Computers & Education 161 (2021) 104059

Contents lists available at ScienceDirect

Computers & Education

journal homepage: http://www.elsevier.com/locate/compedu

Same benefits, different communication patterns: Comparing Children's reading with a conversational agent vs. a human partner

Ying Xu [a,*], Dakuo Wang [b], Penelope Collins [a], Hyelim Lee [a], Mark Warschauer [a]

[a] School of Education, University of California, Irvine, USA
[b] IBM Research AI, Cambridge, USA

ARTICLE INFO

Keywords:
Conversational agents
Language development
Storybook reading
Communication
Young children

ABSTRACT

Storybook reading accompanied by adult-guided conversation provides a stimulating context for children's language development. Conversational agents powered by artificial intelligence, such as smart speakers, are prevalent in children's homes and have the potential to engage children in storybook reading as language partners. However, little research has explored the effectiveness of using conversational agents to support children's language development. This study examined how an automated conversational agent can read stories to children via a smart speaker while asking questions and providing contingent feedback. Using a randomized experiment among 90 children aged three to six years, this study compared these children's story comprehension and verbal engagement in storybook reading with a conversational agent versus an adult. The conversational agent's guided conversation was found to be as supportive in improving children's story comprehension as that provided by an adult language partner. At the same time, this study uncovered a number of differences in children's verbal engagement when interacting with a conversational agent versus with an adult. Specifically, children who read with the conversational agent responded to questions with better intelligibility, whereas those who read with an adult responded to questions with higher productivity, lexical diversity, and topical relevance. And the two groups responded to questions with a similar level of accuracy. In addition, questions requiring high cognitive demand amplified the differences in of verbal engagement between the conversational agent and adult partner. The study offers important implications for developing and researching conversational agent systems to support children's language development.

Children's development of language skills in preschool years has a profound impact on their later literacy proficiency and overall academic success. Early language skills center on the ability to understand and convey meaning in oral language form (Kim, 2017; Kim, Park, & Wagner, 2014). Extensive research shows that children's development of language skills begins in homes long before children start formal instruction (Fan, Antle, Hoskyn, Neustaedter, & Cramer, 2017; Gest, Freeman, Domitrovich, & Welsh, 2004; Roth, Speece, & Cooper, 2002; Whorrall & Cabell, 2016). **Storybook reading** by family members, typically parents, provides a comfortable environment for stimulating children's language skills. During storybook reading, parents sit together with and read to their children, ideally engaging the children in **guided conversation** where parents serve as children's language partner by posing questions and providing responsive feedback (Golinkoff, Hoff, Rowe, Tamis-LeMonda, & Hirsh-Pasek, 2019; Lever & Sénéchal, 2011). This kind of

* Corresponding author. School of Education at the University of California, Irvine, CA, 92697, USA.
E-mail address: ying.xu@uci.edu (Y. Xu).

https://doi.org/10.1016/j.compedu.2020.104059
Received 29 May 2020; Received in revised form 14 October 2020; Accepted 18 October 2020
Available online 23 October 2020

guided conversation substantially amplifies the learning benefits associated with storybook reading (for a review, see Mol, Bus, de Jong, & Smeets, 2008). However, parents may not always have the language skills, time, or inclination to engage in such conversation-rich storybook reading with their children (Cooter, 2006; Manz, Hughes, Barnabas, Bracaliello, & Ginsburg-Block, 2010; Zevenbergen & Whitehurst, 2003).

In recent years, researchers believe **intelligent systems** with a conversational interface[1] can potentially provide children with additional language learning opportunities, as they have become increasingly powerful and are capable of simulating some interpersonal communications. A growing body of research has developed conversational interfaces that can engage children in a variety of conversations as part of the experiences (see Belpaeme, Kennedy, Ramachandran, Scassellati, & Tanaka, 2018; Kennedy, Baxter, Senft, & Belpaeme, 2016, for review). Some intelligent systems developed in these studies can perform storybook reading tasks adaptable to a child's language level (e.g., Kory & Breazeal, 2014; Kory, Jeong, & Breazeal, 2013); others can employ game-like interactions for vocabulary and language learning (e.g., Freed, 2012; Movellan, Eckhardt, Virnes, & Rodriguez, 2009). Studies have demonstrated the feasibility and educational potential of intelligent systems as language partners (Gordon, Spaulding, Westlund, Lee, Plummer, Martinez et al., 2016; Kanero et al., 2018; Kory & Breazeal, 2014).

Most of the intelligent systems in existing studies have an embodied representation as a virtual avatar (Mack, Cummings, Rembert, & Gilbert, 2019; Pauchet et al., 2017) or as a physical robotic body (Belpaeme et al., 2018; Freed, 2012; Movellan et al., 2009; Kory & Breazeal, 2014; Shamekhi, Liao, Wang, Bellamy, & Erickson, 2018). These experimental systems are often designed for narrowly specific scenarios (e.g., a robot to teach food related French vocabularies; Freed, 2012), and thus are rarely adopted by the general public (de Graaf, Ben Allouch, & van Dijk, 2017; Jacques, Følstad, Gerber, Grudin, Luger, Monroy- Hernández, & Wang, 2019). On the contrary, **conversational agents (CAs)** in a smart speaker form,[2] such as Google Home and Amazon Echo, are already used by many families as consumer-oriented voice assistants (Brush, Hazas, & Albrecht, 2018). According to a report, over 150 million households in the U.S. owned smart speakers in early 2020 (Kinsella, 2020). Studies have found that children enjoy their spontaneous interactions with CAs in their homes; children initiated questions (e.g., "Hey Google, does unicorn exist?"; Lovato, Piper, & Wartella, 2019) or commanded CAs to perform small tasks (e.g., "Hey Alexa, play a Christmas song"; Sciuto, Saini, Forlizzi, & Hong, 2018). Despite the popularity of these affordable and versatile smart speakers, little research has been carried out to build CA systems based on smart speakers to support children's language development. Therefore, the ultimate objective of this research is to examine the potential of fully automated CAs in the form of a widely-adopted smart speaker that engages children in guided conversation in storybook reading (Blewitt, Rump, Shealy, & Cook, 2009; Chien, 2013; Zhou & Yadav, 2017).

1. Related work

1.1. Storybook reading with guided conversations for Children's language learning

Storybook reading is an effective way of fostering children's language development in their early years (Bus, 2001; Change & Huang, 2016; Yen, Y. Chen, Cheng, S. Chen, Y.-Y. Chen, Ni, & Hiniker, 2018). For young children who are not able to decode text independently, storybook reading typically involves them listening to their parents reading out loud a picture book while looking at images. This activity cultivates young children's ability to comprehend oral narratives, thus laying the foundation for understanding the more complex text in higher grade levels. Storybook reading by parents, such as bedtime stories, is a highly routinized activity engaged in by families across cultures (Shanahan & Lonigan, 2010).

In its basic form, storybook reading involves children merely listening to their parents reading the text verbatim (Lenhart, Lenhard, Vaahtoranta, & Suggate, 2018). But this form can be enriched with additional interactive strategies. One effective interactive strategy is to engage children in **guided conversation** (Zevenbergen & Whitehurst, 2003), in which parents ask children prepared questions and provide responsive feedback with a goal of stimulating children's active participation in the reading process. Through back-and-forth conversation, children reflect on and vocally express their understanding of the story. A meta-analysis that reviewed 16 studies has suggested an added value on children's language development resulting from incorporating guided conversation in storybook reading activities (Mol et al., 2008).

Specifically, researchers believe that guided conversation benefits both children's **story comprehension** as well as **verbal engagement** during the storybook reading activity (Vukelich, 1976). In these studies, story comprehension is typically assessed by a battery of questions developed specifically for a story. For example, Lever and Sénéchal found that children who had guided conversation with an experimenter performed significantly better in retelling story elements than those who were not asked any questions during the storybook reading (Lever & Sénéchal, 2011). When analyzing children's verbal engagement in guided conversation, researchers commonly focus on the quality of children's responses to questions asked by parents along one or more of the five aspects, namely language productivity, lexical diversity, topical relevance, accuracy, and intelligibility (Westerveld & Roberts, 2017). A study, for example, suggested that children were more engaged in a guided conversation, as indicated by greater quantity and topical relevance in their responses, if they had higher language proficiency (Westerveld & Roberts, 2017). These prior studies have established useful metrics for evaluating the effectiveness of guided conversation, which guides the development of measures used in this study.

[1] This paper does not consider text-based conversational agents (i.e. "chatbots") (e.g., Hu, Xu, Liu, You, Guo, & Sinha et al., 2018; Xu, Liu, Guo, Sinha, & Akkiraju, 2017), given its focus on young children who cannot read or type.

[2] In this paper, "disembodied conversational agents", "CAs", and "smart speakers" are used interchangeably.

Some studies have investigated the types of questions parents asked in a storybook reading exercise (Birbili & Karagiorgou, 2009). In general, studies suggest that parents should ask questions at different **cognitive demand levels** (Blewitt et al., 2009). Low-cognitive-demand questions typically revolve around a specific story fact, and high-cognitive-demand questions require children to make predictions and inferences based on information that is only implicit in the text. The different cognitive processes required to answer low- and high-demand questions lead to specific patterns in children's verbal engagement (Raphael, 1986). For example, a study found that the children's responses to low-cognitive-demand questions are more concise and simpler than those to a high-cognitive-demand question (Raphael, Highfield, & Au, 2006). This differential pattern has suggested that researchers should take questions' cognitive demand level into consideration when designing and evaluating CAs that engage children in story-related dialogues.

In summary, traditional research suggests that an effective reading partner can increase children's language development through engaging children in guided conversation. Yet, this kind of guided conversation is not as common as may have been expected: Parents do not always pause the story, ask questions, and comment on their children's response. This could be due to parents either assuming their child can learn well enough by simply listening to parent reading, or lacking the skills or time to incorporate such interactive opportunities (Golinkoff, Hoff, Rowe, Tamis-LeMonda, & Hirsh-Pasek, 2019). The recent development of intelligent systems with voice interfaces that can carry out natural conversation may provide an alternative approach to enrich children's in-home reading experiences.

1.2. Embodied intelligent systems as Children's language learning partners

Using embodied intelligent systems, both robots and virtual avatars, to enhance children's language learning through a voice interface has been a popular research topic in recent years (Papadopoulos et al., 2020). A number of initiatives have developed robotic intelligent systems to carry out structured language learning activities. For example, Kory and Breazeal (2014) developed a robotic learning companion for preschool children's oral language development. The robot was designed to tell children stories with different vocabulary complexities and teach children these words. The study found that children learned the vocabulary words that the robot had introduced in their conversation. Michaelis and Mutlu (2017) implemented a robot that was designed to make pre-programmed comments at particular points in a story as a child read aloud. Another group of researchers developed a robotic intelligent agent to support children's French language learning (Freed, 2012). The robot played a food-selection game with children and then talked about that food item with the children in French. The study found that this game-like conversation helped children learn these French words (Freed, 2012). Some other projects have developed intelligent agents embodied in avatars. Allen and colleagues utilized an avatar agent to speak with students in authentic situations, with a goal of improving students' comprehension, pronunciation, and vocabulary in a foreign language (Allen, Divekar, Drozdal, Balagyozyan, Zheng, Song et al., 2019). Authors incorporated an agent in a children's science animation series to teach children scientific vocabularies, and this agent was embodied in the series' main character (Xu & Warschauer, 2020d).

In addition to this prior work contributing to system development, other studies have evaluated the effectiveness of embodied intelligent systems in children's learning context by comparing systems' performance with human learning partners. In terms of learning outcomes, for example, Westlund and colleagues found that children learn unfamiliar words equally well whether with a robot or with a human interlocutor (Westlund et al., 2017). Hong and colleagues also suggested that incorporating a robot teaching assistant in a classroom led to students' similar level of reading and writing improvement as compared to having a human assistant (Hong, Huang, Hsu, & Shen, 2016). In terms of children's verbal engagement with intelligent systems, for example, Hyde and colleagues found that children produced a comparable amount of utterances whether their on-screen conversation was with another human or with an avatar whose speech was operated by an experimenter (Hyde, Kiesler, Hodgins, & Carter, 2014). Tewari and Canny found in their study that children produced even more utterances that were relevant when playing a game with an animal character agent as compared to children playing a game with a familiar human (Tewari & Canny, 2014). They speculated that children's high-level language productions with this particular agent may stem from the more immersive experience of conversing directly with the game's character.

However, these aforementioned embodied systems relied heavily on non-verbal communication (e.g., eye gaze, body orientation) or anthropomorphism features (Tan, Wang, & Sabanovic, 2018) to engage children in learning activities, and these features are not supported by CAs. Despite many studies suggesting that embodied systems' non-verbal cues help establish social relationships with learners and thus positively affect learning (e.g., Gordon et al., 2016; Kennedy et al., 2016), such non-verbal behaviors may also place more cognitive load on the children, which may inhibit children's capacity to process information related to the learning and concentrate on the conversation (Kennedy, Baxter, & Belpaeme, 2015). These two results lead to conflicting hypotheses regarding how the effectiveness of disembodied CAs without non-verbal capacity may compare to that of embodied systems.

1.3. Disembodied conversational agents (CAs) and Child-CA interaction

The research on children's interaction with smart speakers has been growing due to these devices' increasing prevalence in many households over the past few years. These studies utilized various methodologies, including parent or child interviews, observations, diary instruments, or in-home audio recordings, and the majority of them focused on unstructured conversations initiated by children with general voice assistant tools (e.g., Amazon Alexa, Google Assistant, Apple Siri). In general, studies found that children commonly either command the voice assistant to perform specific tasks or ask questions to receive answers (Lovato & Piper, 2019). For example, through analyzing audio recordings of children talking with the smart speakers deployed in their home, Beneteau and colleagues

categorized children's interactions into three themes, namely entertainment, assistance, and information seeking (Beneteau, Richards, Zhang, Kientz, Yip, & Hiniker, 2020). This categorization scheme was echoed in Lovato's survey study (Lovato & Piper, 2015) and in Garg's study combining interviews with user log data (Garg & Sengupta, 2020a). Another study conducted by Lovato and colleagues specifically focused on children's information seeking behaviors with smart speakers and found that children turned to the CAs for information on a variety of topics, including language, culture, science, and math (Lovato et al., 2019). Although the CA studies reviewed above did not involve educational systems tailored for age-appropriate learning, they still indicated some learning opportunities for children as children initiated conversations with CAs (Garg & Sengupta, 2020b).

In addition to exploring how children readily use CAs, some studies also investigate how children perceive CAs. At least two studies have found that children generally perceive CAs as having cognitive ability; children in both studies indicated that the CAs they interacted with were "smart" and "knowledgeable" ((Xu & Warschauer, 2020c); Druga, Williams, Breazeal, & Resnick, 2017). Children in these two studies also perceived CAs as "friendly," truthful," and sociable companions (Xu & Warschauer, 2020c); Druga et al., 2017).

Nevertheless, children were found to sometimes encounter challenges when they interact with CAs. Some children were not aware that smart speakers can not capture or interpret non-verbal expression; thus they attempted to use both verbal and non-verbal communication when responding to the CA. As CAs could not register the non-verbal responses, the conversation flow may have suffered. However, the CAs' reliance on speech may actually be positive, since this reliance—once understood by children—encourages children to practice verbal communication that is vital for their language development Xu & Warschauer, 2020b. Indeed, two studies found that preschool-aged children made efforts to have their speech understood by CAs; they adjusted sentence structures, modified word choice, or spoke more articulately (Beneteau et al., 2019; Cheng, Yen, Y. Chen, S. Chen, & Hiniker, 2018).

Research on children's interactions with CAs shows that CAs are being seamlessly integrated into children's lives and into the family unit. This favorably positions CAs to be adapted to engage children in focused learning experiences. Indeed, an emerging yet limited body of research develops smart speaker CAs to support specific language learning goals (Smutny & Schreiberova, 2020). Most of these studies leveraged CAs to teach adults foreign languages (Fryer, Ainley, Thompson, Gibson, & Sherlock, 2017; X. L. Pham, T. Pham, Q. M Nguyen, T. H. Nguyen, & Cao, 2018), yet more research is needed to understand how young children respond to learning activities scaffolded by CA partners.

1.4. Research questions

This project focus on the design of a CA that can engage with children in a guided conversation in storybook reading and the evaluation of the effectiveness of this system. The evaluation is guided by two sets of questions that focus on children's comprehension after reading and verbal engagement during reading:

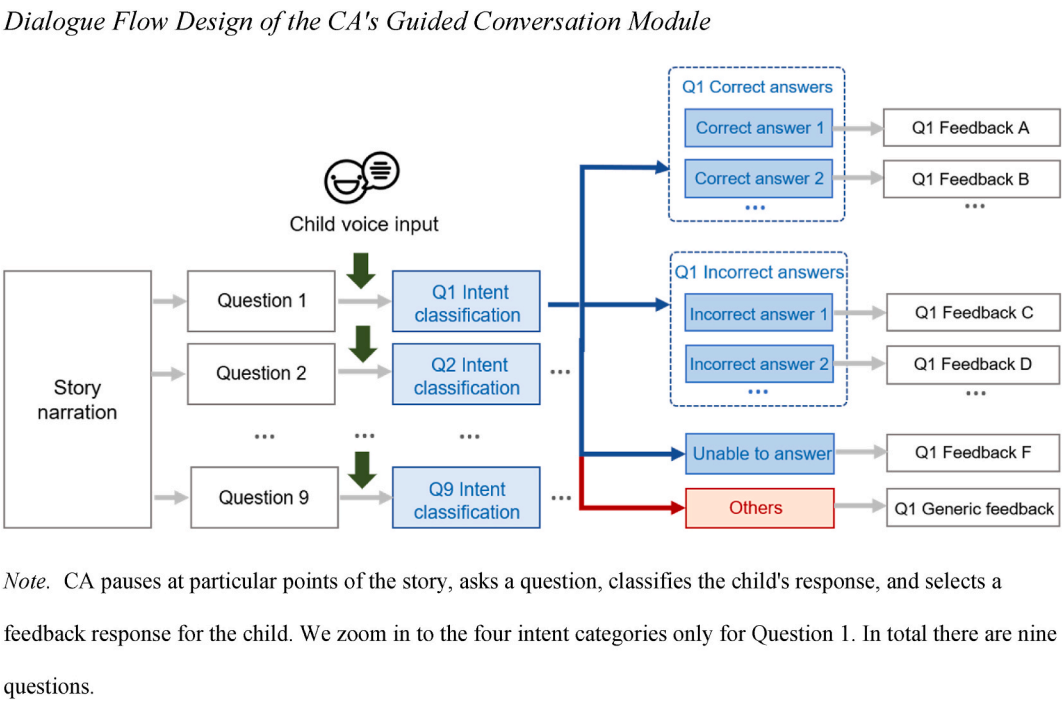


Fig. 1. Dialogue Flow Design of the CA's Guided Conversation Module, *Note.* CA pauses at particular points of the story, asks a question, classifies the child's response, and selects a feedback response for the child. We zoom in to the four intent categories only for Question 1. In total there are nine questions.

RQ1: Does guided conversation with the CA improve children's story comprehension? If so, how does this improvement compare to that resulting from conversing with a human partner?
RQ2a: Do children's verbal engagement behaviors with a human partner resemble or differ from their behaviors with a CA partner?
RQ2b: Does the similarity or difference in verbal engagement with a CA versus human partner apply to both low- and high-cognitive-demand questions?

2. Development of the CA reading partner

The automated CA system was developed to simulate the dialogue flow of a human conversational partner. The system is built upon Dialogflow open source client library (Google Cloud, 2020). The CA's natural language understanding module was based on a generic pretrained model built in the Dialogflow engine, then retained with training utterances specific to the CA's conversation context (Lee, 2018; Sabharwal & Agrawal, 2020, pp. 13–54). These training utterances were collected from a pilot study of what children might say as a response to a particular question prompt. The CA was then able to learn from a small set of training utterances and naturally expand them to many more similar phrases so that the intent of children's verbal responses can be accurately captured and classified.

The CA engages children in a fantasy story *Three Bears in a Boat* authored by David Soman. This story was chosen because of the appropriate level of narrative complexity for the age group and potential story interest. To eliminate the confounding effects of the CA with the effects of voice quality (Cambre & Kulkarni, 2019; O'Neal et al., 2019), the CA used a female recorded voice instead of machine synthetic voice.

Nine open-ended questions were asked throughout the storytelling. Six of these were low-cognitive-demand questions, while the other three were high cognitive-demand questions. For example, the following is a paragraph from the story: "One day, when their mother was out, the three bears did something they really shouldn't have, and with a crash, their mother's beautiful blue seashell lay scattered in pieces across the floor." A low-cognitive-demand question asked, "What did the bears break?" And the answer to that question was "seashell", which was found directly in the text. A high-cognitive-demand question asked children to make an inference based on the given information in the story or to summarize the information (e.g., "How did the bears search for the seashell?")

The CA performs end-to-end language processing that transcribes children's voice input into text utterance, classifies the utterance's intent, and selects a response to that intent. As indicated in Fig. 1, for each of the nine questions, four intent categories were defined to classify a child's response utterances. These four categories were (1) a set of intents for correct answers, (2) a set of intents for incorrect answers, (3) an intent for when children explicitly express their inability to answer a question, and (4) an intent category for classifying all other intents (e.g., a child does not respond to the question at all or provides an off-topic response). For each intent within each question, there could be many variations of utterances that contained similar semantic meaning. After classifying a child's responses as belonging to one of the intents, the agent then provided differential feedback that specifically addressed that response.

The CA's language model was optimized during a three-round field testing involving 20 children. These children's various responses to the CA's nine pre-defined questions were collected. For example, the correct answers to the question "What do you think is going to happen with the weather?" describe inclement weather. Possible answers to the question may be "Stormy", "Bad", "Windy", "Rainy", etc. and thus these intents were created in the initial CA. However, during the pilot run, children were also found to commonly refer to the inclement weather as being scary (e.g., "It's kind of scary."; "The bears are afraid of this weather."; "The bears are too scared and they closed their eyes."). Thus, "Scary" was added as another intent to capture this group of utterances. This iterative process lasted three rounds, and the agent achieved an inter-rater reliability with a human coder of 0.88, assessed by Cohen's Kappa. A Cohen's Kappa above 0.80 has been considered as excellent agreement (McHugh, 2012).

3. Method

This section describes the experimental design, measures, and participants of the study.

3.1. Experimental design

This study used a three-condition between-subject experimental design, where participants were randomly assigned to one of the three conditions:

- **"Human-Story"** where children were read a story by a human partner without any guided conversation;
- **"Human-Conversation"** where children were read the same story, plus engaged in guided conversation with a human partner; or
- **"CA-Conversation"** where children were read the same story, and engaged in the same guided conversation with the CA.

In all conditions, children met individually with a trained human experimenter in a designated quiet area at their school. Prior to the experiment session, the participant received an expressive vocabulary assessment (Expressive One Word Picture Vocabulary Test [Martin & Brownell, 2011], see Section 3.2.2 for more detail) as their baseline language proficiency.

At the beginning of the experiment session, the CA or a human experimenter had casual conversation with children about the child's age and favorite colors, following the same protocol. This activity aimed to build rapport between the child and their reading partner (Human or CA).

During the storybook reading activity, children in "Human-Story" group were only read the story by a human experimenter without being asked questions, whereas children in "Human-Conversation" and "CA-Conversation" groups were asked a same list of questions

and received scripted feedback based on their answers. The smart speaker used in the "CA-Conversation" condition was a Google Home Mini device. There was a human experimenter present in the room in the "CA-Conversation" condition to ensure the child's safety but not to interact with the child.

In all of the three conditions, a physical copy of the storybook was placed in front of the child as that the child could look at the pages as they followed along the narration. Fig. 2 shows the experiment session setup of the "Human-Conversation" and the "CA-conversation" condition.

The reading activity took about 20 min. Immediately after the reading activity, children's comprehension was assessed using an assessment battery developed by the research team (see Section 3.2.3 for more details). The whole experimental session was video recorded with consent from parents or legal guardians in order to conduct video coding to analyze children's verbal engagement patterns (see Section 3.2.4 for more details).

3.2. Experiment measurements

3.2.1. Background information

A parent survey was utilized to collect background information on children's date of birth (month and year) and home language (i.e., English only, English as second language, bilingual). These two factors have been traditionally shown to associate with children's learning and engagement in storybook reading (Cain, Oakhill, & Bryant, 2000; Farnia & Geva, 2013). This survey also asked for information about children's prior experience with CAs, since this factor has been found to influence children's interactions with the CA system (Bartneck, Suzuki, Kanda, & Nomura, 2007). A child was classified as a heavy CA user if parents indicated that the child used CAs more than a few days a week.

3.2.2. Baseline language proficiency

Children's baseline oral language skills were measured by the Expressive One Word Picture Vocabulary Test Fourth Edition (EOWPVT-4), which is an experimenter-administered, norm-referenced picture-naming assessment. Each child was asked to name objects, actions, and concepts that were depicted graphically, and the test lasted 15–20 min depending on the child's English proficiency. The internal reliability (Cronbach's coefficient alpha) of EOWPVT-4 for 3- to 6-year-olds is 0.95 (Martin & Brownell, 2011). Children's oral language skills are positively associated with children's performances in storybook reading activities (Kendeou, Van den Broek, White, & Lynch, 2009).

3.2.3. Story comprehension

Children's comprehension level of the story after the storybook reading was measured as an indicator of a proximal learning outcome, similar to the research approach in Zhou and Yadav (2017). A questionnaire was developed, with a total of 10 items to measure how much a child understands the story.[3] Together, these items aim to assess children's ability to 1) memorize main story events and make inferences, 2) sort narrative sequence, and 3) retell part of the story. There were eight items on memorization and inferences. Children were first asked to freely recall the answers. If they could not recall the answer correctly, the researcher provided three multiple-choice options for children to select from. Two points were given to each item that was answered correctly through free recall and one point was given if answered correctly with multiple-choice options. There was one narrative sequence sorting item, where children were asked to place images from the book in the order they occurred in the story. Children earned two points for completely correct order and one point for partially correct order. There was one item to prompt children to retell a part of the story, where children could earn one point for mentioning each key element in their answer up to four points.

An overall story comprehension score was calculated by summing the number of points across all the items and used this score as a dependent variable for the analysis. The range is from 0 to 22 points (16 points maximum for the 8 memorization and inference-making items, 2 points maximum for the single sequence sorting item, and 4 points maximum for the story-retelling item). Cronbach's coefficient alpha is 0.87 for this story comprehension assessment.

3.2.4. Verbal engagement

Children's verbal engagement is a measure of how children responded to the CA's questions during storybook readings, which was coded from the video-taped interaction sessions. Only the Human-Conversation and CA-Conversation sessions have this measurement, because the Human-Story condition does not have guided conversation. Five sub-dimensions of verbal engagement were coded, based on the literature on parent-child storybook reading (Vukelich, 1976; Westerveld & Roberts, 2017), namely productivity, lexical diversity, topical relevance, accuracy, and intelligibility. The unit of coding was a child's response to a single prompt, and each child had nine responses.

The reliability of the coding was established using two coders. These two coders, both native English speakers, were undergraduate research assistants. Neither of them were authors of this paper. Coder A coded all of the videos, while Coder B coded a subset of the videos (30%). Coders met once every week to compare codes and discuss any discrepancies in coding. The operationalization and inter-rater reliability (i.e., Inter-class correlation) for each sub-dimension are detailed below.

Productivity. Children's language productivity was captured by the length of utterances in words. The total number of words was

[3] These 10 questions are different from the nine questions asked during the guided conversation activity.

Experiment Session Setup

Note. A child participant in the Human-Conversation group (left); and another child participant in the CA-Conversation group, the smart speaker system is highlighted (right)

Fig. 2. Experiment Session Setup, *Note.* A child participant in the Human-Conversation group (left); and another child participant in the CA-Conversation group, the smart speaker system is highlighted (right).

counted in each response, including repetitive words. The length of utterances is counted as 0 if the response does not contain verbal expressions. Meaningless speech input (e.g., filter words like Uhhh, Umm, Ahha) was also excluded from the word count. Inter-class correlation = 1.

Lexical diversity. Children's lexical diversity was captured by the number of unique words in children's responses. The repetitive words in the utterance were removed, and only the unique words were counted. Lexical diversity was coded as 0 if no verbal expression was present, and meaningless speech input was excluded from the word count. Inter-class correlation = 1.

Topical relevance. The relevance of children's response to a prompt will be coded using three categories, which indicate how well children's responses maintain the semantic flow of conversation. Childish language, imperfect grammar, or answer correctness was not penalized within the relevance code. A response that was directly addressed to the question received a score of 2, a response that was not directly addressed to the question but aligned with the overall theme of the story received a score of 1, and a response that was not related to the question or overall theme received a score of 0. Responses that did not contain verbal expressions was considered as irrelevant and received a score of 0. Inter-class correlation = 0.94.

Accuracy. The correctness of children's response to a prompt was coded as a dichotomous variable, indicating whether a response is correct or incorrect. Specifically, correct answers received a score of 1 and incorrect answers received a score of 0. Responses that did not contain verbal expressions were considered as incorrect and received a score of 0. Inter-class correlation = 1.

Intelligibility. The intelligibility of children's utterances for each prompt was rated by a 0 to 2 scale, following the method proposed by Flipsen (FlipsenJr, 2002). A score of 0 indicated that a child's utterance was largely unintelligible, and the coders could understand less than 50% of the utterance; a score of 1 indicated that a child's utterance was mostly intelligible except for one or two words; a score of 2 indicated a child's utterance was articulate, and the coder could understand every single word. Responses that did not contain verbal expressions were excluded from this coding. Inter-class correlation = 0.87.

In the analyses of this paper, these five sub-dimensions were analyzed separately, with each of them being a dependent variable.

3.3. Participants

The study targeted to include 90 children aged 3 to 6 to participate in the study. This sample size was pre-determined by a power analysis in order to detect a minimum effect size of 0.2 (Cohen, 2013). Upon approval by the Institutional Review Board, children were recruited from five local childcare programs affiliated with or nearby a research university in the United States. These programs predominately enroll children of university students and faculty and include children from a wide range of economic, ethnic, cultural and linguistic backgrounds. The research team set up a recruitment booth at the school during pickup hours, explained the study procedure to the parents, and collected consent. There were 102 parents who provided consent for their child to participate; 90 of these children actually participated and completed the experiment, and the remaining 12 children did not participate due to absence.

Background information of the 90 children who participated in the study is displayed in Table 1. Children's average age was 58 months (4.8 years old) with a standard deviation of 9 months. From the full sample, 26 children were randomly assigned into the "Human-Story" group, 31 were in the "Human-Conversation" group, and 33 were in the "CA-Conversation" group. The breakdown of background information of each of the three groups is displayed in Table 1. Randomization check using binomial regressions indicated that there were no significant differences in child baseline characteristics among the experimental condition assignments. After randomization had already occurred, during the study, children were asked if they had already read the book *Three Bears in a Boat*. A total of 5 of the 90 children, spread out across the three conditions (1 in Human-Story, 2 in Human-Conversation, 2 in CA-

Table 1
Participant background information by experimental condition.

	Full Sample	Human-Story	Human-Conversation	CA-Conversation
Age in months	58.1 (9.33)	56.9 (9.5)	58.3 (9.1)	59.5 (8.8)
Home language				
English only	67.4%	60.0%	71.0%	69.7%
Bilingual	10.1%	16.0%	9.7%	6.1%
ESL	22.5%	24.0%	19.3%	24.2%
Heavy CA use	36.2%	33.3%	34.2%	37.9%
EOWPVT	68.2 (17.5)	68.8 (17.1)	66.7 (17.1)	70.6 (17.4)
N	90	26	31	33

Note: Standard deviations in parentheses.

Conversation), indicated they had done so. The 5 were then asked to describe what the book was about; all 5 said they couldn't remember the story.

4. Results

This section first presents the full sample descriptive statistics of the outcomes measures. Findings were then reported regarding the CA's effects on story comprehension (RQ1), verbal engagement behaviors with CA versus with human partner (RQ2a), and the interaction effects of questions' cognitive demand on verbal behaviors (RQ2b).

4.1. Descriptive statistics of outcome measures for full sample

The descriptive statistics for the full sample are presented in Table 2. Children's average score in story comprehension was 11.2, indicating that these children in the sample correctly answered half of the comprehension items correctly. In terms of children's verbal engagement as they engaged in guided conversation, the average length of utterance (i.e., productivity) was 4.4 words and the average number of unique words (i.e., lexical diversity) was 3.7 words. The average score of topical relevance was 1.5 out of 2, indicating that most children were able to generate answers that addressed the questions. Children in this study on average responded to half of the in-story questions accurately, evidenced by an accuracy rate of 0.5. Also, these children generally articulated their answers with good intelligibility, resulting in an intelligibility score of 1.9 out of 2.

The Pearson correlation coefficients between study variables and their significance levels are displayed in Table 2. Children's story comprehension was significantly positively correlated with all verbal engagement measures. Among the verbal engagement variables, productivity, diversity, relevance, and accuracy were significantly correlated with each other, while intelligibility was only significantly correlated with relevance and accuracy but not productivity or diversity.

4.2. The effect of CAs on story comprehension

The first research question examined the extent to which having guided conversation with a learning partner during storybook reading may enhance children's story comprehension and whether the benefits of guided conversation differed depending on the nature of the learning partner (i.e., a CA or a human partner).

Descriptively, children who had guided conversation with either a CA or human language partner correctly answered story comprehension questions more frequently than did children in the group without guided conversation (see Table 3). When comparing the performance of the two groups with guided conversation, children in two groups answered approximately the same number of items correctly. The difference in score was only 0.13, which was much smaller than 1 (i.e., 1 item).

The regression analyses first compared whether the two groups of children who had guided conversation performed better in story comprehension than their counterparts who did not engage in guided conversation (i.e., Human-Story group). This was considered as a baseline analysis to validate the benefits of guided conversation with low- and high-cognitive-demand questions, regardless of the

Table 2
Intercorrelations among study variables.

	Comp.	Pro.	Div.	Rel.	Acc.	Int.	Mean	SD	Range
Comprehension	1	0.28**	0.35**	0.59***	0.66***	0.32**	11.17	5.23	(0, 22.00)
Productivity		1	0.96***	0.52***	0.48***	0.03	4.38	2.55	(0, 13.78)
Diversity			1	0.55***	0.49***	0.02	3.73	1.89	(0, 11.33)
Relevance				1	0.87***	0.27*	1.47	0.53	(0, 2)
Accuracy					1	0.34*	0.54	2.19	(0, 1)
Intelligibility						1	1.90	1.65	(0, 2)

Note: Pearson correlation coefficients and significance levels reported.
$p < 0.05$ denoted as *, $p < 0.01$ denoted as **, $p < 0.001$ denoted as ***.

Table 3
Descriptive statistics of story comprehension and verbal engagement variables by experimental condition.

		Human-Story		Human-Conversation		CA-Conversation	
Comprehension		9.62	(4.62)	11.86	(5.78)	11.73	(5.03)
Productivity	Low cog.			3.24	(2.71)	3.10	(1.99)
	High cog.			7.84	(4.67)	5.81	(4.05)
	Combined			4.77	(2.53)	4.00	(2.56)
Diversity	Low cog.			2.89	(2.57)	2.62	(1.44)
	High cog.			6.35	(3.32)	4.89	(2.57)
	Combined			4.11	(1.99)	3.38	(1.75)
Relevance	Low cog.			1.56	(0.44)	1.40	(0.59)
	High cog.			1.53	(0.63)	1.36	(0.68)
	Combined			1.55	(0.46)	1.39	(0.59)
Accuracy	Low cog.			0.58	(0.25)	0.59	(0.28)
	High cog.			0.46	(0.28)	0.44	(0.29)
	Combined			0.52	(0.27)	0.51	(0.30)
Intelligibility	Low cog.			1.89	(0.17)	1.94	(0.12)
	High cog.			1.82	(0.30)	1.89	(0.27)
	Combined			1.87	(0.17)	1.93	(0.15)

Note: Standard deviations in parentheses. Verbal engagement measures not applicable in Human-Story condition.

nature of language partners who carried out the conversation. The "Human-Story" group was used as the reference group in our regression models (Model 1 in Table 4). The results indicated that both the "CA-Conversation" ($\beta = 0.44$, $p = 0.03$) and the "Human-Conversation" groups ($\beta = 0.64$, $p = 0.01$) scored significantly higher than the "Human-Story" group. The higher comprehension score achieved by the two groups with guided conversation confirmed the advantage of incorporating guided conversation in storybook reading.

A post-hoc analysis was then conducted to compare the comprehension scores between children who had guided conversation with the CA and those who had guided conversation with a human partner, by using "Human-Conversation" as the reference group in the original regression model. The result indicated that the comprehension scores of children in the "CA-Conversation" group were not statistically different from those of children in the "Human-Conversation" group ($\beta = -0.20$, $p = 0.29$). These results suggested that the guided conversation carried out by CA could yield similarly effective learning as a human partner.

4.3. The effect of CAs on verbal engagement in guided conversation

The second set of research questions (RQ2a and RQ2b) focused on children's verbal engagement behaviors in guided conversation. RQ2a examined whether children conversing with CA partners exhibited similar or different verbal engagement patterns as they would when talking with a human partner, and RQ2b examined whether any difference in verbal engagement varied based on the question's cognitive demand level. Table 3 presents descriptive statistics of children's verbal engagement between CA-Conversation and Human-Conversation conditions, as well as disaggregates the between-condition verbal engagement measures by low- and high-cognitive-demand questions. Descriptively, children produced longer, more lexically diverse, and more relevant responses when conversing with a human partner, yet children conversing with a CA language partner responded more intelligibly. The accuracy rate between the two conditions resembled each other. When the questions' cognitive demand level is taken into consideration, the difference in productivity and lexical diversity between the CA and Human groups became larger for high-cognitive-demand questions.

Table 4
Linear regression model on story comprehension measures.

	Comprehension
Human-Conv	**0.64**** (0.22)
CA-Conv	**0.44**** (0.22)
EOWPVT	**0.62***** (0.11)
Age	**0.27*** (0.12)
English only	−0.07 (0.30)
ESL	0.23 (0.31)
Heavy CA use	0.02 (0.19)
R^2	0.62

Note: "Human-Story" is the reference group. Coefficients are standardized. Standard error in parentheses.
$p < 0.05$ denoted as *, $p < 0.01$ denoted as **, $p < 0.001$ denoted as ***. Significant coefficients bolded.

Multilevel linear analyses were employed to formally test whether children conversing with a CA exhibited similar verbal engagement patterns as they would when talking with a human partner, and whether any difference in verbal engagement varied based on the question's cognitive demand. For each of the five engagement metrics, the analyses first focused on the effects of language partner and questions' cognitive demand (Models 1, 3, 5, 7, and 9 in Table 5). The analyses then focused on examining the interaction effects between the nature of language partner and questions' cognitive demand, by including an additional cross-level interaction term between the nature of language partner and questions' cognitive demand (i.e., CA-Conv × High cog; Models 2, 4, 6, 8, and 10 in Table 5). All models were controlled for children's age, expressive vocabulary score, home language, and prior CA experiences, given the documented relations between these variables and children's verbal responses.

4.3.1. Productivity

The multilevel model analysis suggested a significant effect of the nature of learning partners on language productivity. Specifically, reading with a human partner resulted in children responding in greater length ($\beta = -0.28$, $p = 0.04$) than they did to the CA partner (Model 1 in Table 5), suggesting some benefits of a human language partner in promoting language productivity over CAs. Questions that require high cognitive demand elicited responses that were significantly longer ($\beta = 0.78$, $p = 0.00$) than low-demand questions (Model 1 in Table 5).

Furthermore, the cross-level interaction between the nature of language partner and the questions' cognitive demand was significant (Model 2 in Table 5). The difference in language productivity between "CA-Conversation" and "Human-Conversation" conditions was significantly amplified when children were answering high-cognitive-demand questions ($\beta = -0.44$, $p = 0.002$; Fig. 3-A). Specifically, children's response length did not differ significantly for questions that required low cognitive demand ($\beta = -0.14$, $p = 0.36$). However, human partners' advantages in eliciting more language production became prominent when the conversation was cognitively challenging.

4.3.2. Lexical diversity

Regarding the effect of a CA versus human partner on lexical diversity (Model 3 in Table 5), children's responses contained more unique words when conversing with a human partner than with the CA ($\beta = -0.35$, $p = 0.01$), indicating some advantages of a human partner in encouraging more lexically diverse utterances. In terms of the effect of questions' cognitive demand levels (Model 3 in Table 5), children were found to respond to high-cognitive-demand questions using more unique words ($\beta = 0.83$, $p = 0.00$).

Furthermore, the model with the interaction effect (Model 4 in Table 5) indicated that the difference in lexical diversity between "CA-Conversation" and "Human-Conversation" was significantly larger among questions that require high cognitive demand ($\beta = -0.35$, $p = 0.01$; Fig. 3-B). Conversing with a human partner did not elicit more lexically diverse responses from children than did the CA if the questions required low cognitive demand ($\beta = -0.24$, $p = 0.11$). Yet among the questions that were cognitively challenging, human partners were more likely to invite responses with higher lexical diversity than were CAs.

4.3.3. Topical relevance

In terms of the effect of a CA versus human partner on topical relevance (Model 5 in Table 5), children's responses were more topically relevant when they conversed with a human partner than with a CA partner ($\beta = -0.30$, $p = 0.04$). However, when focusing on the effect of questions' cognitive demand (Model 5 in Table 5), a child's ability to generate topically relevant answers did not

Table 5
Multilevel linear models on verbal engagement measures.

	Productivity		Diversity		Relevance		Accuracy		Intelligibility	
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
CA-Conv	**−0.28***	−0.14	**−0.35***	−0.24	**−0.30***	**−0.30***	−0.08	−0.07	**0.23***	0.20
	(0.14)	(0.15)	(0.14)	(0.15)	(0.15)	(0.15)	(0.11)	(0.12)	(0.11)	(0.13)
High cog	**0.78*****	**0.58*****	**0.83*****	**0.67*****	−0.04	−0.04	**−0.28*****	**−0.30*****	−0.14	−0.11
	(0.07)	(0.1)	(0.07)	(0.1)	(0.07)	(0.1)	(0.08)	(0.11)	(0.09)	(0.12)
CA-Conv × High cog		**−0.44****		**−0.35***		−0.01		−0.04		0.07
		(0.14)		(0.14)		(0.15)		(0.16)		(0.17)
EOWPVT	**0.21***	**0.21***	**0.26***	**0.26***	**0.24***	**0.24***	**0.30*****	**0.30*****	0.13	0.13
	(0.11)	(0.11)	(0.10)	(0.10)	(0.11)	(0.11)	(0.08)	(0.08)	(0.08)	(0.08)
Age	0.05	0.05 (0.10)	0.04	0.04	0.19	0.19	0.10 (0.08)	0.10 (0.08)	0.02	0.02
	(0.10)		(0.10)	(0.10)	(0.11)	(0.11)			(0.08)	(0.08)
English only	−0.09	−0.09	−0.16	−0.16	0.13	0.13	0.03 (0.23)	0.03 (0.23)	0.16	0.16
	(0.31)	(0.31)	(0.30)	(0.30)	(0.32)	(0.32)			(0.26)	(0.26)
ESL	0.20	0.20 (0.32)	0.08	0.08	0.35	0.35	0.23 (0.24)	0.23 (0.24)	0.20	0.16
	(0.32)		(0.32)	(0.32)	(0.33)	(0.33)			(0.27)	(0.26)
Heavy CA use	−0.09	−0.09	−0.05	−0.05	0.02	0.02	−0.10	−0.10	−0.01	−0.01
	(0.17)	(0.17)	(0.16)	(0.16)	(0.17)	(0.17)	(0.13)	(0.13)	(0.13)	(0.13)

Note: Each row is an independent variable, and each column is a dependent variable. For each dependent variable, two models were performed: the latter one has an interaction effect (CA-Conv × High cog). EOWPVT, Age, English only, ESL, and Heavy CA use are 5 control variables. Coefficients are standardized coefficient. Standard error in parentheses.
$p < 0.05$ denoted as * is considered statistically significant, $p < 0.01$ denoted as **, $p < 0.001$ denoted as ***. Significant coefficients bolded.

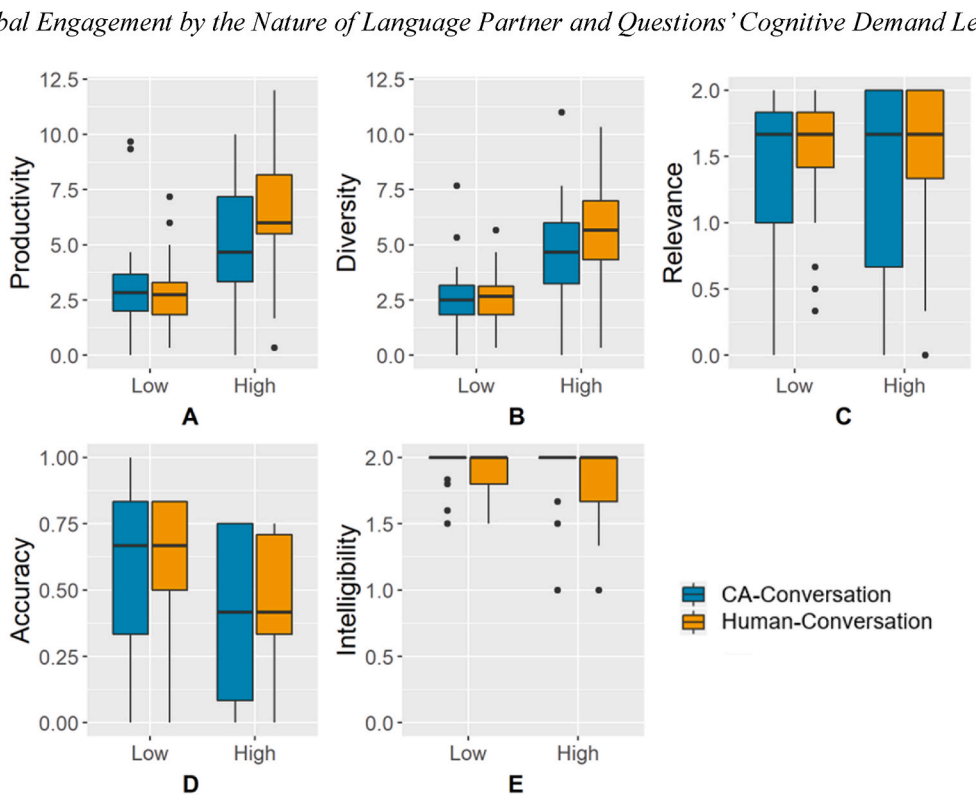


Fig. 3. Verbal engagement by the nature of language partner and questions' cognitive demand levels.

significantly differ by low- and high-cognitive-demand questions ($\beta = -0.04$, $p = 0.60$).

The cross-level interaction between questions' cognitive demand and learning condition was not significant, as indicated in Model 6 in Table 5 ($\beta = -0.01$, $p = 0.95$). Specifically, this non-significant interaction effect suggested that human learning partners, in general, were more likely to elicit relevant responses from children, and this pattern was consistent for both low- and high-cognitive-demand questions (Fig. 3-C).

4.3.4. Accuracy

In terms of the effect of a CA versus human partner on response accuracy (Model 7 in Table 5), there was no significant differences in response accuracy between children conversing with the CA and those conversing with a human partner ($\beta = -0.08$, $p = 0.38$). In terms of the effect of questions' cognitive demand level (Model 7 in Table 5), unsurprisingly, children were more likely to answer lower cognitive demand questions accurately compared to higher cognitive demand questions ($\beta = -0.28$, $p = 0.00$). In terms of the interaction effect between questions' cognitive demand and the nature of language partner (Model 8 in Table 5), our finding indicated that, when answering low-cognitive-demand and high-cognitive-demand questions, children had a comparable level of accuracy in the "CA-Conversation" and "Human-Conversation" conditions ($\beta = -0.04$, $p = 0.79$; Fig. 3-D).

4.3.5. Intelligibility

In terms of the effects of a CA versus human partner on intelligibility of children's responses (Model 9 in Table 5), children's responses appeared to be more intelligible when conversing with a CA than with a human partner ($\beta = 0.23$, $p = 0.04$). This suggested CAs' advantages in encouraging children to articulate their utterances. Questions' cognitive demand level did not significantly influence the intelligibility of children's responses ($\beta = -0.14$, $p = 0.10$; Model 9 in Table 5). When including the interaction effect (Model 10 in Table 5), children showed a similar level of intelligibility when conversing with the CA or a human partner, regardless of whether they answered low- or high-cognitive-demand questions ($\beta = 0.07$, $p = 0.67$; Fig. 3-E).

4.4. Robustness check

We ran a robustness check excluding the 5 children who had indicated they had already read the story. All results remained consistent with the findings reported above.

5. Discussion

This section discusses the interpretation of the findings with regard to why children can learn from a CA partner and demonstrated certain verbal engagement behaviors in the guided conversations. As one of the first studies to design and evaluate a CA learning partner, the findings of this study provide novel design implications for further improving CAs in an affordable smart speaker format and deploying such systems to enrich young children's everyday literacy learning.

5.1. The learning benefits of conversing with disembodied CAs

This study demonstrated that the children who had guided conversation, whether with the CA or a human, comprehended the story better than the group who did not engage in guided conversation. This result was not surprising given that the vast education literature documenting the added value of guided conversation over non-interactive storybook reading where parents merely read the text (for reviews, see Arnold & Whitehurst, 1994; Mol et al., 2008). However, the study's findings also extend this traditional line of research by demonstrating that a CA can potentially facilitate children's language learning as effectively as a human partner in this study's context. The positive effects of guided conversation in this study also validated the CA dialogue design strategy that incorporates high- and low-cognitive-demand questions (Raphael, 1986), proving the usefulness of developing intelligent systems grounded in education theories (Callaghan & Reich, 2018). One important factor enabling the CA in this study to replicate these benefits is its capacity to respond to children adaptively based on the children's answers. This kind of adaptive response helped identify and clarify children's misconceptions and reinforce an accurate understanding of the story (Aksan, Kochanska, & Ortmann, 2006; Funamoto & Rinaldi, 2015).

The fact that the CA, with only a voice interface, can benefit children's story comprehension as much as face-to-face human partners can reinforces the importance of verbal dialogue in promoting children's language skills. Yet, this result should not be interpreted as undermining the role non-verbal cues ordinarily play in boosting learning effectiveness (Dunn, Rodriguez, Miller, Gerhardt, Vannatta, Saylor et al., 2011; Kahlbaugh & Haviland, 1994; Negi, 2009). Instead, this result arises out of the storybook reading scenario in general. During storybook reading, young children typically look at a picture book while listening to the story. Therefore, children's visual channel primarily concentrates on the illustrations (Paivio, 1991), which substantially facilitates their understanding of the narration (Takacs & Bus, 2018), thus leaving limited room for processing other non-verbal information provided by a human partner (Hanson, 1989). The minimal non-verbal information children do receive from human partners during storybook reading may not be sufficient to translate into the short-term learning benefits this study assessed, such as immediate recall of story elements. Yet it is plausible that non-verbal cues may influence how children verbally engage with their reading partner, which is discussed below.

5.2. Verbal engagement behaviors with disembodied CAs

The findings of this study revealed nuances in how children verbally respond differently to a natural human and to a CA. Specifically, children were found to generate longer and more lexically diverse responses when conversing with a human partner than with a CA. The human partner's ability to leverage social cues (e.g., looking at children as children formulate responses [Guo & Feng, 2013]) could contribute to this difference. Moreover, children were found to provide more relevant responses to a human partner. One speculation is that the social presence of a human partner may have encouraged children to make an effort to maintain the conversation flow (Groom, 2008; Kim et al., 2013; Zhou, Mark, Li, & Yang, 2019). Yet interestingly, despite the differences in response relevance, children answered questions from the CA and the human partner with a similar level of accuracy, corroborating the finding on post-storybook reading comprehension that the CA benefits children's learning as well as the human does. Taken together, this suggests that the lower relevance of children's responses to CAs' questions was not due to cognitive factors but may be related to social or behavioral factors. Lastly, CAs were found to enhance children's intelligibility. This may be due to children's perceptions of the CAs' listening ability: Studies have suggested that people are likely to talk more clearly and slowly if they perceive their partners as needing additional support in interpreting their utterances (Rooy, 2009). This pattern was also identified in children's communication with CAs: Young children adjusted their speech style if they perceived CAs as encountering difficulties in understanding them (Beneteau et al., 2019; Cheng, Yen, Chen, Chen, & Hiniker, 2018).

There was evidence that the cognitive demand required to participate in the conversation, on top of the nature of the language partner, jointly shapes children's verbal engagement. Specifically, high-cognitive-demand questions amplified the effects of human partners on eliciting longer and more lexically diverse responses from children. This may also be attributed to the social presence of a human partner discussed before. An adult figure is often perceived as more authoritative by children, which may encourage children to devote greater cognitive effort in attempting challenging questions (Davis, 2003).

Despite the nuances discussed above, the descriptive statistics suggest that children's responses to the CA were not fundamentally different from their responses to a human partner. Children in both conditions replied to the prompting with multi-word responses, kept their responses quite relevant to the question, and uttered their responses intelligibly. This implies that children, regardless of whether they are conversing with a CA or a human partner, follow a shared convention during the conversation. This finding resonates with the prominent "Computers as Social Actors (CASA)" paradigm (Nass, Moon, & Morkes, 1997; Nass, Steuer, & Tauber, 1994), which suggests that human users, especially children, tend to treat intelligent systems as human beings. Numerous studies on children's interactions with embodied intelligent systems (e.g., robots and avatars) have supported this paradigm (Admoni & Scassellati, 2017; Fink, Lemaignan, Dillenbourg, Rétornaz, Vaussard, Berthoud et al., 2014; Heerink et al., 2012; Melson, Kahn, Friedman,

Roberts, Garrett, & Gill, 2009; Spolaôr & Benitti, 2017; Tewari & Canny, 2014). The current study thus extends the application of the CASA paradigm to include disembodied CAs that do not have anthropomorphic figures and are restricted to verbal communication. This extension is also supported by other theories that suggest an intelligent system's verbal ability is a central factor that shapes how users judge the system's sociability and intelligibility and thus how they interact with that system (Araujo, 2018).

5.3. Designing better CAs for early childhood language development

The current study sheds light on three implications for future design of CA language partners for young children.

First, it is important to design CAs with a clear theoretical rationale for meeting children's unique learning needs. In this study, the CA was tailored to a storybook reading context, incorporating evidence-based strategies that take into consideration the cognitive demand required by the conversation. The CA's ability to improve children's learning confirms the importance of a theoretically-driven design approach. Unfortunately, according to a recent review of over 500 voice-based apps on the market ((Xu & Warschauer, 2020a)), many of the apps purporting to benefit language learning were not grounded in research, thus limiting these apps' abilities to fulfill their intended educational goal.

Second, it is important to fully leverage a CA's conversation capacity to compensate for its inability to utilize non-verbal expressions. In the current study, the CA did not fully simulate a human partner in eliciting children's elaborate, complex, and relevant responses. As discussed before, it is possible that the CA's disadvantages may arise from its disembodiment which prevents it from leveraging non-verbal cues. Developers can compensate for this lack by improving on such CAs' conversational expressiveness (Lin, Ginns, Wang, & Zhang, 2020). For example, CAs may be designed to clearly explain to children how to best answer a question (e.g., "Listen to the question carefully and try to say as much as you can!") or provide follow-up prompts to encourage longer or more appropriate responses (e.g., "Great job! Can you say some more?" or "That is a good idea! But how about what we've just talked about?"). CAs may also leverage natural acoustic features (i.e., tone, prosody, speech speed), such as asking a question with a tone of genuine curiosity, which may entice children to more thoroughly express their thoughts.

Third, it is also important for developers to recognize the unique properties of CAs that make them especially useful learning tools regardless of whether they precisely mimic a human partner. Some researchers have proposed that CAs may be particularly valuable for providing children with opportunities to practice their language skills since CAs require children to communicate verbally (Vaquero, Saz, Lleida, Marcos, & Canalís, 2006). This supports the claim of Clark and colleagues (Clark, Munteanu, Wade, Cowan, Pantidi, Cooney et al., 2019), who suggest that developers need not and should not attempt to develop CAs that exactly emulate human-to-human interactions. Instead, CAs should be envisioned as a new form of language partner, one that could complement and enrich children's everyday conversational experiences.

5.4. Generalizability

The findings of this study are intended to be generalized to broader scenarios of using conversational interface to support young children's engagement and learning from storybook reading. First, storybook reading is one of the most important sources of language and literacy development during early childhood, and the activity simulated in this study followed a general reading format that children normally engage with in their everyday lives: listening to a story narrated by a reading partner while looking at a hard-copy book. Second, the conversation in this study took place via a smart speaker, which is a commercial product that is already widely available among children in the U.S. and other developed countries. Third, the participants of this study (i.e., 90 children aged 3–6) represented children from a variety of backgrounds, including approximately one third who are speakers of languages other than English and two thirds who had little prior experience with CAs.

However, there are also limitations to the generalizability including the fact that the experiment took place in a single well-educated community in one country and other issues of study design discussed below.

5.5. Limitations and future directions

This study was among the first to utilize smart speakers as language partners, thus demonstrating the potential of this kind of CA. Future research will want to build from what this study has shed light on by attending to some of the limitations of this initial study.

First, this study only focused on children's immediate learning outcomes—comprehension of the story they have just listened to. While this finding provides important evidence regarding the positive learning effects of smart speaker CAs, it is unclear whether interacting with a CA reading partner may lead to long-run benefits to children's language development. Future studies may want to include a delayed post-test or carry out a more intense intervention to examine long-term effects of CAs on children's general language abilities.

Second, the reading activity was adapted from a commercial book that is available to the public. It is possible that some children had prior exposure to this book, which may influence their verbal engagement and story comprehension. Given that this study was a randomized experiment, the expected likelihood of children who had prior exposure to the book across the three conditions should be equivalent. Thus, children's prior exposure should not have confounded with the results regarding the comparisons across the three reading conditions. Nevertheless, in the future it would be of value to replicate this research with an original book created for the purpose of the research.

Third, this study was conducted in an experimental manner where the human partner's conversational behaviors were scripted. This design increased the internal validity of the findings, reducing the confounding effects resulting from the variations in ways of

human partners actually carrying out the guided conversation. Nevertheless, in a naturalistic setting, parents may unitize guided conversation in varying degrees and with varying approaches. Future studies may compare children's learning and engagement with CAs with those with their parents and family members who routinely read with them. In addition, in this study, children had brief casual conversation with the CA, which might be viewed as a training opportunity that familiarized children with the scheme of interacting with the CA. Future studies may want to further explore how to best design such warm-up interactions to better support children who were either reluctant to participate in the conversation or encountered obstacles (e.g., relied on non-verbal expressions) during their conversation with the CA.

Fourth, the current study has demonstrated that CAs can facilitate storybook reading through conversing with children individually. However, this study is not intended to develop agents that stand in the role of parents to read to their children. Rather, it paves the way towards a new computing paradigm of "Human-AI Collaboration" (Grudin, 2017) where CAs (or other AI systems) serve as a collaborator for parents to support more involved parental guidance during storybook reading. In fact, a recent study has suggested that CAs like smart speakers have helped augment parent-child interaction as a third-party mediator (Beneteau et al., 2020; Scoito et al., 2018). Future studies may examine the feasibility of using CAs as a training system to model to parents the beneficial strategies of guided conversation or redesigning prompts to include parents in the conversation. In addition, future systems may consider how parents, CAs and children may form a "conversation triad", where parents and CAs collaboratively engage children in discussions related to a story. By doing so, CAs' potential could be maximined through to fully mobilizing other elements in family contexts that work together to support children's language development.

6. Conclusion

This study demonstrated the potential of smart speaker CAs in carrying out guided conversation in storybook reading activities to nurture children's language development. Given that smart speakers are already accessible in many homes, the endeavor to augment smart speakers' usefulness as learning tools may have profound impact on children and on the market. Encouragingly, this study demonstrated that the CA developed in this study based on education literature was equally supportive as a human partner in enhancing children's story comprehension. However, nuanced patterns in children's verbal engagement were also identified: CAs and human partners have their own advantages in some respects. Among the first experimental studies comparing children's learning and engagement with CAs versus adults, this study provides initial evidence on the potential of smart speakers as effective reading partners. Nevertheless, the precious parent-child interactions cannot be replaced by artificially intelligent systems; yet CAs may supplement parents' current practices and thus enrich children's early literacy experiences. Understanding how children learn from and engage with CAs is an important step in gaining a more complete picture of the role that intelligent systems play in children's educational landscape in today's world.

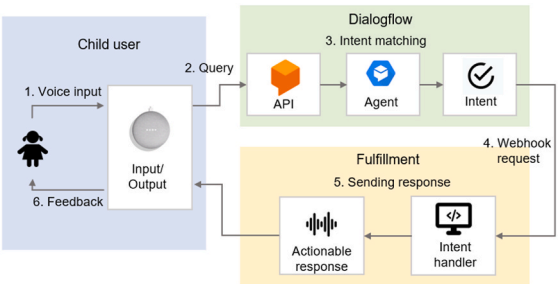
Credit author statement

Ying Xu: Conceptualization, Methodology, Data Collection and Analysis, Writing- Original draft preparation; Dakuo Wang: Conceptualization, Writing- Reviewing and Editing; Penelope Collins: Data Interpretation; Hyelim Lee: Data Analysis, Writing - Editing; Mark Warschauer: Data Interpretation, Writing- Reviewing and Editing.

Author note

We have not known conflict of interest to disclose.

Appendix A



The figure illustrates the system architecture of the automated CA system.

1. End user (the child) generates a voice input to the Google Home Mini device to respond to a question.
2. The voice input is streamed to the Dialogflow API.

3. Dialogflow matches the child voice input to an intent, utilizing the agent's language training model.
4. Dialogflow sends a webhook request to call functions (intent handlers) hosted on Google Cloud.
5. These intent handlers define the following actions (feedback) for a given intent.
6. The child user hears the feedback.

References

Admoni, H., & Scassellati, B. (2017). Social eye gaze in human-robot interaction: A review. *Journal of Human-Robot Interaction, 6*(1), 25–63. https://doi.org/10.5898/jhri.6.1.admoni

Aksan, N., Kochanska, G., & Ortmann, M. R. (2006). Mutually responsive orientation between parents and their young children: Toward methodological advances in the science of relationships. *Developmental Psychology, 42*(5), 833–848. https://doi.org/10.1037/0012-1649.42.5.833

Allen, D., Divekar, R. R., Drozdal, J., Balagyozyan, L., Zheng, S., Song, Z., et al. (2019). The rensselaer Mandarin project — a cognitive and immersive language learning environment. *Proceedings of the AAAI Conference on Artificial Intelligence, 33*, 9845–9846. https://doi.org/10.1609/aaai.v33i01.33019845

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior, 85*, 183–189. https://doi.org/10.1016/j.chb.2018.03.051

Arnold, D. H., & Whitehurst, G. J. (1994). *Accelerating language development through picture book reading: A summary of dialogic reading and its effect*.

Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. (2007). The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI & Society, 21*(1–2), 217–230. https://doi.org/10.1007/s00146-006-0052-7

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics, 3*(21), eaat5954. https://doi.org/10.1126/scirobotics.aat5954

Beneteau, E., Boone, A., Wu, Y., Kientz, J. A., Yip, J., & Hiniker, A. (2020). Parenting with Alexa: Exploring the introduction of smart speakers on family dynamics. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–13). https://doi.org/10.1145/3313831.3376344

Beneteau, E., Richards, O. K., Zhang, M., Kientz, J. A., Yip, J., & Hiniker, A. (2019). Communication breakdowns between families and Alexa. In *Proceedings of the 2019 CHI conference on human factors in computing systems - CHI '19* (pp. 1–13). https://doi.org/10.1145/3290605.3300473

Birbili, M., & Karagiorgou, I. (2009). Helping children and their parents ask better questions: An intervention study. *Journal of Research in Childhood Education, 24*(1), 18–31. https://doi.org/10.1080/02568540903439359

Blewitt, P., Rump, K. M., Shealy, S. E., & Cook, S. A. (2009). Shared book reading: When and how questions affect young children's word learning. *Journal of Educational Psychology, 101*(2), 294–304. https://doi.org/10.1037/a0013844

Brush, A. J., Hazas, M., & Albrecht, J. (2018). Smart homes: Undeniable reality or always just around the corner? *IEEE Pervasive Computing, 17*(1), 82–86. https://doi.org/10.1109/mprv.2018.011591065

Bus, A. G. (2001). Joint caregiver-child storybook reading: A route to literacy development. *Handbook of Early Literacy Research*, 171–191.

Cain, K., Oakhill, J., & Bryant, P. (2000). Investigating the causes of reading comprehension failure: The comprehension-age match design. *Reading and Writing, 12*(1–2), 31–40.

Callaghan, M. N., & Reich, S. M. (2018). Are educational preschool apps designed to teach?: An analysis of the app market. *Learning, Media and Technology, 43*(3), 280–293. https://doi.org/10.1080/17439884.2018.1498355

Cambre, J., & Kulkarni, C. (2019). One voice fits all?: Social implications and research challenges of designing voices for smart devices. In *Proceedings of the ACM on human-computer interaction* (Vol. 3, pp. 1–19). CSCW. https://doi.org/10.1145/3359325.

Change, C.-J., & Huang, C.-C. (2016). Mother–child talk during joint book reading in two social classes in Taiwan: Interaction strategies and information types. *Applied PsychoLinguistics, 37*(2), 387–410. https://doi.org/10.1017/s0142716415000041

Cheng, Y., Yen, K., Chen, Y., Chen, S., & Hiniker, A. (2018). Why doesn't it work?: Voice-driven interfaces and young children's communication repair strategies. *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 337–348. https://doi.org/10.1145/3202185.3202749

Chien, C.-W. (2013). Using Raphael's QARs as differentiated instruction with picture books. *English Teaching Forum, 51*, 20–27 (ERIC).

Clark, L., Munteanu, C., Wade, V., Cowan, B. R., Pantidi, N., Cooney, O., et al. (2019). What makes a good conversation?: Challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI conference on human factors in computing systems - CHI '19* (pp. 1–12). https://doi.org/10.1145/3290605.3300705

Cloud, G. (2020). *Dialogueflow documentation. Technical report*.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic Press.

Cooter, K. S. (2006). When mama can't read: Counteracting intergenerational illiteracy. *The Reading Teacher, 59*(7), 698–702. https://doi.org/10.1598/rt.59.7.9

Davis, H. A. (2003). Conceptualizing the role and influence of student-teacher relationships on children's social and cognitive development. *Educational Psychologist, 38*(4), 207–234. https://doi.org/10.1207/s15326985ep3804_2

Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017). "Hey Google is it ok if I eat you?": Initial explorations in child-agent interaction. In *Proceedings of the 2017 conference on interaction design and children* (pp. 595–600). https://doi.org/10.1145/3078072.3084330

Dunn, M. J., Rodriguez, E. M., Miller, K. S., Gerhardt, C. A., Vannatta, K., Saylor, M., et al. (2011). Direct observation of mother-child communication in pediatric cancer: Assessment of verbal and non-verbal behavior and emotion. *Journal of Pediatric Psychology, 36*(5), 565–575. https://doi.org/10.1093/jpepsy/jsq062

Fan, M., Antle, A. N., Hoskyn, M., Neustaedter, C., & Cramer, E. S. (2017). Why tangibility matters: A design case study of at-risk children learning to read and spell. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 1805–1816). https://doi.org/10.1145/3025453.3026048

Farnia, F., & Geva, E. (2013). Growth and predictors of change in English language learners' reading comprehension. *Journal of Research in Reading, 36*, 389–421. https://doi.org/10.1111/jrir.12003

Fink, J., Lemaignan, S., Dillenbourg, P., Rétornaz, P., Vaussard, F., Berthoud, A., et al. (2014). Which robot behavior can motivate children to tidy up their toys?: Design and evaluation of "ranger. *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction - HRI, '14*, 439–446. https://doi.org/10.1145/2559636.2559659

FlipsenJr, P. (2002). Longitudinal changes in articulation rate and phonetic phrase length in children with speech delay. *Journal of Speech, Language, and Hearing Research, 45*(1), 100–110. https://doi.org/10.1044/1092-4388(2002/008

Freed, N. A. (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. thesis]).

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior, 75*, 461–468. https://doi.org/10.1016/j.chb.2017.05.045

Funamoto, A., & Rinaldi, C. M. (2015). Measuring parent-child mutuality: A review of current observational coding systems. *Infant Mental Health Journal, 36*(1), 3–11. https://doi.org/10.1002/imhj.21481

Garg, R., & Sengupta, S. (2020a). 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), 1–24. https://doi.org/10.1145/3381002

Garg, R., & Sengupta, S. (2020b). 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* (pp. 1–13). https://doi.org/10.1145/3313831.3376631

Gest, S. D., Freeman, N. R., Domitrovich, C. E., & Welsh, J. A. (2004). Shared book reading and children's language comprehension skills: The moderating role of parental discipline practices. *Early Childhood Research Quarterly, 19*(2), 319–336. https://doi.org/10.1016/j.ecresq.2004.04.007

Golinkoff, R. M., Hoff, E., Rowe, M. L., Tamis-LeMonda, C. S., & Hirsh-Pasek, K. (2019). Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child Development, 90*(3), 985–992. https://doi.org/10.1111/cdev.13128

Gordon, G., Spaulding, S., Westlund, J., Lee, J., Plummer, L., Martinez, M., et al. (2016). Affective personalization of a social robot tutor for children's second language skills. In *Thirtieth AAAI conference on artificial intelligence*.

de Graaf, M., Ben Allouch, S., & van Dijk, J. (2017). Why do they refuse to use my robot?: Reasons for non-use derived from a long-term home study. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, 224–233. https://doi.org/10.1145/2909824.3020236

Groom, V. (2008). What's the best role for a robot. In *Proceedings of the fifth international conference on informatics in control, automation and robotics service* (pp. 323–328). Automation and Robotics (ICINCO). https://doi.org/10.5220/0001507103230328.

Grudin, J. (2017). From tool to partner: The evolution of human-computer interaction. *Synthesis Lectures on Human-Centered Informatics, 10*(1). https://doi.org/10.2200/s00745ed1v01y201612hci035. i–183.

Guo, J., & Feng, G. (2013). How eye gaze feedback changes parent-child joint attention in shared storybook reading? *Eye Gaze in Intelligent User Interfaces*, 9–21. https://doi.org/10.1007/978-1-4471-4784-8_2

Hanson, L. (1989). Multichannel learning research applied to principles of television production: A review and synthesis of the literature. *Educational Technology, 29*, 15–19.

Heerink, M., Diaz, M., Albo-Canals, J., Angulo, C., Barco, A., Casacuberta, J., et al. (2012). A field study with primary school children on perception of social presence and interactive behavior with a pet robot. In *2012 IEEE RO-MAN: The 21st IEEE international symposium on robot and human interactive communication* (pp. 1045–1050). https://doi.org/10.1109/roman.2012.6343887

Hong, Z., Huang, Y., Hsu, M., & Shen, W. (2016). Authoring robot-assisted instructional materials for improving learning performance and motivation in EFL classrooms. *Journal of Educational Technology & Society, 19*(1), 337–349.

Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., et al. (2018). Touch your heart: A tone-aware chatbot for customer care on social media. In *Proceedings of the 2018 CHI conference on human factors in computing systems - CHI '18* (pp. 1–12). https://doi.org/10.1145/3173574.3173989

Hyde, J., Kiesler, S., Hodgins, J. K., & Carter, E. J. (2014). Conversing with children: Cartoon and video people elicit similar conversational behaviors. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1787–1796). https://doi.org/10.1145/2556288.2557280

Jacques, R., Følstad, A., Gerber, E., Grudin, J., Luger, E., Monroy-Hernández, A., et al. (2019). Conversational agents: Acting on the wave of research and development. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems* (pp. 1–8). https://doi.org/10.1145/3290607.3299034

Kahlbaugh, P. E., & Haviland, J. M. (1994). Nonverbal communication between parents and adolescents: A study of approach and avoidance behaviors. *Journal of Nonverbal Behavior, 18*(1), 91–113. https://doi.org/10.1007/bf02169080

Kanero, J., Geçkin, V., Oranç, C., Mamus, E., Küntay, A. C., & Göksun, T. (2018). Social robots for early language learning: Current evidence and future directions. *Child Development Perspectives, 12*(3), 146–151. https://doi.org/10.1111/cdep.12277

Kendeou, P., Van den Broek, P., White, M. J., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: The independent contributions of oral language and decoding skills. *Journal of Educational Psychology, 101*(4), 765–778. https://doi.org/10.1037/a0015956

Kennedy, J., Baxter, P., & Belpaeme, T. (2015). The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction - HRI '15* (pp. 67–74). https://doi.org/10.1145/2696454.2696457

Kennedy, J., Baxter, P., Senft, E., & Belpaeme, T. (2016). Social robot tutoring for child second language learning. In *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)* (pp. 231–238).

Kim, Y.-S. G. (2017). Why the simple view of reading is not simplistic: Unpacking component skills of reading using a direct and indirect effect model of reading (DIER). *Scientific Studies of Reading, 21*(4), 310–333. https://doi.org/10.1080/10888438.2017.1291643

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., et al. (2013). Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders, 43*(5), 1038–1049. https://doi.org/10.1007/s10803-012-1645-2

Kim, Y.-S., Park, C. H., & Wagner, R. K. (2014). Is oral/text reading fluency a "bridge" to reading comprehension? *Reading and Writing, 27*(1), 79–99. https://doi.org/10.1007/s11145-013-9434-7

Kinsella, B. (2020, April 28). *Nearly 90 million U.S. adults have smart speakers, adoption now exceeds one-third of consumers*. Voicebot.Ai.

Kory, J., & Breazeal, C. (2014). Storytelling with robots: Learning companions for preschool children's language development. In *The 23rd IEEE international symposium on robot and human interactive communication* (pp. 643–648). https://doi.org/10.1109/roman.2014.6926325

Kory, J. M., Jeong, S., & Breazeal, C. L. (2013). Robotic learning companions for early language development. In *Proceedings of the 15th ACM on international conference on multimodal interaction - ICMI '13* (pp. 71–72). https://doi.org/10.1145/2522848.2531750

Lee, H. (2018). *Voice user interface projects: Build voice-enabled applications using Dialogflow for Google home and Alexa skills kit for Amazon Echo*. Packt Publishing.

Lenhart, J., Lenhard, W., Vaahtoranta, E., & Suggate, S. (2018). Incidental vocabulary acquisition from listening to stories: A comparison between read-aloud and free storytelling approaches. *Educational Psychology, 38*(5), 596–616. https://doi.org/10.1080/01443410.2017.1363377

Lever, R., & Sénéchal, M. (2011). Discussing stories: On how a dialogic reading intervention improves kindergartners' oral narrative construction. *Journal of Experimental Child Psychology, 108*(1), 1–24. https://doi.org/10.1016/j.jecp.2010.07.002

Lin, L., Ginns, P., Wang, T., & Zhang, P. (2020). Using a pedagogical agent to deliver conversational style instruction: What benefits can you obtain? *Computers & Education, 143*, 103658. https://doi.org/10.1016/j.compedu.2019.103658

Lovato, S., & Piper, A. M. (2015). "Siri, is this you?": Understanding young children's interactions with voice input systems. In *Proceedings of the 14th international conference on interaction design and children - IDC '15* (pp. 335–338). https://doi.org/10.1145/2771839.2771910

Lovato, S. B., & Piper, A. M. (2019). Young children and voice search: What we know from human-computer interaction research. *Frontiers in Psychology, 10*. https://doi.org/10.3389/fpsyg.2019.00008

Lovato, S. B., Piper, A. M., & Wartella, E. A. (2019). Hey Google, do unicorns exist?: Conversational agents as a path to answers to children's questions. In *Proceedings of the 18th ACM international conference on interaction design and children* (pp. 301–313). https://doi.org/10.1145/3311927.3323150

Mack, N. A., Cummings, R., Rembert, D. G. M., & Gilbert, J. E. (2019). Co-Designing an intelligent conversational history tutor with children. In *Proceedings of the 18th ACM international conference on interaction design and children*. https://doi.org/10.1145/3311927.3325336

Manz, P. H., Hughes, C., Barnabas, E., Bracaliello, C., & Ginsburg-Block, M. (2010). A descriptive review and meta-analysis of family-based emergent literacy interventions: To what extent is the research applicable to low-income, ethnic-minority or linguistically-diverse young children? *Early Childhood Research Quarterly, 25*(4), 409–431. https://doi.org/10.1016/j.ecresq.2010.03.002

Martin, N. A., & Brownell, R. (2011). *Expressive one-word picture vocabulary test-4 (EOWPVT-4)*. Academic Therapy Publications.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica: Biochemia Medica, 22*(3), 276–282.

Melson, G. F., Kahn, P. H., Beck, A., Friedman, B., Roberts, T., Garrett, E., et al. (2009). Children's behavior toward and understanding of robotic and living dogs. *Journal of Applied Developmental Psychology, 30*(2), 92–102. https://doi.org/10.1016/j.appdev.2008.10.011

Michaelis, J. E., & Mutlu, B. (2017). Someone to read with: Design of and experiences with an in-home learning companion robot for reading. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 301–312). https://doi.org/10.1145/3025453.3025499

Mol, S. E., Bus, A. G., de Jong, M. T., & Smeets, D. J. H. (2008). Added value of dialogic parent–child book readings: A meta-analysis. *Early Education & Development, 19*(1), 7–26. https://doi.org/10.1080/10409280701838603

Movellan, J., Eckhardt, M., Virnes, M., & Rodriguez, A. (2009). Sociable robot improves toddler vocabulary skills. In *Proceedings of the 4th ACM/IEEE international conference on human robot interaction - HRI '09*. https://doi.org/10.1145/1514095.1514189

Nass, C. I., Moon, Y., & Morkes, J. (1997). Computers are social actors: A review of current. *Human Values and the Design of Computer Technology, 72*, 137.

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI conference on human factors in computing systems celebrating interdependence - CHI '94* (pp. 72–78). https://doi.org/10.1145/191666.191703

Negi, J. S. (2009). The role of teachers' non-verbal communication in ELT classroom. *Journal of NELTA*, 101–110. https://doi.org/10.3126/nelta.v14i1.3096
O'Neal, A. L. (2019). *Is Google Duplex too human?: Exploring user perceptions of opaque conversational agents* ([Ph.D. thesis]).
Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 45*(3), 255–287. https://doi.org/10.1037/h0084295
Papadopoulos, I., Lazzarino, R., Miah, S., Weaver, T., Thomas, B., & Koulouglioti, C. (2020). A systematic review of socially assistive robots in pre-tertiary education. *Computers & Education*, 103924. https://doi.org/10.1016/j.compedu.2020.103924
Pauchet, A., Şerban, O., Ruinet, M., Richard, A., Chanoni, É., & Barange, M. (2017). Interactive narration with a child: Avatar versus human in video-conference. *Intelligent Virtual Agents*, 343–346. https://doi.org/10.1007/978-3-319-67401-8_44
Pham, X. L., Pham, T., Nguyen, Q. M., Nguyen, T. H., & Cao, T. T. H. (2018). Chatbot as an intelligent personal assistant for mobile language learning. In *Proceedings of the 2018 2nd international conference on education and E-learning* (Vol. 2018, pp. 16–21). ICEEL. https://doi.org/10.1145/3291078.3291115.
Raphael, T. E. (1986). Teaching question answer relationships, revisited. *The Reading Teacher, 39*(6), 516–522.
Raphael, T. E., Highfield, K., & Au, K. H. (2006). Question-Answer relationships. *Scholastic*.
Rooy, S. C.-V. (2009). Intelligibility and perceptions of English proficiency. *World Englishes, 28*(1), 15–34. https://doi.org/10.1111/j.1467-971x.2008.01567.x
Roth, F. P., Speece, D. L., & Cooper, D. H. (2002). A longitudinal analysis of the connection between oral language and early reading. *The Journal of Educational Research, 95*(5), 259–272. https://doi.org/10.1080/00220670209596600
Sabharwal, N., & Agrawal, A. (2020). *Cognitive virtual assistants using Google Dialogflow*. Apress. https://doi.org/10.1007/978-1-4842-5741-8
Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). "Hey Alexa, what's up?": A mixed-methods studies of in-home conversational agent usage. In *Proceedings of the 2018 on designing interactive systems conference* (pp. 857–868). https://doi.org/10.1145/3196709.3196772
Shamekhi, A., Liao, Q. V., Wang, D., Bellamy, R. K. E., & Erickson, T. (2018). Face value? Exploring the effects of embodiment for a group facilitation agent. In *Proceedings of the 2018 CHI conference on human factors in computing systems - CHI '18* (pp. 1–13). https://doi.org/10.1145/3173574.3173965
Shanahan, T., & Lonigan, C. J. (2010). The national early literacy panel: A summary of the process and the report. *Educational Researcher, 39*(4), 279–285. https://doi.org/10.3102/0013189x10369172
Smutny, P., & Schreiberova, P. (2020). Chatbots for learning: A review of educational chatbots for the facebook messenger. *Computers & Education*, 103862. https://doi.org/10.1016/j.compedu.2020.103862
Spolaôr, N., & Benitti, F. B. V. (2017). Robotics applications grounded in learning theories on tertiary education: A systematic review. *Computers & Education, 112*, 97–107. https://doi.org/10.1016/j.compedu.2017.05.001
Takacs, Z. K., & Bus, A. G. (2018). How pictures in picture storybooks support young children's story comprehension: An eye-tracking experiment. *Journal of Experimental Child Psychology, 174*, 1–12. https://doi.org/10.1016/j.jecp.2018.04.013
Tan, H., Wang, D., & Sabanovic, S. (2018). Projecting life onto robots: The effects of cultural factors and design type on multi-level evaluations of robot anthropomorphism. In *2018 27th IEEE international symposium on robot and human interactive communication* (pp. 129–136). RO-MAN). https://doi.org/10.1109/roman.2018.8525584.
Tewari, A., & Canny, J. (2014). What did spot hide?: A question-answering game for preschool children. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1807–1816). https://doi.org/10.1145/2556288.2557205
Vaquero, C., Saz, O., Lleida, E., Marcos, J., & Canalís, C. (2006). Vocaliza: An application for computer-aided speech therapy in Spanish language. *IV Jornadas En Tecnologia Del Habla*, 321–326.
Vukelich, C. (1976). The development of listening comprehension through storytime. *Language Arts, 53*(8), 889–891.
Westerveld, M. F., & Roberts, J. M. A. (2017). The oral narrative comprehension and production abilities of verbal preschoolers on the autism spectrum. *Language, Speech, and Hearing Services in Schools, 48*(4), 260–272. https://doi.org/10.1044/2017_lshss-17-0003
Whorrall, J., & Cabell, S. Q. (2016). Supporting children's oral language development in the preschool classroom. *Early Childhood Education Journal, 44*(4), 335–341. https://doi.org/10.1007/s10643-015-0719-0
Xu, A., Liu, Z., Guo, Y., Sinha, V., & Akkiraju, R. (2017). A new chatbot for customer service on social media. In *Proceedings of the 2017 CHI conference on human factors in computing systems - CHI '17*. https://doi.org/10.1145/3025453.3025496
Xu, Ying, & Warschauer, Mark (2020a). A content analysis of voice-based apps on the market for early literacy development. In *Proceedings of the 19th ACM International Conference on Interaction Design and Children*. https://doi.org/10.1145/3392063.3394418
Xu, Ying, & Warschauer, Mark (2020b). Exploring young children's engagement in joint reading with a conversational agent. In *Proceedings of the 19th ACM International Conference on Interaction Design and Children (IDC '20)*. https://doi.org/10.1145/3392063.3394417
Xu, Ying, & Warschauer, Mark (2020c). 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*. https://doi.org/10.1145/3313831.3376416
Xu, Ying, & Warschauer, Mark (2020d). "Elinor is talking to me on the screen!" Integrating conversational agents into children's television programming. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. https://doi.org/10.1145/3334480.3383000
Yen, K., Chen, Y., Cheng, Y., Chen, S., Chen, Y.-Y., Ni, Y., et al. (2018). Joint media engagement between parents and preschoolers in the U.S., China, and Taiwan. In *Proceedings of the ACM on human-computer interaction* (Vol. 2, pp. 1–19). CSCW. https://doi.org/10.1145/3274461.
Zevenbergen, A., & Whitehurst, G. (2003). Dialogic reading: A shared picture book reading intervention for preschoolers. *On Reading Books to Children: Parents and Teachers*, 177–200.
Zhou, M. X., Mark, G., Li, J., & Yang, H. (2019). Trusting virtual agents: The effect of personality. *ACM Transactions on Interactive Intelligent Systems, 9*(2–3), 1–36. https://doi.org/10.1145/3232077
Zhou, N., & Yadav, A. (2017). Effects of multimedia story reading and questioning on preschoolers' vocabulary learning, story comprehension and reading engagement. *Educational Technology Research & Development, 65*(6), 1523–1545. https://doi.org/10.1007/s11423-017-9533-2