"Elinor's Talking to Me!":Integrating Conversational AI into Children's Narrative Science Programming

Ying Xu
University of California, Irvine
Irvine, California, USA
ying.xu@uci.edu

Valery Vigil
University of California, Irvine
Irvine, California, USA
vigilv@uci.edu

Andres Bustamante
University of California, Irvine
Irvine, California, USA
asbustam@uci.edu

Mark Warschauer
University of California, Irvine
Irvine, California, USA
markw@uci.edu

ABSTRACT

Video programs are important, accessible educational resources for young children, especially those from an under-resourced backgrounds. These programs' potential can be amplified if children are allowed to socially interact with media characters during their video watching. This paper presents the design and empirical investigation of interactive science-focused videos in which the main character, powered by a conversational agent, engaged in contingent conversation with children by asking children questions and providing responsive feedback. We found that children actively interacted with the media character in the conversational videos and their parents spontaneously provided support in the process. We also found that the children who watched the conversational video performed better in the immediate, episode-specific science assessment compared to their peers who watched the broadcast, non-interactive version of the same episode. Several design implications are discussed for using conversational technologies to better support child active learning and parent involvement in video watching.

CCS CONCEPTS

• **Social and professional topics → Children**; • **Human-centered computing → Empirical studies in interaction design**.

KEYWORDS

Conversational AI, conversational agents, science learning, children, educational media

ACM Reference Format:
Ying Xu, Valery Vigil, Andres Bustamante, and Mark Warschauer. 2022. "Elinor's Talking to Me!":Integrating Conversational AI into Children's Narrative Science Programming. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3491102.3502050

1 INTRODUCTION

Early childhood is an important time for the development of knowledge, skills, and attitudes that help young children become scientifically literate citizens. Yet science learning, especially when it involves abstract concepts like force, matter, and energy, can be challenging for young children [31]. One effective strategy for making science learning more accessible is to integrate science lessons into narrative stories [24, 58]. A compelling narrative contains rich contexts, interesting story plotlines, and attractive characters, making otherwise abstract scientific ideas more relatable and concrete for young learners.

Children are most commonly exposed to science narratives through educational videos and television shows [5, 50]. For example, the *Magic School Bus*[1] takes children on field trips with Ms. Frizzle under the sea, to outer space, and even inside human bodies. Another popular show, *Sid the Science Kid*[2], introduces children to everyday science concepts from the perspective of a curious young boy as he learns new things about the world around him, such as why bananas get mushy. Decades of research have shown that this type of science narrative programming is a useful resource for young children's learning.

However, these conventional television and video programs usually take a linear, non-interactive form and have not effectively leveraged technology that could provide children with a more personalized and enriching experience [19, 63, 82]. Yet modern Internet-connected devices, such as tablets, smartphones, and smart TVs, have become the primary means through which children consume video media [70], increasing the opportunities for media producers to integrate interactive technologies into science narrative programming.

Recently, voice-based interaction has been introduced to enhance children's educational media. A growing number of projects have integrated conversational agents into smart speaker apps [28, 92, 93], social robots [85], Internet-connected toys [21], and intelligent learning systems [32, 66]. However, few studies have incorporated conversational agents as part of immersive science narratives that allow children to interact with story characters in a natural, social way.

In this paper, we report a two-year research and development project in partnership between University of California Irvine and PBS KIDS. In this project, we interrogated the design, usability, and effectiveness of integrating conversational agents into a children's science television show, so that children can interact with the show's main character and help them find solutions to problems they encounter. The media character is capable of comprehending a child's speech input and responding to the child contingently

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04.
https://doi.org/10.1145/3491102.3502050

[1]This popular animated educational series originally aired in the United States on PBS in the mid-1990s. It has been rebroadcast multiple times and is now available on Netflix. For more details, please see https://www.netflix.com/title/70264612
[2]This animated series originally aired in the US on PBS KIDS from 2008 to 2013. For more details, please see https://pbskids.org/sid

to advance the interaction. We engaged in an iterative design process, conducted usability testing, and carried out a randomized experiment to examine the benefits incorporating conversational interactions had on children's learning and engagement. Overall, we found that these "conversational videos" could be feasibly used by young children, and the children and their parents held positive perceptions about such programs. Having conversation with the program character also improved children's performance in an immediate post-test on science learning. Among the first studies of this kind, our study offers insights to the design, usability, and effectiveness of interactive television and video programming.

2 RELATED WORK

2.1 Science Learning, Narrative Programming, and Interactive TV

There is a long tradition of research that focuses on narrative-oriented science learning in which lessons are integrated into stories [22, 24, 47]. This strategy is thought to make science learning more intuitive for and accessible to young children, with the narrative structure serving as the underlying foundation that facilitates children's understanding and memorization. Fisch's capacity model of children's comprehension posits that presenting educational content in narrative forms lessens the cognitive burden of processing new information, leading to enhanced learning [1].

Science television shows have been one major source of narrative-based science learning widely available to young children. In 2020, three- to six-year-olds, on average, spent about two hours per day watching television or videos, with much of this being educational content [71]. Eight out of the ten most popular preschool STEM shows in the U.S., according to Common Sense Media, use stories to convey information, and studies have shown that this approach increases children's learning and engagement [5]. For example, Bonus and Mares found that children learned better from watching an episode containing narrative descriptions (i.e., day-night cycles explained as the sun and moon playing hide and seek around the earth) than from watching an episode that was only factual and did not feature narrative content [5].

Researchers also stress that children's engagement with and comprehension of educational content can be further promoted by adding in-the-moment interaction to television programming [67, 79–81]. In fact, almost all guidelines developed by individual research groups or institutions recommend including interactivity when designing children's educational media [41, 65, 84]. Since the early days of television, various efforts have been made to increase the interactivity of television programming (for an overview of "interactive TV," see [46]). One of the earliest experiments with interactive TV for children can be dated to the 1950s, when a U.S. children's show *Winky Dink and You* [3] developed supplemental materials that included a thin sheet of plastic that adhered to television screens so that children can use an erasable crayon to draw on the sheet. The program featured a cartoon character who encountered a series of problems, and the host invited child viewers to interact with the program character by completing connect-the-dot drawings of different objects to help the character solve the problem (e.g., drawing a bridge to help the character cross a chasm) [77].Another form of interactivity that quickly gained popularity was a technique that features pseudo-conversation, in which the character asks the viewer a question, pauses for a moment, and finally provides a generic response. Many children's television programs, including Mr. Roger's Neighborhood in the 1980s, Blue's Clues and Dora the Explorer in the 2000s, and the more recent Ask the Storybots, have adopted this technique in a generally similar fashion. Although there is some evidence suggesting that this kind of pseudo-conversation has certain learning benefits [30, 52], more evidence claims that such learning benefits are minimal or non-existent [11, 53, 68, 72, 83], especially if the pseudo-conversation fails to correct children's misunderstandings [80]. Moreover, this technique has limited effects on children's engagement since children often realize that they cannot actually influence the story or the characters' responses [13]. Nevertheless, television programs' reliance on pseudo-conversation is partially due to the technical constraints of traditional cabled television and its one-way information stream.

It is noteworthy that in addition to making television shows interactive, media producers have also developed apps or websites to accompany educational shows and to provide other types of interactivity. For example, McCarthy and colleagues developed a mobile mathematics game featuring familiar characters from PBS KIDS shows, and their evaluation suggested that this interactive component improved children's math ability [57].

In recent years, Internet-connected smart devices have become the standard for television and video watching, and these devices offer an opportunity to truly realize real-time, two-way interaction in video programming. Most of the innovations involving interactive programming have targeted adult audiences. For example, live stream videos often allow viewers to communicate with the host and with each other through open audio channel or instant messaging [40], choose-your-own-adventure programs enable individual users to decide how the story progresses [60, 87], and augmented reality television allows individual viewers to explore movie scenes from a variety of perspectives [86]. Moreover, there is an emerging interest in incorporating "conversational characters" so that viewers can interact with on-screen characters using speech, and several studies have suggested its feasibility for adult audiences (for an overview, see [25]).

Only recently have media producers begun to explore integrating this kind of voice-based interaction into children's television shows, largely because of the advances in natural language processing technology along with the dramatic increase in young children's use of voice interfaces (e.g., Amazon Alexa, Apple Siri). For example, Gray and colleagues carried out a project to embed conversational agents within an app-based video featuring Sesame Street characters. The goal of the project was to strengthen children's parasocial relationship for the show's characters, thereby enhancing children's motivation and learning from the video content [37]. While this paper detailed the design of the conversational characters, it did not report results of testing with children and thus did not provide substantive evidence about the program's usability and effectiveness.

[3]Winky Dink and You was a Columbia Broadcasting System (CBS) children's television show that aired from 1953 to 1957.

Nevertheless, numerous studies using the Wizard of Oz approach have confirmed that integrating conversational characters into children's television narratives can result in positive engagement and learning outcomes. For example, Carters found that characters responding with appropriate timing and repeating questions that children did not answer increased children's verbal participation [13]. Similarly, Calvert found that children who had opportunities to talk about math with a conversational character embedded within a popular animated children's series, Dora the Explorer, performed better in a math assessment task than did children not given such opportunities [8]. These studies collectively point to the potential benefits of automating such interactions so that children's contingent interactions with video characters in educational programming can be implemented at scale.

2.2 Conversational Agent-based Systems

Although automated conversational agents have not been fully integrated into children's television shows, such technology, more broadly, has been incorporated in many children's toys, social robots, and intelligent learning systems. While these conversational agent-based applications have different features and functionalities, one key common affordance is their ability to enable two-way contingent dialogue with a child user. For example, the smart toy Hello Barbie begins the user interaction with a preset list of prompts (e.g., "How are you doing?") and then provides relevant replies based on its understanding of children's responses [59]; the dinosaur robot CogniToys Dino invites children to ask questions and provides child-friendly answers (just like a child-friendly version of Amazon Alexa or Google Assistant) [45]; and the pedagogical agent AutoTutor simulates dialogue with a human tutor by asking students curriculum-based questions and following up to either reinforce students' correct responses or to clarify their misconceptions [35]. Studies have revealed that children can form parasocial relationships with the conversational agents they interact with. For example, our prior study suggested that children who conversed with a conversational agent embedded within a smart speaker personified the agent and regarded it as sociable and smart [94]. Garg and Sengupta also found that children aged five to seven who had a smart speaker device in their home developed emotional attachments with the conversational agent [29].

In general, researchers are optimistic that conversational agents embedded in children's toys, robots, and learning systems can foster children's information recall, learning, and engagement. Many of the existing studies focus on using these voice-based systems as children's learning companions or to provide direct instruction. Breazeal developed a robot embedded within a plush doll that was intended to teach children about exotic animals by engaging children in dialogue [6]. The robot described the animal (e.g., "*I like how it's white with such big antlers!*") and intermittently asked children questions to allow children to respond verbally (e.g., "*Did you know it can go for weeks without drinking water?*"). A posttest revealed that children were able to successfully learn the information the robot had taught them. Another group of researchers developed a robot that played a food-selection game with children and then talked about that food item with the children in French [26]. The study found that this game-like conversation helped children learn the French words introduced by the robot. Similar results were found in Ryokai, Vaucelle and Cassell, in which children learned storytelling strategies from a virtual peer they interacted with [14, 74].

In addition to learning benefits, agent-based conversation can also increase children's engagement in specific subject domains (e.g., [90]). For example, Shiomi placed a social robot in a classroom setting and found that the robot, which encouraged children to ask science-related questions and then provided answers, enhanced the science curiosity of the children who interacted with the robot [76]. Michaelis and Mutlu developed a robot that reacted to children as they read science books aloud. The robot provided expressive speech, nonverbal cues, and personal comments, which effectively cultivated children's interest in science reading [61].

In addition, there is a large body of studies that incorporate speech-based conversational agents to support student learning from intelligent learning systems (for an overview, see [42]). Empirical evidence has suggested that conversational agents engaged in focused pedagogy yielded learning gains comparable to those of trained human tutors [36, 88]. While these studies are distinct from our own in that these intelligent learning systems position the conversational agents as tutors or teachers within lesson-based curriculum, they are also relevant to our study since the general dialogue structure is similar (i.e., the agent asking questions and providing contingent feedback based on specific learning objectives) [34]. One notable exception of integrating agents into narrative-based learning was Ruan [73], in which children learned math concepts through helping solve problems in an adventure. The authors found that including a character-based chatbot tutor (using Wizard of Oz method) boosted children's engagement and learning outcomes.

Despite these promises, several studies have suggested that children, in general, are likely to encounter obstacles as they interact with conversational agents [15]. These obstacles may influence how much children engage with and learn from such interactions. For example, if children's responses are not accurately registered by the agent, children may become frustrated and would also not likely receive appropriate feedback. Studies have also suggested that a child's individual characteristics may play a role in how they interact with conversational agents. For example, Xu and Warschauer found that younger children (3- and 4-year-olds) more frequently ignored a conversation prompt or used gestures instead of speech to respond, compared to older children (5- and 6-year-olds) [93]. Monarca et al. also examined how children's language ability affected their speech production when talking to a conversational agent [62]. Specifically, the authors found that children with lower language ability generated shorter utterances and needed more time to formulate their responses than did those with higher language ability. As such, some children may need additional support to effectively interact with voice-based agents.

Some studies have begun to investigate how to best design such support. Based on a close analysis of children's communication breakdowns with Alexa devices and parents' repair strategies, Beneteau proposed that future systems incorporate "discourse scaffolding" that emulates how parents ask their children follow-up questions to guide them in clarifying their previous response for the agent [4]. In fact, this strategy has been implemented in Xu and Warschauer, in which a conversational agent was designed to

present children with multiple response options as a fallback mechanism during storybook reading with the conversational agent [93]. This strategy was shown to increase children's response rates and response quality while also preventing potential breakdowns.

3 DESIGN OF THE CONVERSATIONAL VIDEOS

The conversational videos developed in our study are adapted from a popular children's science animation show *Elinor Wonders Why* which debuted in September 2020. *Elinor Wonders Why* centers around Elinor, a young rabbit who has many characteristics of a budding scientist, such as curiosity, perseverance, and willingness to experiment. Each episode typically begins with Elinor encountering an interesting problem, and her scientific discoveries unfold within a rich plotline that includes other characters, events, and settings. The conversational videos allow children to speak directly with Elinor, thus priming them to engage in observation, prediction, pattern finding, and problem solving as the story progresses.

We have developed three conversational videos corresponding to three episodes of *Elinor Wonders Why*. The first episode teaches the concepts of liquid viscosity, concentration, and dilution as Elinor enjoys the delicious honey and ketchup at a town picnic. The second episode demonstrates the idea of aerodynamics through Elinor experimenting ways of building a fast cardboard car to win a race. The third episode focuses on the concepts of reptile molting as Elinor observes how and why snakes shed their skin during a soccer game. In the rest of this section, we will illustrate the design of the conversational videos using the third episode on snake molting as an example.

3.1 Conversation Design Principles

Given that the main objective of the conversational videos is to promote learning and engagement, the design emphasizes educational and playful perspectives. The design principles were grounded in the literature in communication and STEM learning in face-to-face settings, as well as studies that shed light on the design implications of conversational agents for learning purposes (e.g., [28, 91, 93]). Figure 1 display an example conversation between Elinor and a child. Appendix A consists of a complete list of questions included in each of the episodes.

3.1.1 The Educational Principle. To achieve educational goals, we design the dialogue flow based on two relevant frameworks: general strategies for effective person-to-person dialogue with children (e.g., [78]) and the Next Generation Science Standard (NGSS) that lays out the core science learning goal posts for young children [7].

The existing research on general interpersonal communication strategies points to three key components of effective dialogue with children (for an overview, see [78]). First, asking children **questions** stimulates their thoughtful responses. Second, offering elaborative **feedback** continues the conversation around the particular topic. Third, providing **scaffolding** helps children better participate in the conversation. Several studies have applied these strategies to the development of conversational agents for preschool-aged children and have shown them to be useful and appropriate (e.g., [93]).

In our conversational videos, Elinor asks children questions that are open-ended in nature and start with *"why," "how,"* or *"what"*.

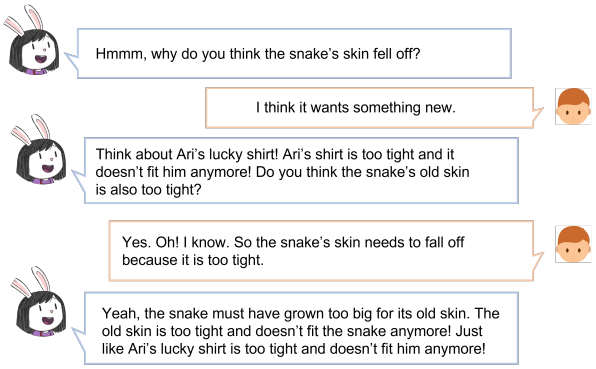

Figure 1: An example conversation between Elinor and a child

For example, after noticing how a snake's once-shiny skin had become dull, Elinor wonders aloud whether the snake is sick. She encourages children to describe their own observations by asking them "What do you notice about this snake's skin?" This question challenges children to identify the key change in the snake's appearance related to molting apart from any unchanged characteristics they may also notice (e.g., the stripes).

After children respond to each of Elinor's questions, Elinor provides feedback to help children deepen their understanding of the topic being explored. For example, after asking children to describe the change in the snake's skin, Elinor first acknowledges their specific response. Thus, if a child identifies the key change (e.g., *"the skin is dull"* or *"the snake is flaky"*), Elinor says *"Yeah, I saw that too."* And if the child responds with some other characteristics (e.g., the stripes), Elinor provides one of several tailored replies (e.g., *"Yeah, the snake does have stripes, but…."*). After acknowledging children's responses, Elinor then adds, *"the snake's skin is so flaky and not shiny. I wonder if the snake's flaky skin will fall off and it will have shiny skin again."* Elinor's additional explanations are designed to clearly articulate scientific concepts in an easy-to-understand way (e.g., contrasting changes such as flaky vs. shiny) while also foregrounding the episode's broader learning topic (e.g., snake's skin falling off as they grow).

Elinor also provides scaffolding to facilitate learning and conversation using a "hint and rephrasing" strategy that is commonly applied in building conversational agents [32, 33]. Specifically, if a child provides an unanticipated response or does not respond, Elinor offers additional information and also rephrases the original open-ended question into a multiple choice question. For example, when discussing why snakes shed their skin, Elinor originally asks children, *"Why do you think the snake's skin falls off?"* If a child does not respond or if their response is off topic, Elinor first makes an analogy involving another story character's shirt being too small and then rephrases the question (e.g., *"Think about Ari's lucky shirt! Ari's shirt is too tight and it doesn't fit him anymore! Do you think the snake's old skin is also too tight?"*). This scaffolding that refers to previous narrative elements of the episode helps children make logical connections to make sense of a new phenomena. Other types of scaffolding utilize visual hints. For example, to observe a snake's molted skin, Elinor uses a magnifying glass and she asks children

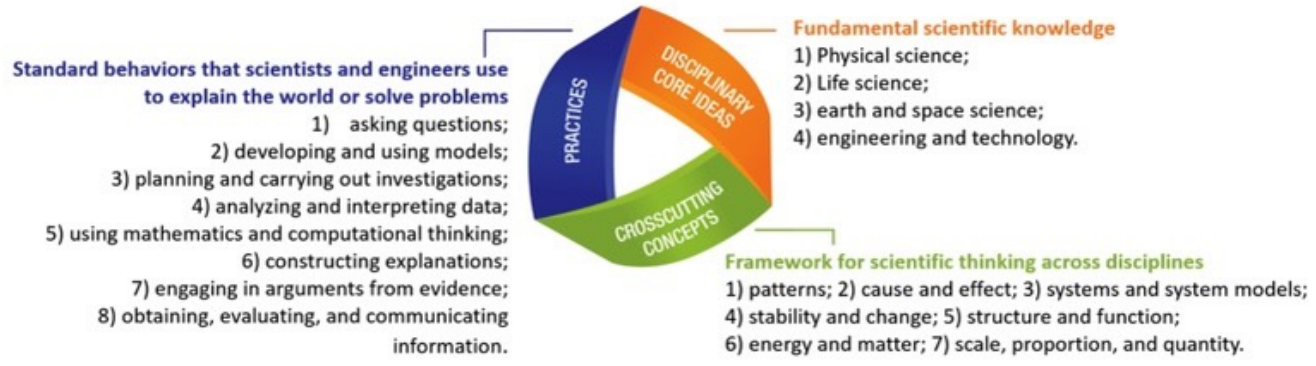

Figure 2: NGSS Overarching Scientific Goals [64].

how this tool helps her see better. If a child's response does not fall along the line of enlarging an object, a close-up shot appears to show a magnifying glass enlarging the molted skin and making its pattern clearer. With this visual, Elinor then rephrases the question as *"Did you see how this tool helped me observe the snake? Did it make the snake look bigger or did it make the snake look smaller?"*

In addition to using the question-feedback-scaffolding strategy, we design the conversation moments around NGSS's three overarching scientific goals: foundational knowledge, scientific thinking, and inquiry practices (see Figure 2). For example, the question *"Why do you think the snake's skin fell off?"* helps children build their life sciences knowledge, encourages them to think scientifically about structure and function, and leads them to construct explanations. In line with NGSS's recommendations, each conversation moment is designed to address each of the three overarching goals whenever possible.

3.1.2 The Playful Principle. Although a number of scholars have utilized slightly different "playful principles" to effectively design children's learning environments (e.g., [38, 43, 56, 89]), one common thread in connecting these applications is the general notion of learning through play-based activities with peers. Along this line, many children's AI learning systems are designed to take on the role of peers (e.g., [3, 66]), and this approach appeared to result in children's heightened autonomy and more enjoyable learning experiences as compared to systems that are designed as "teacher" or "guide" (for a review, see [51]). Thus, we design Elinor to be an ideal peer to facilitate this kind of learning. First, Elinor is characterized as a playful young child at an age similar to our target audience, which is consistent with the original *Elinor Wonders Why* show. Second, in some respect, the language Elinor uses mimics the way a child may talk to their peers during play, but in other respects, her language is carefully crafted to encourage children's curiosity and confidence. Specifically, Elinor's questions are worded as if she were discussing a new problem with a friend (e.g., *"Hmmm. How many snakes do you see in the box?" or "Oh, the snake's skin is so flaky and not shiny. I wonder what is happening. What do you think is going to happen?"*). Elinor's feedback to children's responses uses subjective language (e.g., *"Hmm, I'm not sure if that is a good idea. We just tried but it didn't work."* or *"Wow! That sounds like a great idea! Let's see if it works."*), rather than explicit judgement (e.g., *"you are right/wrong."*) to make it appear as if Elinor is learning alongside the child.

3.2 Dialogue Flow Architecture

Based on the design principles described above, we built the conversational agents for each episode using Google's cloud services Dialogflow. The conversational agent performs end-to-end language processing that classifies children's speech utterances into semantic intents (i.e., categorization of intended meaning). The agent's natural language understanding module was based on a generic pre-trained model built in the Dialogflow engine and then refined with utterances specific to the conversational moments in our video. By calibrating the generic model with the context-specific model, we enabled the agent to more precisely and accurately extract semantic intent from children's dialogue with Elinor.

As shown in Figure 3, given that children can respond to a particular question in a variety of ways, we trained the agent to associate more than one semantic intent with each conversational opportunity. These intents were created based on predicted responses formulated by the research team, as well as children's actual utterances during field testing. Creating multiple intents for each question helps to increase the specificity of Elinor's feedback.

3.3 Development Process

The development of the conversational videos for each episode consisted of multiple steps. The development was led by University of California Irvine and PBS KIDS, facilitated by a group of external consultants from the academia and media industry.

3.3.1 Development of conversation scripts. The research team used the educational and playful principles described above to draft a variety of conversation moments and integrate them into scripts for each episode. The scripts were then reviewed by a second group to ensure that they were consistent with the show's branding. We tested the conversational scripts with ten children aged four to six years (Mean age = 5.2 years; seven girls; six predominantly speaking English at home). Each child watched the original *Elinor Wonders Why* episode alongside a human experimenter who interacted with the child using the conversation script devised for the episode. The

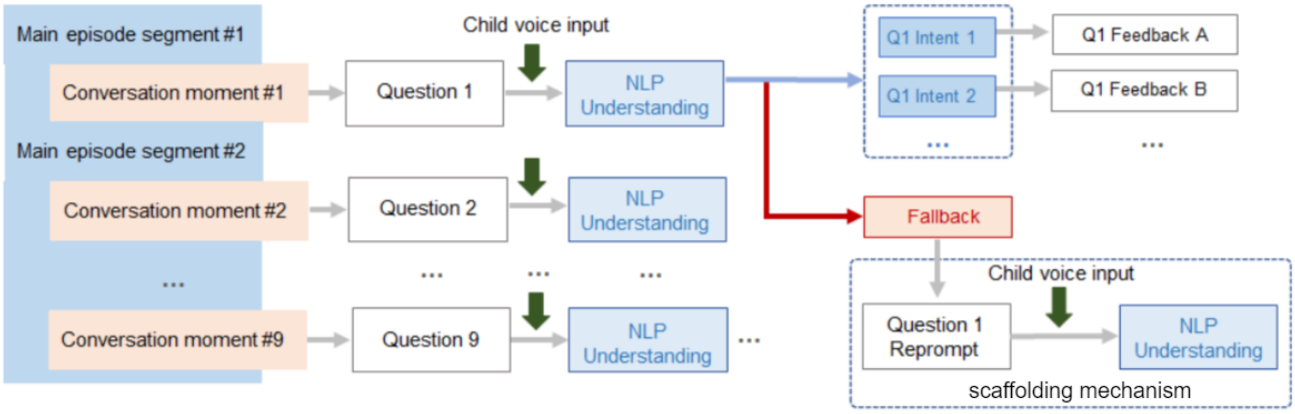

Figure 3: Nine conversation moments were inserted throughout the main episode, splitting the main episodes into multiple segments. Dialogue flow of the one question was displayed.

purpose of this testing was to ensure that Elinor's language was clear to children. We made minor revisions to the scripts based on this testing (e.g., to better contextualize Elinor's question involving a magnifying glass and make it easier for children to understand, we changed *"What is this tool used for?"* to "*How does this tool help me observe the snake"*).

3.3.2 Voice-over and animation development. After the conversation scripts were finalized, they were sent to the *Elinor Wonders Why* production studio for voice-over and animation. We tested two different types of visual backgrounds for the conversation moments with Elinor (see Figure 4). The first type (one the left) involves backgrounds that transition seamlessly to the existing episode, with the advantage being a strong perception that the viewer was experiencing story events together with Elinor in her world. The second type (one the right) involves using the same background with question marks for all conversation moments, with the advantage being the opportunity to transport viewers to a "question world" where they expect to interact with Elinor. Testing these two options with children suggested that the second approach was more suitable. Children were less likely to miss conversation moments when the "question world" background was used. In addition, using the same background for each conversation moment in each episode is more cost effective, especially for large-scale development.

Figure 4: Two different background designs for the conversational moments.

3.3.3 System development and field testing. After finalizing the conversation scripts and animation files, we began developing the mobile application consisting of three conversational videos. After prototyping the application, the research team tested the conversation moments using the actual cloud-based natural language processing engine to identify areas needing further refinement (e.g., adding more training phrases to increase the agent's accuracy). We then conducted a small field test with ten children and made additional refinements along with improvement to usability (e.g., adding a sound effect to prompt children to answer questions).

4 USABILITY STUDY

After developing and testing the application, we conducted a usability study with children who watched the conversational videos in their own homes. This usability study involved two episodes, one on molting and the other on aerodynamics, and the remaining episode on liquid viscosity was used in a randomized experiment reported in the next section. Our usability focused on four questions:

(1) How do children respond to Elinor's questions?
(2) How does the conversational agent underlying Elinor respond to children's answers?
(3) How do children perceive these conversational videos?
(4) How do parents perceive these conversational videos?

Children were instructed to watch the conversational episodes in their home using a laptop we provided. Additionally, the family used their own device to join a video conference that allowed an experimenter to observe the child's interaction with Elinor and then asked interview questions after the episode's completion. We scheduled a separate video conference with the parents for an interview regarding their child's experience interacting with the conversational videos. This procedure was pre-approved by the IRB of the authors' institution, and the participants were compensated $50 for their participation. Figure 5 shows a child interacting with Elinor during the video watching.

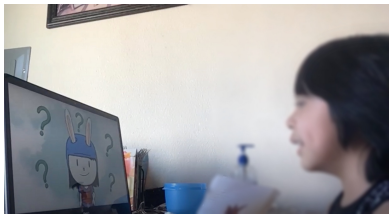

Figure 5: A child participant interacting with Elinor in the usability testing.

4.1 Participants

Twenty participants aged four through six were recruited through snowball sampling with the assistance of a community non-profit organization serving local families with preschool-aged children. The average age of the participants was 5.1 years, and 13 of the 20 participants were girls. Half of the participants predominantly spoke English at home, and the rest spoke another language, mainly Spanish. The majority of participants (n = 14) identified as Latino, while two identified as White, one as Asian, and three as more than one ethnicity. Half of the participants' parents had a Bachelor's degree or higher. Eight participants reported that they had never used voice assistants before, while six used them monthly, and another six used them more frequently (i.e., weekly or daily).

4.2 Evaluation Metrics

4.2.1 Child Response. We coded three items to capture the quantity and quality of children's responses to Elinor's questions during the video watching. For the response quantity, we coded both whether a child verbally responded to the prompt and the total number of words in each of the child's responses. For the response quality, we coded the topical relevance of each response into one of three categories: an on-topic response that directly answered Elinor's question, an off-topic response that did not relate to the question, and meta-comments that describe the child's thought process (e.g., *"I don't know."* or *"I'm still thinking."*). This coding was conducted by two trained research assistants, with one coding the video footage from all participants and the other coding 30% of the participants as a quality check. The agreement rate between these two coders was 100% for all items.

In addition, we also took detailed notes for each conversation moment. We documented children's behaviors, including visual orientation (i.e., looking at the screen, looking at family members, or looking somewhere else), facial expressions (e.g., smiling, laughing, or frowning), and non-verbal expressions (e.g., gestures like nodding and shaking their head). We also took note of the participants' environment (e.g., whether there was excessive noise or distractions from siblings). Five coders were involved in taking these field notes. Over the span of ten weeks, the research team discussed the notes and reviewed the video footage to ensure the notes taken reflected the actual interactions.

4.2.2 Agent Performance. The agent's performance was evaluated by comparing the record of the agent's speech-to-text translation and intent classification with a transcription of interactions completed by two trained human coders. We calculated the accuracy rate of the agent's speech-to-text translation and intent classification, using human coding as the benchmark.

4.2.3 Child Perception. We used a survey to elicit children's perceptions of the conversational videos and Elinor. This survey included four dimensions: children's enjoyment of the conversational videos (four items, adapted from [18]), children's trust in Elinor (three items, adapted from [69]), children's social attitude toward Elinor (four items, adapted from [69]), and children's perception of Elinor's capability (five items, adapted from [94]). For all items, children were first asked to indicate whether they agree with a statement (yes/no) and then asked to clarify the magnitude to which they agree or disagree (a bit/definitely), leading to four possible ordinal responses: definitely no, a bit no, a bit yes, and definitely yes.

4.2.4 Parent Perception. Sometime after the child's usability session, we conducted a semi-structured interview with the child's parents. The interview asked about parents' perceptions of the learning benefits and limitations of the conversational videos. We also asked parents to share their experience of the usability session (e.g., whether they needed to assist their children as they watched the conversational videos). Lastly, we elicited parents' suggestions for improvements to future iterations of the conversational videos. The interview was audio recorded and then transcribed verbatim.

We used an inductive process to analyse the transcribed interview. We began with qualitative memoing, in which members of our research team viewed the same portion of the transcript together and then individually wrote their own notes. Every five minutes, researchers paused and discussed with one another the themes that emerged from the data in terms of the parents' perceived benefits and limitations, usability, and suggestions for improving the conversational videos. After reviewing all the transcripts, we systematically coded all the interview based on the themes we generated. Coding was periodically cross-checked by two coders to ensure accuracy.

4.3 Results from the Usability Testing

4.3.1 Children's Response. On average, children verbally responded to 92.8% of Elinor's questions, with average length of verbal responses being 4.6 words (*SD* = 5.2 words). Most of the children's responses were categorized as a direct, on-topic response (87.8%), a smaller percentage (11.6%) was categorized as meta comments, and less than 1% were categorized as off-topic. Children's response patterns were consistent across the two episodes, and the breakdown of responses by episodes is presented in Table 1.

Recall that Elinor asked children reprompt questions (i.e., scaffolding with hints and rephrased questions) if children either responded with an unanticipated answer, indicated they did not know the answer, or did not verbally respond at all. On average, reprompt questions were triggered among 22% of the conversation moments, which equated to each child receiving about two reprompt questions per episode (each episode contains about ten conversation moments). Forty-eight percent of reprompt questions were triggered by children's unanticipated responses, 21% by the agent mistranslating or classifying children's originally valid responses, 18% by children not verbally responding, and 12% by children indicating they did not know the answer. Among the original-reprompt

Table 1: Breakdown of children's responses by episode.

			Relevancy		
	Response rate	**Average length**	**On-topic**	**Off-topic**	**Meta-comments**
Overall	92.8%	4.64 (5.23)	87.8%	0.6%	11.6%
Molting	93.3%	4.88 (5.74)	86.4%	0.3%	13.2%
Aerodynamics	92.4%	4.38 (4.56)	89.5%	0.9%	9.6%

question pairs, only 65% of children's responses to initial questions were categorized as on-topic, while 85% of responses to scaffolding questions were on-topic. Here are several examples of how the re-prompt questions had helped children generate on-topic responses. In one example, when Elinor initially asked, *"Do boxy things or pointy things move through the air slower?"* a child simply repeated the question by answering *"slower, slower."* Elinor then followed up by asking, *"Think about what my dad just said! The air pushes against boxy things, but moves around pointy things. So which one does the air slow down more, boxy things or pointy things?"* This time the child answered the question directly, *"slow down...boxy."* Another child initially responded to Elinor's question about how magnifying glass helped by saying *"because it is made out of glass,"* and then changed his answer to *"make things bigger"* after Elinor rephrased the question.

We then examined the transcriptions of children's actual responses. While the majority of the children's responses were succinct phrases with five words or fewer, a considerable portion of the responses were complete sentences with more than ten words. These longer answers typically contained details beyond what the question intended to elicit, commonly with children's spontaneous reasoning of their answers. For example, in response to the question *"Do you think the snake will move faster?"*, a five-year-old girl first answered *"maybe faster"* and then articulated her reasoning by adding *"because the old skin is so tight but now that old skin taked (taken) off the new skin can move now."* Similarly, a six-year-old girl was prompted to predict whether Elinor's car would go faster after they painted the car yellow. The girl replied that it would not *"because it doesn't have any engine and painting it just makes it look different but it's not actually faster."* These kinds of elaborate responses suggest the usefulness of prioritizing open-ended questions that leave room for children's free expression.

Another interesting phenomenon we observed was children using gestures alongside their verbal responses when communicating with Elinor. The most typical gestures were nodding their head while saying yes and shaking their head while saying no; this happened about 50% of the time when children responded *"yes"* or *"no"*. Children also frequently used gestures to describe motions and shapes, as a supplemental visualization to their verbal response. For example, a six-year-old girl moved her hands to gesture air going around an object as she responded to Elinor, *"[the car] is pointy so the air goes around it."* Another six-year-old girl drew a zigzag with her forefinger in the air as she said *"[the snake] has some stripes that are like this."* Interestingly, however, according to the post-viewing interviews, most of the children, including those who frequently utilized non-verbal expressions, accurately perceived that Elinor could not actually see them. This is consistent with the Media Equation Theory which suggests that people, including children, automatically apply real-world behavioral norms to their interactions with technology in mediated worlds [27].

4.3.2 Agent Performance. The agent was able to satisfactorily decipher and interpret children's responses and initiate its feedback with appropriate timing. The speech-to-text translation accuracy rate was 81%, meaning that the agent correctly translated about eight out of every ten words spoken by a child. Common factors that had led to errors in speech-to-text translation included: children initiating responses before Elinor finished her question; multiple family members attempting to respond to the agent at the same time; and children's fuzzy pronunciation. The intent classification accuracy rate was higher at 89%, meaning that the agent was able to accurately map almost nine out of ten responses, on average, to the correct intent and thus give correct feedback. The intent classification accuracy rate was higher because the language processing model used semantic-based understanding, and thus errors in specific words did not necessarily influence the overall meaning of a given response. For example, the meaning is similar between a child's actual response *"that's too tight"* and the response *"that just too tight"* as translated by the machine. Note that in all cases, the intent classification errors occurred when the agent classified a child's valid response as "fallback" and provided feedback that was generic but not inappropriate (i.e., the agent had not interpreted a child's incorrect answer as correct and followed with feedback that praised the answer). The agent's performance deciphering and interpreting children's responses was consistent with the state-of-the-art natural language processing models reported in other studies focusing on children in this age range ([17]).

4.3.3 Child Perception. Overall, children reported positive perceptions of the conversational videos. All but one of the items received an average perception score between 3 (positive) and 4 (very positive). The average score of the four items in the enjoyment dimension was 3.5, indicating that children generally enjoyed watching the videos (3.6) and would like to do it again (3.3), felt that the video watching experience was interactive (3.5), and believed that they learned new things from the experience (3.5). The average score of the three trustworthiness items was 3.4. Most children reported that they believed what Elinor said in the video (3.3), that they thought Elinor made good choices (3.6), and that Elinor was a good scientist (3.3). The social dimension received an average score of 3.7 across four items. Children believed that Elinor was friendly (3.9) and wanted to make friends with them (3.6), that Elinor would feel upset if they did not help her solve problems when she asked for help (3.7), and that Elinor would also help them solve a problem if they had one (3.7). In terms of the capacity dimension, the average

score across five items was 3.3. Children perceived that Elinor had the ability to hear (3.3), to understand (3.7), to remember (3.2), and to solve problems (3.3). The only item that received an average score below 3 was Elinor's perceived ability to see (2.7).

Children sometimes spontaneously brought up their rationale of the choice. Children commonly referred to their interactions with Elinor and to her contingent feedback when explaining their positive perceptions. For example, when asked why Elinor wanted to make friends with them, one child answered *"because I had helped her, and she was happy."* and another child said, *"Elinor's talking to me!"* Another child explained why they thought Elinor was good at solving problems, indicating that they were initially unable to help Elinor solve a problem but that Elinor soon figured it out and shared the solution with them (this occurred during Elinor's feedback to the child's response).

4.3.4 Parent Perception. Parents' reports of their children's experience with the program were overwhelmingly positive. Not only did all parents share their thoughts on the potential for this program to aid in their children's science learning and vocabulary development, but they also mentioned that this program is helpful for parents to learn how to interact with their children surrounding science topics, and also potentially use this program as an educational enrichment for their children as they work or complete chores around the house.

In terms of the usability of the conversational videos, all parents described that the program was easy for children to use and most children were capable of using it alone, if necessary. The parents mentioned that their child's experiences with other digital apps contributed to their smooth interaction with the conversational videos. For example, a parent of a 5-year-old child shared *"It wasn't hard at all. She uses the PBS KIDS app to play games, so she is very familiar."* Parents also suggested that their child had become proficient in navigating interactive digital devices due their experiences of online schooling during the pandemic. For example a parent told us that *"In this past year, in the pandemic, the kids all had to go online and she became very computer savvy."* All parents reported that they have access to smart devices and stable Internet connection in order to use the conversational videos.

In terms of benefits, multiple parents recognized and appreciated the potential of this program to help in their child's learning of science topics. A parent of a 4-year-old reported *"I 100% think this can help her learn science, especially because I'm not a science or math person. I think this is an interactive, fun way to learn while participating in a story versus just having to discuss the terms."* Parents shared that their children are using higher vocabulary after their participation with this program. A parent of a 5-year-old communicated that they *"really enjoyed that she's using the vocabulary. She's not just saying 'the bees are eating honey', she'll say, like, 'the bees are drinking the nectar from the flower.'" These interactions may also spark children's curiosity in other languages, as one parent reported "she'll use the words in Spanish, so she can translate it, or she'll ask me 'Oh, how do you say it in Spanish?'"* One parent of a 4-year-old described how this opportunity attuned her to her daughter's potential science interest, stating *"I don't know what might capture her interest or what she might pick up on at this age. Before, my first thought might have been to take her to the park, versus now I realize she would enjoy visiting a science or discovery center. Maybe we should start doing that."*

Although parents indicated that children were able to use the conversational videos with minimal parental supervision, some parents preferred to co-view the videos with their child. Two parents of 4-year-olds commented that their child could benefit from a co-viewer encouraging them to pay attention and reprompt them to respond to Elinor throughout the program. Moreover, two parents suggested that our team design post-video watching activities and projects, which will further facilitate parents' involvement in children's science learning and will also reinforce the learned concepts. While parents communicated that this show has the potential to encourage parents to build more knowledge with their children, and have more normal conversations surrounding science topics, one parent shared that *"there's nothing like the real experience of building a car. As we observe it on TV, we learn about it, and plant a little seed of curiosity, but we would love to actually try it at home."*

Some parents suggested that we incorporate more small talk moments between Elinor and child viewers to get them familiar with Elinor. Two parents noticed differences in how their younger (4-year-old) and older (6-year-old) children participated in the conversation based on how familiar they were with the main character. One parent reported that her 6-year-old daughter was familiar with the cartoon *Elinor Wonders Why* and, *"with the many interesting facts she was fine learning right away, but for him (4-year-old) I think it was a little bit slower to ease him into really being engaged into the cartoon."* Another parent suggested beginning the conversational episode by having Elinor introduce herself and spend some time familiarizing children with the interaction. *"Elinor can start by asking them questions like 'What's your name?' 'How old are you?' or 'Do you like Mickey Mouse?' just to get them familiar and comfortable with her,"* they suggested.

5 RANDOMIZED STUDY TO ASSESS EFFECTIVENESS

The usability study suggested that the conversational videos could feasibly be used by children at home. We conducted a randomized study to further understand the added benefits of these conversational videos compared to the standard version of *Elinor Wonders Why* that is currently available on the network. In this study, children were randomly assigned either to the conversational video condition in which children had contingent interaction with Elinor as they watched the episode on liquid viscosity or to the control condition in which children watched the standard version of the show, the same episode but without the opportunity for contingent interaction with the character. This experiment setup would also provide the show's producers with evidence of conversational interactivity's benefits, as compared to the show's current format. We focused on two questions in this study:

(1) Does a conversational video improve children's learning of science concepts?
(2) Does a conversational video increase children's engagement?

5.1 Procedure

The randomized study included two sessions that were scheduled one week apart. In the first session, children's English language

proficiency was assessed using a computer-based assessment (i.e., Quick Interactive Language Screener [55]). In the second session, children watched one episode on liquid viscosity with or without conversational agents. Children were then asked questions to assess how much they learned from the show. Both of the study sessions were carried out remotely; children participated in the study from their home and communicated with the experimenter via video conferencing. The entire study session was video recorded. Similar to the usability study, participant used a laptop we provided for watching the episode and used their own device for the video conference.

5.2 Participants

Seventy-seven children, different from those in the usability study but drawn from the same recruitment method, completed the study. Among these children, there are 21 four-year-olds, 28 five-year-olds, and 28 six-year-olds. Forty-nine children were Latino (63.6%), 12 were White (15.6%), and 16 were Asian or mixed race (20.8%). Fifty-five of the children (71.4%) spoke predominantly English at home, whereas 22 spoke other languages at home, including Spanish, Chinese, and Japanese (28.6%). Forty-nine were girls (63.6%). Twenty-eight children were reported to have more than one monthly experience talking with smart speakers or other voice assistants on smartphones, and the rest of the forty-nine children never or rarely had such experience. Table 2 presents the participant information.

5.3 Evaluation Metrics

5.3.1 Immediate Assessment of Episode-Specific Science Learning. To assess children's learning from the episode, the research team developed a 10-item questionnaire (different from the conversational prompts embedded in the episode) that was aligned with the NGSS and the US Department of Education's Ready to Learn Science Framework. We vetted the items on the questionnaire with an advisory board of science curriculum consultants from the Ready to Learn program and other prominent experts on young children's science learning. The questionnaire assessed children's problem solving skills and their understanding of vocabulary and science facts introduced in the episode. The actual assessment items are displayed in Appendix B.

For ten out of the twelve questions, we first asked children to freely formulate the answer, and if children were not able to generate the correct answer, we prompted them with three options to choose from. Children received a score of 2 if they answered correctly without prompting, a score of 1 if they required prompting to answer correctly, and a score of 0 if they could not answer correctly. For the remaining two questions that required children to provide explanations of their answers, we scored their answers from 0 to 4 points. A score of 0 indicated an answer was completely incorrect; a score of 1 indicated an answer was incorrect but included some correct ideas; a score of 2 indicated an answer was almost correct but the language was inappropriate; a score of 3 indicated a correct answer with proper language; and a score of 4 indicated a correct answer with additional correct details to support the answer. Based on this scoring system, we calculated a total score by summing the points across all items, with a possible range from 0 to 28. The Cronbach's alpha of the learning outcome items was 0.81.

5.3.2 Engagement. The global scale was based on coders' broader holistic assessments of each child's engagement. For each time segment, we provided a 5-point rating based on a child's posture, facial expression, eye gaze, distractibility, verbal and nonverbal comments, and responsiveness to the adult or agent's direction ([48]). A score of 5 indicated the highest level of engagement (e.g., showing clear signs of excitement that stems from the video, making large movements with hands to illustrate a point). A score of 3 indicated a medium level of engagement where a child did the minimum work required to follow protocols (e.g., remaining seated). A score of 1 was the lowest level of engagement where a child was clearly distracted and had little interest in the video. An average global engagement rating was calculated by the mean of the ratings across all time segments in each child's reading session. The IRR calculated by Intraclass Correlation was 0.82 for this global coding.

5.4 Results from the Randomized Study

5.4.1 Effects on Immediate Science Assessment. The maximum score of the science assessment was 28 points, and children in our sample got an average score of 14.8 points (standard deviation of 6.0, slightly over half of the full score; see Table 3). As shown in Table 3, children in the experimental group outperformed those in the control group by 2.5 points, which equates to correctly answering one more question (out of 12 questions) via free recall. ANCOVA analysis controlling for children's age, English language proficiency, and prior conversational agent usage suggested that children who watched the conversational video scored significantly higher than those who watched the broadcast video, $F(1, 70) = 7.86$, $p < 0.01$. The eta squared η^2 effect size was 0.05, which was at the medium range.

5.4.2 Effects on Engagement. Children across the two conditions were scored 3.01 in the engagement rating, which was very close to the neutral engagement state. This suggested that, on average, they were able to stay on task and comply with the study procedure. Children who watched the conversational video received an average rating of 3.04 ($SD = 0.15$), which was higher than those in the broadcast group ($M = 2.98$, $SD = 0.15$). ANCOVA analysis controlling for age, language proficiency, and prior experience with conversational agents failed to suggest a significant benefit of conversational videos on engagement, $F(1,70) = 3.22$, $p = 0.07$, although the effect size was moderate, $\eta^2 = 0.04$.

6 DISCUSSION

This paper presents the design and evaluation of conversational videos adapted for a popular science narrative program, in which children are allowed to interact with the program character by answering questions and receiving feedback. Our usability study showed that the conversational videos could feasibly be used by young children in their homes, and a separate randomized study suggested that children's science learning was improved from their in-the-moment interactions with the program character. In this section, we first discuss our findings in relation to existing research. We then discuss design implications that could inform future developments. Finally, we discuss some potential limitations of our conversational videos and future research agenda.

Table 2: Participant information by study conditions.

	Full Sample	Conversational	Standard	Difference
QUILS	64.97 (30.78)	66.47 (30.99)	63.43 (30.90)	$t(77) = 0.43$ $p = 0.67$
Female	63.64%	66.67%	60.53%	χ^2 (1) = 0.10, $p = 0.74$
Age				$\chi^2(2) = 0.99$, $p = 0.61$
4-year-olds	27.27%	23.08%	31.58%	
5-year-olds	36.36%	35.90%	36.84%	
6-year-olds	36.36%	41.03%	31.58%	
Race/Ethnicity				χ^2 (2) = 1.59, $p = 0.45$
White	15.58%	25.51%	10.52%	
Latino	63.64%	61.54%	65.79%	
Others	20.78%	17.95%	23.68%	
Predominant Home Language				χ^2 (1) = 0.11, $p = 0.74$
English	71.43%	74.36%	68.42%	
Other	28.57%	25.64%	31.58%	
Prior Voice Assistant Usage				χ^2 (1) = 0.00, $p = 0.95$
Heavy users	36.36%	38.46%	34.21%	
Non-heavy users	63.64%	61.54%	65.79%	
Mother's Education				χ^2 (2) = 3.26, $p = 0.20$
Less than high school	14.29%	20.51%	7.89%	
Above high school	20.78%	15.38%	26.32%	
Above Bachelor's degree	64.94%	64.10%	65.79%	
Typical TV Time During Weekdays				χ^2 (2) = 0.49, $p = 0.78$
Less than 30 minutes	41.56%	43.59%	39.47%	
30-60 minutes	49.35%	46.15%	52.63%	
More than 60 minutes	9.09%	10.26%	7.89%	
N	77	39	38	

Table 3: Outcomes by study conditions.

	Full sample	Conversational	Standard	ANCOVA
Science Learning	14.80 (6.02)	16.00 (5.26)	13.5 (6.54)	F(1,70) = 7.86 **
Engagement	3.01 (0.15)	3.04 (0.15)	2.98 (0.15)	F(1,70) = 3.22†

6.1 The Promise of Conversational Narrative Programming

In general, our conversational videos were able to fulfill our intended design goal of building two-way interactivity into television or video content. The conversational videos elicited a high level of verbal engagement from children, as children responded to almost all of Elinor's questions. In turn, Elinor's replies to children were based on their particular responses to her questions, leading to a personalized video watching experience for each child. Thus, our study presents a case for AI-powered conversational characters unlocking the potential inherent in currently non-interactive narrative television programming. Our usability study also suggested that conversational videos were perceived positively by children and their parents. The children in our study generally agreed that watching the videos was enjoyable and that Elinor was trustworthy, sociable, and intelligent. They also pointed to their interactions with Elinor when explaining their perceptions. While our study was based on one study session, it is possible that children's sustained interaction with Elinor over longer periods of time could result in them forming parasocial relationships, one-sided emotional attachments with media characters that could foster more attentive and motivated learning [10, 44]. The parents in our study viewed the children's interactions with Elinor positively, and they repeatedly mentioned the benefits of science learning and language development. The parents' appreciation of our conversational videos and their interactivity contrasts markedly with their perception of standard television programming as passive [49].

Our study also provided preliminary evidence for the benefits automated conversational interactivity can have on children's learning. We found that children who watched the conversational video performed better in an episode-specific science assessment than did those who watched the broadcast version. This result is consistent with other studies on dialogic interactions during children's television watching (e.g., [75, 80, 81]), which prior to this study have relied on direct human involvement in the interaction (e.g., a Wizard of Oz approach or a human co-viewer). For example, Calvert and colleagues used a Wizard of Oz approach and found

that children learned math concepts presented in an animated video program significantly better when the video's main character asked children questions and replied in a timely and responsive manner as compared to when they watched the same video without this interaction [9]. Our study, however, involved a fully automatic conversational agent that could readily be implemented in children's television shows.

While we expected that the conversational interactivity would enhance children's overall engagement, we did not find a statistically significant difference in this regard between children who interacted with Elinor and those who watched the standard version without such interactions. Yet prior studies, for example, found that that when an on-screen actor responded contingently to children through a live video feed, children were more likely to respond to prompts than those viewing a pre-recorded video without live video chat [81]. The difference between our study and prior studies may be due to the different approaches to measuring engagement. To gain a more nuanced understanding of these results, we conducted a follow-up analysis that differentiated between the various engagement indicators in our initial holistic coding scheme. We found that children in the conversational video group showed a much higher level of verbal engagement (consistent with [81] described above) and positive affect (consistent with [54]), despite spending less time looking at the screen compared to children who watched the broadcast video. One possible explanation for the children's lower visual attention levels is the fact that children who interacted with Elinor tended to look away from the screen during the conversational moments (i.e., they looked toward a nearby family member or simply raised their head as they contemplated their response to Elinor's questions). Nevertheless, in the following section, we discuss several ways that future iterations of conversational videos could more effectively engage children.

6.2 Some Design Considerations

In this section, we discuss some design implications of our study. Our team has already begun improvements to our conversational videos based on the considerations below.

6.2.1 Supporting multimodal interaction. In our testings, we commonly saw children, particularly the youngest participants, using non-verbal expressions to supplement, and sometimes replace, their verbal expressions. However, the conversational agent registered children's speech only. Prior to the video watching sessions in the current study, we told children about Elinor's inability to see them and encouraged the children to respond only verbally. This might have also been reflected in children perceiving Elinor as not being able to see in the perception test. Yet in retrospect, we believe it would be better to instead design a conversational agent that could also register non-verbal expressions (i.e., gestures). Studies have suggested that multimodal interaction is more natural and beneficial for children who are still developing their communication skills. For example, Crowder and Newman found that when children explain science concepts (e.g., seasonal changes), they often use gestures to enhance the ideas they express through speech [16]. Another study found that encouraging children to use gestures to explain novel concepts solidified their learning as compared to children who were not taught to use gestures [23].

Thus, our team has embarked on follow-up research to combine natural language processing with gesture recognition. As a starting point, we have begun incorporating models to recognize several gestures, including children nodding or shaking their head and using their fingers to show the numbers one through nine. Though this work is still preliminary and involves many challenges, we believe it is a promising direction that is worth exploring.

6.2.2 Enabling bilingual input. In the current study, our conversational character was trained to recognize only English responses. However, we occasionally observed that some children responded to Elinor in their home language (e.g., one child said "aqua" instead of "water"). Even though "water" was the correct answer, Elinor interpreted the child's Spanish response as an unanticipated answer and replied with the relevant fallback mechanism. We recognize that it is common for bilingual children to switch between their home language and English, particularly in their home environment [12]. In addition, we noticed in our testing that bilingual children were less likely to respond to Elinor and that Elinor was less successful in interpreting their responses. As such, we have begun investigating the possibility of allowing for multilingual speech input in the next iteration of our conversational videos. Specifically, because our project currently involves a large number of Spanish-English bilingual children, we have refined the agent's natural language processing model to include Spanish responses. Including multilingual conversational characters could potentially make interactive science learning videos more linguistically relevant to a wider range of children. Other studies have also recognized the importance of developing educational media that meets the needs of bilingual children (e.g., [39]).

6.2.3 Allowing Multi-user Participation. Our conversational videos were initially designed for individual users, as this is the most common way children watch television [2]. However, during our testing, we observed that some children's siblings—who were not actually watching the episode but could hear it—would also respond to Elinor's questions. The overlapping speech from the child user and their sibling interfered with the agent's ability to process the speech input. While we might be able to prevent this situation by instructing children to view the conversational video in a space where they will not be disturbed, this is not an optimal solution as children generally benefit from social engagement during media usage (for an overview, see [20]). As such, we should proactively design the agent so that multiple users can participate together. For example, it is possible to support "trialogue" such that a child and their sibling can take turns responding to the agent. The agent could invite one child to respond to a question, and then invite the second child to comment on the first child's response. It is also possible to design the agent so that it can utilize voice recognition to automatically distinguish responses from different users.

6.2.4 Allowing local language processing. In addition, in our current project, the conversational agent was hosted on a cloud service and the users needed to connect their viewing device to the Internet. During our testing, two of our participants' Internet connections were unstable when the agent was processing their speech input through the cloud service, thus leading to an error in Elinor's interpretation of their responses. In this current iteration, we resorted to

hard-coding the conversational agent to respond with generic feedback (the same as when a child provides an unanticipated answer) whenever a user's Internet connection becomes unstable. However, with the rapid development of low-cost on-device computing, it is plausible to use local language processors to carry out speech recognition and intent classification. This would allow users to preload the entire application (i.e., the conversational episodes and language models) and view them with or without Internet connection, giving users more flexibility. This might be particularly relevant to under-resourced households who do not have stable Internet connection.

6.3 Current Limitations and Agenda for Future Research

In this section, we discuss several limitations of this current paper and our agenda for possible future research.

First, in the randomized study, we assessed children's science learning outcomes immediately after their video watching. Thus, our findings did not shed light on children's retention of science knowledge. Future studies may want to include a delayed post-test to address this issue. Second, our randomized study used the standard broadcast version of the *Elinor Wonders Why* episodes as a baseline control group for comparing our conversational videos. The standard episodes do not include any pseudo-interactive features. Although we identified benefits in children's science assessment and engagement from conversational videos, it is possible that such benefits were simply due to Elinor asking questions rather than Elinor's ability to respond contingently. Indeed, our team had planned to include an additional pseudo-interactive condition where Elinor asked children questions but did not provide contingent feedback. However, research restrictions due to the COVID-19 pandemic limited our sample size and conditions. As a practical consideration, one important purpose of our randomized study was to examine the effectiveness of adding conversational interactivity to *Elinor Wonders Why* episodes. The study results would form the basis for our decision whether to expand the development of conversational videos and eventually integrate them into PBS KIDS' user-facing platforms. Following the lifting of pandemic-related restrictions on research, our team resumed in-person research with children, and we are now replicating this study, including the pseudo-interactive condition, with a larger sample size.

Lastly, children in our study watched a small number of conversational videos (i.e., two videos in the usability study and one video in the randomized study) within a short period of time. As such, our results could not shed light on children's long-term interaction patterns or benefits. It is possible that as the novelty of talking to the character wears off (e.g., after children interacting with the same video for multiple times), children may come to expect having more complex conversations beyond how the character's dialogue is designed. When the character does not meet children's such expectations, children may become less willing to continue the interaction, and this decreased engagement may dampen learning benefits. On the other hand, children's interaction with the intelligent character may become more smooth and productive as children get familiarized with the schema of such interaction, thus leading to heightened engagement and improved learning outcomes. Future research should examine this issue by using a longitudinal design where children are provided with sustained access to this type of conversational videos.

6.4 A Note on Ethical Considerations

It is worth discussing potential ethical issues surrounding children's use of conversational AI. First, we should be mindful of the privacy issues that might emerge as some technology companies may store user audio recordings without explicit acknowledgement. This may contravene the US Children's Online Privacy Protection Act (COPPA), which was enacted to regulate the collection and use of personal information from anyone younger than 13. As such, when developing conversational applications for children, it is important to ensure children's voice utterances are deleted immediately (e.g., the data protection option provided by Amazon Lex). Second, many fear that conversational AI might supplant the interpersonal interaction children would otherwise have in their day-to-day lives, since this technology can potentially simulate a conversation partner for children. However, we want to be clear that researchers should develop conversational applications so as to provide additional opportunities to enrich children's language experiences. Conversational applications should not be designed to replace parents, teachers, or peers. Indeed, some of our upcoming work has focused on using conversational AI to promote parent-child interactions.

7 CONCLUSION

As the time that children spend watching video increases and the mode of watching shifts to Internet-connected devices, it is imperative to investigate how new forms of video watching may better support learning. This study leveraged conversational technologies to allow children to interact with the main character of a science animated program. Our findings suggest that enabling this kind of contingent interaction between child viewers and media characters can bring additional educational benefits not available through standard video programming. Though this line of research is just beginning, we believe that conversational AI has the potential to transform traditional video watching into a more active and engaging learning experience.

ACKNOWLEDGMENTS

This research is based upon work supported by the National Science Foundation under Grant No. 1906321 and No. 2115382. Production of the conversational episodes is supported by the Corporation for Public Broadcasting. Funding for *Elinor Wonders Why* is provided in part by a Ready To Learn grant from the U.S. Department of Education [PR/Award No. U295A150003, CFDA No. 84.295A], and by the Corporation for Public Broadcasting.

REFERENCES

[1] Fashina Aladé and Amy I Nathanson. 2016. What preschoolers bring to the show: The relation between viewer characteristics and children's learning from educational television. *Media Psychology* 19, 3 (2016), 406–430.

[2] Daniel R Anderson and Katherine G Hanson. 2017. Screen media and parent–child interactions. In *Media exposure during infancy and early childhood*. Springer, 173–194.

[3] Paul Baxter, Emily Ashurst, Robin Read, James Kennedy, and Tony Belpaeme. 2017. Robot education peers in a situated primary school study: Personalisation promotes child learning. *PloS one* 12, 5 (2017), e0178126.

[4] Erin Beneteau, Olivia K Richards, Mingrui Zhang, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication breakdowns between families and Alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[5] James Alex Bonus and Marie-Louise Mares. 2018. When the sun sings science, are children left in the dark? Representations of science in children's television and their effects on children's learning. *Human Communication Research* 44, 4 (2018), 449–472.

[6] Cynthia Breazeal, Paul L Harris, David DeSteno, Jacqueline M Kory Westlund, Leah Dickens, and Sooyeon Jeong. 2016. Young children treat robots as informants. *Topics in cognitive science* 8, 2 (2016), 481–491.

[7] Rodger W Bybee. 2014. NGSS and the next generation of science teachers. *Journal of science teacher education* 25, 2 (2014), 211–221.

[8] Sandra L Calvert, Marisa M Putnam, Naomi R Aguiar, Rebecca M Ryan, Charlotte A Wright, Yi Hui Angella Liu, and Evan Barba. 2019. Young Children's Mathematical Learning From Intelligent Characters. *Child Development* (2019).

[9] Sandra L Calvert, Marisa M Putnam, Naomi R Aguiar, Rebecca M Ryan, Charlotte A Wright, Yi Hui Angella Liu, and Evan Barba. 2020. Young children's mathematical learning from intelligent characters. *Child development* 91, 5 (2020), 1491–1508.

[10] Sandra L Calvert, Melissa N Richards, A Jordon, and D Romer. 2014. Children's parasocial relationships. *Media and the well-being of children and adolescents* (2014), 187–200.

[11] Sandra L Calvert, Bonnie L Strong, Eliza L Jacobs, and Emily E Conger. 2007. Interaction and participation for young Hispanic and Caucasian girls' and boys' learning of media content. *Media Psychology* 9, 2 (2007), 431–445.

[12] Katja F Cantone. 2007. *Code-switching in bilingual children*. Vol. 296. Springer.

[13] Elizabeth J Carter, Jennifer Hyde, and Jessica K Hodgins. 2017. Investigating the Effects of Interactive Features for Preschool Television Programming. In *Proceedings of the 2017 Conference on Interaction Design and Children*. 97–106.

[14] Justine Cassell and Kimiko Ryokai. 2001. Making space for voice: Technologies to support children's fantasy and storytelling. *Personal and ubiquitous computing* 5, 3 (2001), 169–190.

[15] Yi Cheng, Kate Yen, Yeqi Chen, Sijin Chen, and Alexis Hiniker. 2018. Why doesn't it work? voice-driven interfaces and young children's communication repair strategies. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. 337–348.

[16] Elaine M Crowder and Denis Newman. 1993. Telling what they know: The role of gesture and language in children's science explanations. *Pragmatics & Cognition* 1, 2 (1993), 341–376.

[17] Griffin Dietz, Jimmy K Le, Nadin Tamer, Jenny Han, Hyowon Gweon, Elizabeth L Murnane, and James A Landay. 2021. StoryCoder: Teaching Computational Thinking Concepts Through Storytelling in a Voice-Guided App for Children. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–15.

[18] Griffin Dietz, Zachary Pease, Brenna McNally, and Elizabeth Foss. 2020. Giggle gauge: a self-report instrument for evaluating children's engagement with technology. In *Proceedings of the Interaction Design and Children Conference*. 614–623.

[19] Steffi Domagk, Ruth N Schwartz, and Jan L Plass. 2010. Interactivity in multimedia learning: An integrated model. *Computers in Human Behavior* 26, 5 (2010), 1024–1033.

[20] Rebecca A Dore and Laura Zimmermann. 2020. Coviewing, Scaffolding, and Children's Media Comprehension. *The International encyclopedia of media psychology* (2020), 1–8.

[21] Stefania Druga, Randi Williams, Hae Won Park, and Cynthia Breazeal. 2018. How smart are the smart toys? children and parents' agent interaction and intelligence attribution. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. 231–240.

[22] Kiran Egan. 1993. Narrative and learning: A voyage of implications. *Linguistics and education* 5, 2 (1993), 119–126.

[23] Stacy B Ehrlich, Susan C Levine, and Susan Goldin-Meadow. 2006. The importance of gesture in children's spatial reasoning. *Developmental psychology* 42, 6 (2006), 1259.

[24] Alison Engel, Kathryn Lucido, and Kyla Cook. 2018. Rethinking narrative: Leveraging storytelling for science learning. *Childhood Education* 94, 6 (2018), 4–12.

[25] Silvia Fernandes, Jorge Abreu, Pedro Almeida, and Rita Santos. 2018. A Review of Voice User Interfaces for Interactive TV. In *Iberoamerican Conference on Applications and Usability of Interactive TV*. Springer, 115–128.

[26] Natalie Anne Freed. 2012. *" This is the fluffy robot that only speaks french": language use between preschoolers, their families, and a social robot while sharing virtual toys*. Ph.D. Dissertation. Massachusetts Institute of Technology.

[27] Andrew Gambino, Jesse Fox, and Rabindra A Ratan. 2020. Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication* 1, 1 (2020), 5.

[28] Radhika Garg and Subhasree Sengupta. 2020. Conversational Technologies for In-home Learning: Using Co-Design to Understand Children's and Parents' Perspectives. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[29] Radhika Garg and Subhasree Sengupta. 2020. He Is Just Like Me: A Study of the Long-Term Use of Smart Speakers by Parents and Children. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–24.

[30] Caroline Gaudreau, Yemimah A King, Rebecca A Dore, Hannah Puttre, Deborah Nichols, Kathy Hirsh-Pasek, and Roberta Michnick Golinkoff. 2020. Preschoolers benefit equally from video chat, pseudo-contingent video, and live book reading: implications for storytime during the coronavirus pandemic and beyond. *Frontiers in Psychology* 11 (2020), 2158.

[31] Rochel Gelman and Kimberly Brenneman. 2004. Science learning pathways for young children. *Early Childhood Research Quarterly* 19, 1 (2004), 150–158.

[32] Arthur C Graesser, Mark W Conley, and Andrew Olney. 2012. Intelligent tutoring systems. *APA educational psychology handbook, Vol 3: Application to learning and teaching*. (2012), 451–473.

[33] Arthur C Graesser, Sidney D'Mello, Xiangen Hu, Zhiqiang Cai, Andrew Olney, and Brent Morgan. 2012. AutoTutor. In *Applied natural language processing: Identification, investigation and resolution*. IGI Global, 169–187.

[34] Arthur C Graesser, Haiying Li, and Carol Forsyth. 2014. Learning by communicating in natural language with conversational agents. *Current Directions in Psychological Science* 23, 5 (2014), 374–380.

[35] Arthur C Graesser, Kurt VanLehn, Carolyn P Rosé, Pamela W Jordan, and Derek Harter. 2001. Intelligent tutoring systems with conversational dialogue. *AI magazine* 22, 4 (2001), 39–39.

[36] Arthur C Graesser, Katja Wiemer-Hastings, Peter Wiemer-Hastings, Roger Kreuz, Tutoring Research Group, et al. 1999. AutoTutor: A simulation of a human tutor. *Cognitive Systems Research* 1, 1 (1999), 35–51.

[37] James H Gray, Emily Reardon, and Jennifer A Kotler. 2017. Designing for parasocial relationships and learning: Linear video, interactive media, and artificial intelligence. In *Proceedings of the 2017 Conference on Interaction Design and Children*. 227–237.

[38] Brenna Hassinger-Das, Andres S Bustamante, Kathy Hirsh-Pasek, and Roberta Michnick Golinkoff. 2018. Learning landscapes: Playing the way to learning and engagement in public spaces. *Education Sciences* 8, 2 (2018), 74.

[39] Eduardo Hernández-Campos, Carlos R Jaimez-González, and Betzabet García-Mendoza. 2020. Interactive Mobile Applications to Support the Teaching of Reading and Writing of Spanish for Children in Primary Education. *International Journal of Interactive Mobile Technologies* 14, 14 (2020).

[40] Zorah Hilvert-Bruce, James T Neill, Max Sjöblom, and Juho Hamari. 2018. Social motivations of live-streaming viewer engagement on Twitch. *Computers in Human Behavior* 84 (2018), 58–67.

[41] Kathy Hirsh-Pasek, Jennifer M Zosh, Roberta Michnick Golinkoff, James H Gray, Michael B Robb, and Jordy Kaufman. 2015. Putting education in "educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest* 16, 1 (2015), 3–34.

[42] Sebastian Hobert and Raphael Meyer von Wolff. 2019. Say hello to your new automated tutor–a structured literature review on pedagogical conversational agents. (2019).

[43] Bruce D Homer, Charles Raffaele, and Hamadi Henderson. 2020. Games as Playful Learning: Implications of Developmental Theory for Game-Based Learning. *Handbook of Game-Based Learning, edited by Jan L. Plass, Richard E. Mayer, and Bruce D. Homer* (2020), 25–52.

[44] Alice Ann Howard Gola, Melissa N Richards, Alexis R Lauricella, and Sandra L Calvert. 2013. Building meaningful parasocial relationships between toddlers and media characters to teach early mathematical skills. *Media Psychology* 16, 4 (2013), 390–411.

[45] Pirita Ihamäki and Katriina Heljakka. 2018. The internet of toys, connectedness and character-based play in early education. In *Proceedings of the Future Technologies Conference*. Springer, 1079–1096.

[46] Jens F Jensen. 2008. Interactive television-a brief media history. In *European Conference on Interactive Television*. Springer, 1–10.

[47] Gabi Jerzembek and Simon Murphy. 2013. A narrative review of problem-based learning with school-aged children: implementation and outcomes. *Educational Review* 65, 2 (2013), 206–218.

[48] Joan N Kaderavek, Ying Guo, and Laura M Justice. 2014. Validity of the children's orientation to book reading rating scale. *Journal of Research in Reading* 37, 2 (2014), 159–178.

[49] Robert Kalinowski, Ying Xu, and Katie Salen Tekinbaş. 2021. The ecological context of preschool-aged children's selection of media content. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*.

[50] Connie A Korpan, Gay L Bisanz, Jeffrey Bisanz, Conrad Boehme, and Mervyn A Lynch. 1997. What did you learn outside of school today? Using structured interviews to document home and community activities related to science and technology. *Science Education* 81, 6 (1997), 651–662.

[51] Jacqueline M Kory-Westlund and Cynthia Breazeal. 2019. A long-term study of young children's rapport, social emulation, and language learning with a peer-like robot playmate in preschool. *Frontiers in Robotics and AI* 6 (2019), 81.

[52] Marina Krcmar and Drew P Cingel. 2019. Do young children really learn best from the use of direct address in children's television? *Media Psychology* 22, 1 (2019), 152–171.

[53] Alexis R Lauricella, Tiffany A Pempek, Rachel Barr, and Sandra L Calvert. 2010. Contingent computer interactions for young children's object retrieval success. *Journal of Applied Developmental Psychology* 31, 5 (2010), 362–369.

[54] Michael Sangyeob Lee, Carrie Heeter, and Robert LaRose. 2010. A modern Cinderella story: a comparison of viewer responses to interactive vs linear narrative in solitary and co-viewing settings. *New Media & Society* 12, 5 (2010), 779–795.

[55] Dani Levine, Amy Pace, Rufan Luo, Kathy Hirsh-Pasek, Roberta Michnick Golinkoff, Jill de Villiers, Aquiles Iglesias, and Mary Sweig Wilson. 2020. Evaluating socioeconomic gaps in preschoolers' vocabulary, syntax and language process skills with the Quick Interactive Language Screener (QUILS). *Early Childhood Research Quarterly* 50 (2020), 114–128.

[56] Maria Luce Lupetti, Yuan Yao, Haipeng Mi, and Claudio Germak. 2017. Design for children's playful learning with robots. *Future Internet* 9, 3 (2017), 52.

[57] Betsy McCarthy, Linlin Li, Michelle Tiu, and Sara Atienza. 2013. PBS KIDS mathematics transmedia suites in preschool homes. In *Proceedings of the 12th International Conference on Interaction Design and Children*. 128–136.

[58] Scott W McQuiggan, Jonathan P Rowe, Sunyoung Lee, and James C Lester. 2008. Story-based learning: The impact of narrative on learning experiences and outcomes. In *International Conference on Intelligent Tutoring Systems*. Springer, 530–539.

[59] Emily McReynolds, Sarah Hubbard, Timothy Lau, Aditya Saraf, Maya Cakmak, and Franziska Roesner. 2017. Toys that listen: A study of parents, children, and internet-connected toys. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5197–5207.

[60] Rachel E Meirson. 2020. *Choose Your Own Adventure: The Future of Interactive Media*. Ph.D. Dissertation. Drexel University.

[61] Joseph E Michaelis and Bilge Mutlu. 2019. Supporting interest in science learning with a social robot. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. 71–82.

[62] Ivonne Monarca, Franceli L Cibrian, Angel Mendoza, Gillian Hayes, and Monica Tentori. 2020. Why doesn't the conversational agent understand me? a language analysis of children speech. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*. 90–93.

[63] Roxana Moreno and Richard Mayer. 2007. Interactive multimodal learning environments. *Educational psychology review* 19, 3 (2007), 309–326.

[64] NGSS. 2013. Next Generation Science Standards (NGSS): The Three Dimensions of Science Learning. (2013). https://www.nextgenscience.org/

[65] Stamatios Papadakis, Michail Kalogiannakis, and Nicholas Zaranis. 2017. Designing and creating an educational app rubric for preschool teachers. *Education and Information Technologies* 22, 6 (2017), 3147–3165.

[66] Bhargavi Paranjape, Yubin Ge, Zhen Bai, Jessica Hammer, and Justine Cassell. 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*. 71–78.

[67] Alanna Peebles, James Alex Bonus, and Marie-Louise Mares. 2018. Questions+ answers+ agency: Interactive touchscreens and Children's learning from a socio-emotional TV story. *Computers in Human Behavior* 85 (2018), 339–348.

[68] Jessica Taylor Piotrowski. 2014. Participatory cues and program familiarity predict young children's learning from educational television. *Media Psychology* 17, 3 (2014), 311–331.

[69] Melissa N Richards and Sandra L Calvert. 2017. Measuring young US children's parasocial relationships: Toward the creation of a child self-report survey. *Journal of Children and Media* 11, 2 (2017), 229–240.

[70] Victoria Rideout. 2017. The Common Sense census: Media use by kids age zero to eight. *San Francisco, CA: Common Sense Media* (2017), 263–283.

[71] Victoria Rideout and Michael Robb. 2020. The Common Sense Census: Media use by kids age zero to eight.

[72] Sarah Roseberry, Kathy Hirsh-Pasek, and Roberta M Golinkoff. 2014. Skype me! Socially contingent interactions help toddlers learn language. *Child development* 85, 3 (2014), 956–970.

[73] Sherry Ruan, Jiayu He, Rui Ying, Jonathan Burkle, Dunia Hakim, Anna Wang, Yufeng Yin, Lily Zhou, Qianyao Xu, Abdallah AbuHashem, et al. 2020. Supporting children's math learning with feedback-augmented narrative technology. In *Proceedings of the Interaction Design and Children Conference*. 567–580.

[74] Kimiko Ryokai, Cati Vaucelle, and Justine Cassell. 2003. Virtual peers as partners in storytelling and literacy learning. *Journal of computer assisted learning* 19, 2 (2003), 195–208.

[75] Preeti G Samudra, Rachel M Flynn, and Kevin M Wong. 2019. Coviewing Educational Media: Does Coviewing Help Low-Income Preschoolers Learn Auditory and Audiovisual Vocabulary Associations? *AERA Open* 5, 2 (2019), 2332858419853238.

[76] Masahiro Shiomi, Takayuki Kanda, Iris Howley, Kotaro Hayashi, and Norihiro Hagita. 2015. Can a social robot stimulate science curiosity in classrooms? *International Journal of Social Robotics* 7, 5 (2015), 641–652.

[77] Ian Smith, Fiona Stewart, and Phil Turner. 2004. Winky Dink and you: Determining patterns of narrative for interactive television design. In *Proceedings of the second European interactive television conference: Enhancing the experience, Brighton*. Citeseer, 1–10.

[78] Catherine Snow. 1983. Literacy and language: Relationships during the preschool years. *Harvard educational review* 53, 2 (1983), 165–189.

[79] Gabrielle A Strouse. 2011. *Dialogic video: influence of dialogic reading techniques on preschoolers' learning from video stories*. Ph.D. Dissertation.

[80] Gabrielle A Strouse, Katherine O'Doherty, and Georgene L Troseth. 2013. Effective coviewing: Preschoolers' learning from video after a dialogic questioning intervention. *Developmental psychology* 49, 12 (2013), 2368.

[81] Gabrielle A Strouse, Georgene L Troseth, Katherine D O'Doherty, and Megan M Saylor. 2018. Co-viewing supports toddlers' word learning from contingent and noncontingent video. *Journal of experimental child psychology* 166 (2018), 310–326.

[82] Penelope Sweetser, Daniel Johnson, Anne Ozdowska, and Peta Wyeth. 2012. Active versus passive screen time for young children. *Australasian Journal of Early Childhood* 37, 4 (2012), 94–98.

[83] Georgene L Troseth, Megan M Saylor, and Allison H Archer. 2006. Young children's use of video as a source of socially relevant information. *Child development* 77, 3 (2006), 786–799.

[84] Sarah Vaala, Anna Ly, and Michael H Levine. 2015. Getting a Read on the App Stores: A Market Scan and Analysis of Children's Literacy Apps. Full Report.. In *Joan Ganz Cooney Center at Sesame Workshop*. ERIC.

[85] Rianne van den Berghe, Josje Verhagen, Ora Oudgenoeg-Paz, Sanne Van der Ven, and Paul Leseman. 2019. Social robots for language learning: A review. *Review of Educational Research* 89, 2 (2019), 259–295.

[86] Radu-Daniel Vatavu, Pejman Saeghe, Teresa Chambel, Vinoba Vinayagamoorthy, and Marian F Ursu. 2020. Conceptualizing Augmented Reality Television for the Living Room. In *ACM International Conference on Interactive Media Experiences*. 1–12.

[87] Peter Vorderer, Silvia Knobloch, and Holger Schramm. 2001. Does entertainment suffer from interactivity? The impact of watching an interactive TV movie on viewers' experience of entertainment. *Media Psychology* 3, 4 (2001), 343–363.

[88] Wayne Ward, Ronald Cole, Daniel Bolanos, Cindy Buchenroth-Martin, Edward Svirsky, Sarel Van Vuuren, Timothy Weston, Jing Zheng, and Lee Becker. 2011. My science tutor: A conversational multimedia virtual tutor for elementary school science. *ACM Transactions on Speech and Language Processing (TSLP)* 7, 4 (2011), 1–29.

[89] Nicola Whitton. 2018. Playful learning: tools, techniques, and tactics. *Research in Learning Technology* 26 (2018).

[90] Ying Xu, Joseph Aubele, Valery Vigil, Andres S. Bustamante, Young-Suk Kim, and Mark Warschauer. [n.d.]. Dialogue with a conversational agent promotes children's story comprehension via enhancing engagement. *Child Development* n/a, n/a ([n. d.]). https://doi.org/10.1111/cdev.13708

[91] Ying Xu, Stacy Branham, Xinwei Deng, Penelope Collins, and Mark Warschauer. 2021. Are Current Voice Interfaces Designed to Support Children's Language Development?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.

[92] Ying Xu and Mark Warschauer. 2019. Young children's reading and learning with conversational agents. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.

[93] Ying Xu and Mark Warschauer. 2020. Exploring young children's engagement in joint reading with a conversational agent. In *Proceedings of the Interaction Design and Children Conference*. 216–228.

[94] Ying Xu and Mark Warschauer. 2020. What Are You Talking To?: Understanding Children's Perceptions of Conversational Agents. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

Appendix A ELINOR'S QUESTIONS IN THE CONVERSATIONAL VIDEOS

A.1 Episode 1: Aerodynamics

(1) Hi, I'm Elinor. I love to run around and play outside. I like to run super fast and feel the wind on my ears! Do you like to run fast?
(2) Have you ever ridden in a car going really fast?
(3) Camilla's car is faster. Let's look closely at our car and Camilla's car to see if we can figure out why Camilla's is faster. How is Camilla's car different from my car?
(4) Ari thinks that changing the color of our car will make it as fast as Camila's car. What can we do to test that idea?

(5) Now we painted our car yellow and drew the racing stripe. Let's think about what we know and make a prediction! A prediction is saying what you think will happen. Do you predict that our car will go as fast as Camila's car?
   (a) Why do you think so?/Why don't you think so?
(6) Now we added a cupholder to the car. Let's think about what we know and make a prediction! Do you predict that our car will go as fast as Camila's car?
   (a) Why do you think so?/Why don't you think so?
(7) My dad knows a lot! Do boxy things or pointy things move through the air slower?
(8) Our car looks really fast now! I notice our car and Camila's car have a similar shape! What is the shape of our cars now?
(9) Now our car and Camila's car are both streamlined. Let's think about what we know and make a prediction! Do you predict that our car will go as fast as Camila's car?
   (a) Why do you think so?/Why don't you think so?
(10) We made a really fast car! It's so fun to try different things and figure out what works and what doesn't work to make the car go faster. What did we change to make our car go faster?

A.2 Episode 2: Molting

(1) Hi, I'm Elinor! I like to explore plants and animals in Nature. I think snakes are sooo interesting. Have you ever seen a snake before?
(2) What do you notice about this snake's skin?
(3) It looks like Ari has gotten too big for his lucky shirt. Hmmm, how does it feel to wear a shirt that is too small?
(4) Wow, snakes smell with their tongue? That is so interesting! What body part do you use to smell?
(5) Oh, the snake's skin is so flaky and not shiny. I wonder what is happening. Do you think the snake will have shiny skin again?
(6) Oh no, Ari hasn't stopped any soccer shots yet. Why do you think he can't stop the shots today?
(7) Hmmm. How many snakes do you see in the box?
(8) This is such a cool tool! How does this tool help me observe the snake?
(9) Hmmm, why do you think the snake's skin fell off?
(10) Hmm, molting… That's so interesting! Snakes molt as they get bigger. Their old skin comes off in one big piece! Look at your skin, do you molt like a snake as you get bigger?
(11) We're gonna let the snake out of the box. Now that it molted its old skin, will the snake move faster or slower?

A.3 Episode 3: Viscosity

(1) Do you like honey?
(2) Have you ever touched honey before?
(3) Hmmm, how does it feel when you touch honey?
(4) Did you see that!? The bee just drank some juice, called nectar from this flower! What do you think the bee is gonna make with the juice from this flower?
(5) Haha! I just said a funny word: "goopy". What do you think goopy means?
(6) When Ari flaps his wings, we can feel the air moving and blowing! Bees flap their wings too when they make honey. How does flapping their wings help the bees make honey?
(7) Let's make a prediction! Do you think adding water in, will get all this goopy honey off Ari?
(8) The ketchup isn't coming out of the bottle! I wonder why. Why do you think the ketchup isn't coming out?
(9) Guess what I'm going to do. How am I going to get the ketchup to come out of the bottle?

Appendix B ITEMS IN THE SCIENCE ASSESSMENT IN THE RANDOMIZED STUDY FOR THE EPISODE ON VISCOSITY

(1) Ok, let's talk about how bees make honey! The bees first drink something from the flower. What do bees drink?
(2) Bees then turn the runny nectar into honey. What do they take out of the nectar to turn it into honey?
(3) What do bees do to get rid of the water in the nectar?
(4) I'm trying to get the honey out of this bottle and I'm squeezing the bottle really hard. My arm is so tired! Why do you think the honey is not coming out of the bottle? [Researcher squeezes a bottle of honey]
(5) Think about what you have learned from Elinor, how can I make the honey come out of the bottle more easily?
(6) Pretend that I have two bottles here. This bottle has something goopy in it, and this one has something runny in it. Now I'm going to turn the bottles upside down. What do you think will come out faster, the thick goopy thing or the runny thing?
(7) If you want to dilute ketchup, what do you need to do?
(8) Now I'm gonna ask you to watch a short video carefully, and I will ask you two questions after that [The video shows four marble dropping in four jars of liquids with varying viscosity respectively].
   (a) Which jar is the most goopy? Which two jars are the most runny?
   (b) Why do you think this one is the most goopy? Why do you think this one is the most runny?
(9) At the beginning of the show, Elinor and her friends stood in line for a long time to wait for Mrs Llama's ketchup. Why was the line moving so slowly?
(10) Elinor and her friends had two problems in the show. Take a look at these two pictures [One picture shows honey stuck to Ari's face, and the other picture shows Elinor and friends trying to get ketchup out of a bottle].
   (a) What problem do they have in the first picture?
   (b) What problem do they have in the second picture?
   (c) What is similar, or the same, between these two problems?