencourt, I. I., de Amorim Silva, R., Espinheira, P. L., & Reis, M. (2022). What to do and what to avoid on the use of gamified intelligent tutor system for low-income students. *Education and Information Technologies*, 27(2), 2677–2694.
- Katuka, G. A., Webber, A. R., Wiggins, J. B., Boyer, K. E., Magerko, B., McKlin, T., et al. (2022). The relationship between co-creative dialogue and high school learners' satisfaction with their collaborator in computational music remixing. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW1), 1–24.
- Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical Teacher*, 42(8), 846–854.
- Kim, Y. (2005). Empathetic virtual peers enhanced learner interest and self-efficacy. In *Workshop on motivation and affect in educational software, in conjunction with the 12th international conference on artificial intelligence in education* (pp. 9–16).
- Kim, S. (2019). Playing with machine translation in language classroom: Affordances and constraints. *Multimedia-Assisted Language Learning*, 22(2), 9–28.
- Kim, Y., Baylor, A. L., & Shen, E. (2007). Pedagogical agents as learning companions: the impact of agent emotion and gender. *Journal of Computer Assisted Learning*, 23(3), 220–234.
- Kumar, R., & Rosé, C. P. (2014). Triggering effective social support for online groups. *ACM Transactions on Interactive Intelligent Systems (TiIS)*, 3(4), 1–32.
- Lai, A.-F., & Yang, S.-M. (2011). The learning effect of visualized programming learning on 6th graders' problem solving and logical reasoning abilities. In *2011 international conference on electrical and control engineering* (pp. 6940–6944). IEEE.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.
- Lewis, C. M. (2011). Is pair programming more effective than other forms of collaboration for young students? *Computer Science Education*, 21(2), 105–134.
- Lopes, A. M. M., de Magalhães Netto, J. F., de Souza, R. A., Mourão, A. B., Almeida, T. O., & de Lima, D. P. (2019). Improving students' skills to solve elementary equations in K-12 programs using an intelligent tutoring system. In *2019 IEEE frontiers in education conference* (pp. 1–9). IEEE.
- Mercer, N. (2002). *Words and minds: how we use language to think together*. Routledge.
- Mercer, N., & Howe, C. (2012). Explaining the dialogic processes of teaching and learning: The value and potential of sociocultural theory. *Learning, Culture and Social Interaction*, 1(1), 12–21.
- Mercer, N., Wegerif, R., & Dawes, L. (1999). Children's talk and the development of reasoning in the classroom. *British Educational Research Journal*, 25(1), 95–111.
- 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.
- Olafsson, S., O'Leary, T. K., & Bickmore, T. W. (2020). Motivating health behavior change with humorous virtual agents. In *Proceedings of the 20th ACM international conference on intelligent virtual agents* (pp. 1–8).
- Paranjape, B., Ge, Y., Bai, Z., Hammer, J., & Cassell, J. (2018). Towards automatic generation of peer-targeted science talk in curiosity-evoking virtual agent. In *Proceedings of the 18th international conference on intelligent virtual agents* (pp. 71–78).
- Pezzullo, L. G., Wiggins, J. B., Frankosky, M. H., Min, W., Boyer, K. E., Mott, B. W., et al. (2017). "Thanks Alisha, keep in touch": Gender effects and engagement with virtual learning companions. In *International conference on artificial intelligence in education* (pp. 299–310). Springer.
- Pires, A., Alves-Oliveira, P., Arriaga, P., & Martinho, C. (2017). Cubus: Autonomous embodied characters to stimulate creative idea generation in groups of children. In *International conference on intelligent virtual agents* (pp. 360–373). Springer.

- Raheja, V., & Tetreault, J. (2019). Dialogue act classification with context-aware self-attention. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies: vol. 1 (long and short papers)*, (pp. 3727–3733). Association for Computational Linguistics.
- Renaud, R. D., & Murray, H. G. (2007). The validity of higher-order questions as a process indicator of educational quality. *Research in Higher Education*, 48(3), 319–351.
- Riek, L. D. (2012). Wizard of Oz studies in HRI: A systematic review and new reporting guidelines. *Journal of Human-Robot Interaction*, 1(1), 119–136.
- Rojas-Drummond, S., & Zapata, M. P. (2004). Exploratory talk, argumentation and reasoning in mexican primary school children. *Language and Education*, 18(6), 539–557.
- Ryokai, K., Vaucelle, C., & Cassell, J. (2003). Virtual peers as partners in storytelling and literacy learning. *Journal of Computer Assisted Learning*, 19(2), 195–208.
- Schroeder, N. L., & Gotch, C. M. (2015). Persisting issues in pedagogical agent research. *Journal of Educational Computing Research*, 53(2), 183–204.
- Schlute, C., & Knobelsdorf, M. (2007). Attitudes towards computer science-computing experiences as a starting point and barrier to computer science. In *Proceedings of the third international workshop on computing education research* (pp. 27–38).
- Shah, N., Lewis, C., & Caires, R. (2014). Analyzing equity in collaborative learning situations: A comparative case study in elementary computer science. In *Learning and becoming in practice: the international conference of the learning sciences*. Boulder, CO: International Society of the Learning Sciences.
- Stolcke, A., Ries, K., Coccaro, N., Shriberg, E., Bates, R., Jurafsky, D., et al. (2000). Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational Linguistics*, 26(3), 339–373.
- Sylvia Rojas-Drummond, N. M. (2003). Scaffolding the development of effective collaboration and learning. *International Journal of Educational Research*, 39, 99–111.
- Tenenbaum, H. R., Winstone, N. E., Leman, P. J., & Avery, R. E. (2020). How effective is peer interaction in facilitating learning? A meta-analysis. *Journal of Educational Psychology*, 112(7), 1303.
- Thissen, D., Steinberg, L., & Kuang, D. (2002). Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of Educational and Behavioral Statistics*, 27(1), 77–83.
- Tsan, J., Vandenberg, J., Zakaria, Z., Boulden, D. C., Lynch, C., Wiebe, E., et al. (2021). Collaborative dialogue and types of conflict: An analysis of pair programming interactions between upper elementary students. In *Proceedings of the 52nd ACM technical symposium on computer science education* (pp. 1184–1190).
- T'Sas, J. (2018). *Learning outcomes of exploratory talk in collaborative activities* (Ph.D. thesis), University of Antwerp.
- Vandenberg, J., Rachmatullah, A., Lynch, C., Boyer, K. E., & Wiebe, E. (2021). Interaction effects of race and gender in elementary CS attitudes: A validation and cross-sectional study. *International Journal of Child-Computer Interaction*, 29, Article 100293.
- Vandenberg, J., Rachmatullah, A., Lynch, C., Boyer, K. E., & Wiebe, E. (2021). The relationship of CS attitudes, perceptions of collaboration, and pair programming strategies on upper elementary students' CS learning. In *Proceedings of the 26th ACM conference on innovation and technology in computer science education: vol. 1*, (pp. 46–52).
- Vygotsky, L. (1978). Interaction between learning and development. *Readings on the Development of Children*, 23(3), 34–41.
- Vygotsky, L. S., & Cole, M. (1978). *Mind in society: development of higher psychological processes*. Harvard University Press.
- Wegerif, R., & Dawes, L. (1998). Encouraging exploratory talk around computers. *IT for Learning Enhancement*, 10–29.
- Wegerif, R., Mercer, N., & Major, L. (2019). *The Routledge international handbook of research on dialogic education, Introduction to the Routledge international handbook of research on dialogic education* (pp. 1–8). Routledge.
- Wiggins, J. B., Earle-Randell, T. V., Bounajim, D., Ma, Y., Ruiz, J. M., Liu, R., et al. (2022). Building the dream team: children's reactions to virtual agents that model collaborative talk. In *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents* (pp. 1–8).
- Wiggins, J. B., Wilkinson, J., Baigorria, L., Huang, Y., Boyer, K. E., Lynch, C., et al. (2019). From doodles to designs: participatory pedagogical agent design with elementary students. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (pp. 642–647). ACM.
- Williams, L., Wiebe, E., Yang, K., Ferzli, M., & Miller, C. (2002). In support of pair programming in the introductory computer science course. *Computer Science Education*, 12(3), 197–212.
- Yu, W. J., Jang, J. H., Ahn, J. M., Park, D. R., Yoo, I. H., Bae, Y. K., et al. (2019). Design of artificial intelligence education program based on design-based research. *International Journal of Advanced Smart Convergence*, 8(4), 113–120.
- Zakaria, Z., Vandenberg, J., Tsan, J., Boulden, D. C., Lynch, C. F., Boyer, K. E., et al. (2022). Two-computer pair programming: Exploring a feedback intervention to improve collaborative talk in elementary students. *Computer Science Education*, 32(1), 3–29.
- Zhong, B., Wang, Q., & Chen, J. (2016). The impact of social factors on pair programming in a primary school. *Computers in Human Behavior*, 64, 423–431.