# Young Children's Reading and Learning with Conversational Agents

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

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

## ABSTRACT

Young children increasingly interact with voice-driven interfaces, such as conversational agents (CAs). The social nature of CAs makes them good learning partners for children. We have designed a storytelling CA to engage children in book reading activities. This case study presents the design of this CA and investigates children's interactions with and perception of the CA. Through observation, we found that children actively responded to the CA's prompts, reacted to the CA's feedback with great affect, and quickly learned the schema of interacting with a digital interlocutor. We also discovered that the availability of scaffolding appeared to facilitate child-CA conversation and learning. A brief post-reading interview suggested that children enjoyed their interaction with the CA. Design implications for dialogic systems for young children's informal learning are discussed.

## CCS CONCEPTS

- **Human centered computing** → Interaction Design → Empirical studies in interaction design
- **Human centered computing** → HCI → Interaction paradigms → natural language interfaces

## KEYWORDS

Conversational agents; natural language processing; artificial intelligence; dialogic reading; young children; social interaction

*CHI'19 Extended Abstracts, May 4-9, 2019, Glasgow, Scotland, UK.*

ACM ISBN 978-1-4503-5971-9/19/05.
DOI: https://doi.org/10.1145/3290607. 3299035

**ACM Reference format:**
Ying Xu and Mark Warschauer. Young children's reading and learning with conversational agents. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts (CHI'19 Extended Abstracts), May 4–9, 2019, Glasgow, Scotland, UK.* ACM, New York, NY, USA. 8 pages. https://doi.org/10.1145/3290607.3299035

## 1 INTRODUCTION

Young children learn best from reading when they socially interact with an adult or peer through meaningful conversation. This kind of dialogic reading scaffolds children's learning from the story content and also fosters their language development and long-term interest in literacy. In the recent years, conversational agents (CAs) such as Google Assistant and Amazon Alexa, have become more powerful largely due to the advances in natural language processing technologies. These CAs have the potential to simulate a social partner to engage children in dialogic reading through asking questions, giving feedback, and adjusting questions to the developmental level of the child. However, CAs are still underutilized for the purpose of facilitating young children's everyday learning.

To fill this need, our team has conducted a research and development (R&D) project to build a storytelling CA that engages children in story-related conversation while narrating the story from a picture book via a Google Home device. The CA is designed to facilitate children's comprehension of the story and enhance engagement, with scaffolded conversation tailored for children's still-developing language skills. This case study explores children's reading experiences with this storytelling CA, shedding light on the affordances and obstacles of voice-based apps for early learning. Through observations and interviews with child users, we examine two questions:

- How do children verbally and behaviorally interact with the storytelling CA?
- How do children perceive their reading experience with the storytelling CA?

## 2 CONVERSATIONAL AGENTS FOR EARLY LEARNING

Researchers have long sought to leverage CAs to develop virtual learning companions for children. This line of research, however, has hinged on CAs' previously limited ability to recognize speech (i.e., translate speech to text) and detect intent (i.e., interpret the meaning of unstructured or semi-structured text). Recent advances in natural language processing show a way out of this conundrum, as CAs are becoming more intelligent. A few studies have investigated children's free, spontaneous conversations with CAs for entertainment purposes. Druga et al. found that preschool-aged children interacted naturally with agents and perceived them to be smart and friendly [4]. In comparing child-agent versus child-human interactions, Tewari and Canny suggested that the quantity of child-initiated utterances was comparable [6]. These studies provide important evidence as to the feasibility of building educational applications that revolve around children's verbal interaction with CAs. However, few studies have explored how young children react to a CA companion within a learning task, until now a rare product.

Designing CAs as children's learning companions is a complex endeavor. Researchers first need to tailor the CA to young children's communication patterns. Given that unencumbered conversation is the prerequisite to the educational affordances of dialogic reading, it is important for CAs to avoid breakdowns in communication with children [2]. For example, in cases where CAs fail to understand children's utterances (i.e., children's utterances do not match the intent), CAs may immediately follow-up by rephrasing the prompts in a more accessible fashion.

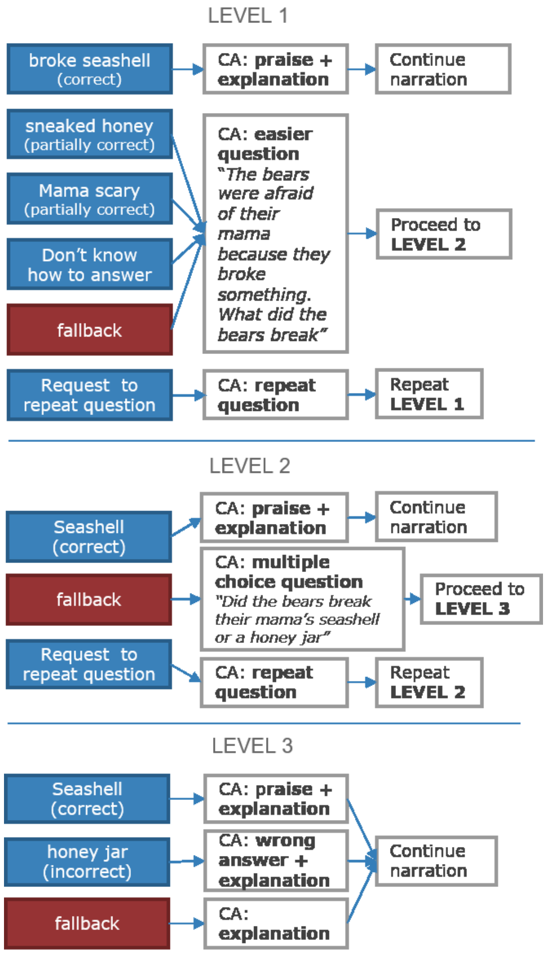

**Figure 1: Dialogue flowchart of an example question "Why do you think the bears were afraid of their Mama?" White text boxes indicate CA speech, blue text boxes indicate child speech, red text boxes represent cases where fall back is triggered.**

Going beyond the communication piece, designers should also consider how to align the conversation with educational principles to support children's learning. The first principle is developmental appropriateness, with the conversation falling within children's "zone of proximal development" (ZPD) [3]. The second principle is adaptivity, referring to the adjustments that the CA may make to attend to children's level of comprehension [1].

## 3 DESIGN OF THE STORYTELLING CA

Based on these communication and educational principles, we iteratively developed a storytelling CA that narrates a picture book, *Three Bears in a Boat*, and pauses at particular points to prompt children to answer a total of ten open-ended questions (e.g., "*Why do you think the bears were afraid of their Mama*?", "*What is the island the bears need to find shaped like*?"). The CA plays a "ding" sound before the question to evoke children's attention. After the narration of each page, children are prompted to turn to the next page by a page-flipping sound.

The conversational agent was built upon Google's Dialogflow Natural Language Understanding (NLU) engine. For each question, three types of intents were created, including the anticipated answers (i.e., correct answers, partially correct answers, incorrect answers), the request to repeat the question (e.g., "*What did you say?*"), and the indication of inability to answer the question (e.g., "*I don't know.*"). For each intent, we included a set of training utterances and defined how the CA should follow up. With the NLU engine's machine learning algorithms, the agent is able to learn from a small set of training utterances and naturally expands them to many more similar phrases so that children's voice input can be accurately mapped to predefined intents. We also built in a fallback function that will be triggered when the CA fails to match children's response with the intents (due to no response, fuzzy pronunciation, or lack of comprehension).

The scaffolding features of the CA consist of back and forth follow-up questions that are adaptive based on children's response to a previous conversational prompt [2]. We designed multiple levels of follow-up prompts for each question, with prompts in each level being more accessible than the one in the previous level. Figure 1 illustrates the dialogue flow of an example question "*Why do you think the bears were afraid of their Mama?"* Specifically, when children give an answer that cannot be categorized as the correct answer for the first time (i.e., response that triggers the "partially correct" intents, "don't know how to answer" intent, or the "fallback" intent), and the CA gives children a second chance by asking a less difficult question (e.g., "*The bears were afraid of their mama because they broke something. What did the bears break?*"). When children's response is not categorized as correct answer a second time, the CA provides scaffolding by rephrasing the question in a multiple-choice format (e.g., "*Did the bears break their mama's blue seashell or did they break a honey jar*?"). In the case of a third failure to provide a correct answer, the CA explains the question and then continues the story. This ensures that children can still enjoy the story even though they are not responsive to the prompts or fail to respond to the prompts appropriately.

**Table 1: Sample description. ESL stands for English as a second language.**

| | Age | Sex | Race/ Ethnicity | ESL | Prior CA use |
|---|---|---|---|---|---|
| P1 | 45m | M | Asian | Y | Y |
| P2 | 59m | F | Asian | Y | N |
| P3 | 72m | M | White | N | N |
| P4 | 59m | F | Hispanic | N | N |
| P5 | 65m | M | Hispanic | N | N |

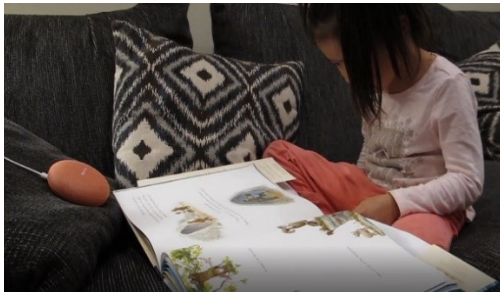

**Figure 2: A child reading with the storytelling CA**

## 4 METHOD

We recruited five children (Mean age = 60 months) from a suburban city in Southern California through convenience sampling (Table 1). Two children spoke a language other than English at home, but they were both fluent in English. Only one child had used voice devices at home. The studies took place at the child's or researcher's home. Children were read to by the CA via a Google Home Mini device while they looked at a hard copy book (see Figure 2). Children were responsible for turning the page to keep up with the narration. The reading sessions were video-recorded for an inductive analysis to generate patterns of the child-CA interaction. After the reading, children's perception was surveyed by the following three questions: "Is it fun to read with Google?", "Do you want to read the story with Google again?", and "How do you think we can make Google more fun?"

## 5 FINDINGS

### 5.1 Child-Agent Verbal Interaction

Most of the time, children in our study appeared to apply similar communication approaches as in face-to-face conversation. However, they occasionally used strategies atypical of face-to-face conversation to adjust for their communication with the CA storyteller. It is encouraging that the CA elicited natural communication, thus implying that children may view the CA as a social partner. The exceptions to this suggest that children may need to develop their schema for interaction with digital entities, as some human-to-human schema may not be suitable in the context of communication with CAs.

*5.1.1 Active response.* On average, children responded to 92.3% of the conversational prompts, indicating children's willingness to communicate with the storytelling CA. For most of the time, children directly answered the CA's questions. In the cases when children did not hear or understand the prompts or did not know the answer, they vocalized their confusion to the CA. For example, we observed that a child (P2) asked the CA to repeat a question by saying "*What? What did you say again?*" Another child (P4) told the CA that "*I don't really know (how to answer it).*" This finding is consistent with prior research on children's free interaction with CAs, which has suggested that CAs elicit utterances at a rate that is comparable with comparable to interaction with a human interlocutor [6].

In addition, a possible age difference was observed regarding the response rate. Four older children responded to all prompts, while the youngest one (P1) had a lower response rate of 61.5%. The older children vocalized their confusion, while P1 expressed confusion using non-verbal cues, such as body movement (shrugging shoulders) or facial expression (frowning). This may be due to younger children's less proficient oral language skills. Another possible explanation may be that younger children intuitively interact as they would in face-to-face conversation rather than intentionally adjust their responses contingent on the nature of digital interlocutor.

**Table 2: Counts of fallback instances for each participant**

| | P1 | P2 | P3 | P4 | P5 |
|---|---|---|---|---|---|
| No input | 5 | 0 | 0 | 0 | 0 |
| Mispronunciation | 1 | 0 | 0 | 1 | 0 |
| Outside of intents | 0 | 2 | 1 | 3 | 2 |
| Total | 6 | 2 | 1 | 4 | 2 |

*5.1.2 Affective reaction to feedback.* The storytelling CA was designed to give specific feedback for children's responses. The CA first plays a sound effect (rising sound for correct answer and falling sound for wrong answer) and then tells children if they get the answer right or wrong (e.g., "*Wow, you're right*!" or "*Hmmmm, it doesn't sound quite right.*"). We observed that children reacted to the feedback accordingly. They showed excitement after getting positive feedback (e.g., by laughing, cheering, clapping, dancing) and showed regret when hearing negative feedback (e.g., by frowning). Interestingly, we observed that children were especially excited with the correct answer feedback if they had answered an earlier question wrong.

The CA also provided an explanation as to why the answer was right or wrong. Children appeared to be more attentive to the explanation for the questions they answered incorrectly.

*5.1.3 Receptive response to scaffolding.* Our observations demonstrated the value of scaffolding within the conversational prompts. When children's utterances failed to match any pre-defined intents, the fallback function was triggered, in which the CA re-prompted children to answer the question by constraining the possible answers. The action is intended to simulate the scaffolding that an engaging adult would provide to children when they need help.

Fallback occurred three times per child on average, mostly because children mispronounced a word, responded outside of the anticipated intents, or did not provide any voice input (P1 only). Table 2 displays the counts of fallback instances for each child. The youngest child P1 appeared to show different patterns from other older participants; while most of P1's fallback instances were triggered by no voice input, other participants encountered the fallback mode mostly due to their responses falling outside of intents (e.g., asking clarification question, giving unanticipated response). We also observed two instances of fallback triggered by mispronunciation.

Follow-up questions appeared to successfully prevent conversational breakdowns. In particular, these questions worked well to scaffold children's pronunciation. For example, we had one question that asks, "*What is the island they need to find shaped like?*" with the correct answer being something related to a hat shape. When P1 responded, the */a/* was slightly mispronounced and the server recorded the response as "a heart" in the dialogue log. As the follow-up, the CA rephrased the question to a multiple-choice format, modelling how to pronounce the term "hat" by saying, "*Is the island they need to find shaped like a **hat** or is it shaped like a crown?*" P1 pronounced "hat" accurately this time. In addition, follow-up questions also appeared to scaffold children's comprehension. For example, some children did not answer questions directly; rather, they asked a clarification question. In responding to the question "*Where did the bears find the blue seashell*?", P3 asked a clarification question, "*Is it the place the old bear asked them to go*?" The CA followed-up with "*Where did the bears find the blue seashell? Did they find it on an island or did they find it under the sea*?" P3 responded appropriately to the follow-up question by selecting one of the two options, "*under the sea.*"

*5.1.4 Learning the turn-taking schema.* Turn-taking, which organizes speakers so as to minimize overlap, is universal in human verbal communication [5]. In human-CA communication, the turn-taking schema is even more prominent: while human speakers can (although rarely) overlap with each other while talking, CAs do not currently allow this flexibility.

As such, awareness of conversational timing is especially important for interacting with CAs, given that CAs can only listen to utterances during a specific, predefined time window (i.e., after the CA finishes talking and gets ready to listen). We observed that children had difficulties in determining the appropriate time to start speaking. This was especially the case during the first few rounds of question and answer.

*Rushed response.* In human-to-human communication, the amount of time between turns is very small, generally less than a few hundred milliseconds [5]. Therefore, the next speaker typically begins planning for their next utterance before the previous speaker has finished. In our study, we observed that children exhibited a similar pattern, as they responded immediately after the CA completed its question. This type of rushed response creates an issue, as the CA requires some time to switch from the speaker (to ask the question) to the listener (to record the child's utterance). As a signal that the microphone is ready, Google Home's four flashing lights prompt children for a response. However, this light signal may be too subtle for young children to notice, as 40% of our first two participants' responses occurred before the CA was ready. We then emphasized the prompting signal by adding a sound effect along with the flashing lights. Although the instances of rushed responses substantially decreased for our last three participants, these participants still sometimes responded before the prompt signal (about 20% for each child). This might be due to children's still developing executive functioning and inhibitory control.

*Interrupting the narration.* We noticed that two children each attempted once to talk with the CA during a pause at the end of a sentence but before the CA asked a question. They either made comments about the story (as P3 saw the thunder image on the book, he asked, "*Can you make the thunder sound?*") or brought up irrelevant topics (P2: "*My name is Oprah, what is your name?*"). As CAs have not been designed to handle this form of interruption, our CA simply ignored children's comments and continued the narration. Children in our study did not show frustration when the CA did not react to their comments, and they also appeared to quickly learn that they cannot disrupt the narration – neither of the two children that interrupted the CA once did so a second time.

*5.1.5 Adjusting for communication uncertainty.* Conversational turns in face-to-face contexts switch rapidly [5]. However, in the context of communication with CAs, there is typically a short delay between the child's utterance and the CA's response, allowing CAs to process the speech data. We observed that children are not accustomed to this delay between turns. They often repeated themselves multiple times until the CA responded. This repetition, which is not common in children's interactions with human partners, implies that children may feel uncertain about whether the CA has heard their responses. Such uncertainty may be due to the fact that the CA cannot provide other social cues (e.g., eye-gaze, facial expression) through which children can infer that their responses have been heard. In addition, children may have learned from their experiences that a human partner is more trustworthy in understanding their utterances than are inanimate objects. As such, the children repeating themselves to the CA can be viewed as an approach to adjusting their interaction with a digital partner.

### 5.2 Child-Agent Behavioral Interaction

Behavioral interactions were only occasionally observed. In our case study, children touched the Google Home device only when they were waiting for the CA's feedback for their responses. P1 nudged the device and said, "*wake up, wake up*". P2 held the device close to her mouth as she repeated her response before the CA reacted. We did not observe other affective behavioral interactions between children and the device. This pattern is different from that found in child-robot interactions where children frequently express their affect to the robots through haptic interactions such as hugging and patting. The unhuman-like appearance of Google Home may discourage children from demonstrating such prosocial behaviors.

### 5.3 Children's Perception

After the story reading session, we asked children about their perception of the storytelling CA. When children were asked whether they had fun reading, four children said yes verbally while the other one nodded. Three children also said that they wanted to read the story again. We then asked children, "Why do you think it is fun to read with Google?" Three participants brought up the idea that they liked talking to Google. For example, one child (P3) said, "I can talk to Google…it's so smart!" and another child (P4) said, "I don't feel bored because it (the CA) talks to me."

We also elicited children's suggestions for the storytelling CA. Children appeared to appreciate opportunities to ask the CA questions. For example, one child (P2) said that she wanted to be able to ask CA questions ("Why can't I ask Google question?"). This resonates with our observation that children sometimes brought up clarification questions. Children also suggested that they valued CA's accuracy in understanding speech. During the study, the CA failed to interpret P1's utterances, and he brought up that "Silly Google doesn't listen! Make it smarter?" This echoes two other studies finding that children view CA's accuracy as an important feature [7, 8].

## 6 DISCUSSION AND CONCLUSION

This project presents an application for leveraging CAs for young children's storybook reading. Through this case study, we identified the affordances and obstacles of a storytelling CA, and thus offer design implications for the future development of this type of product as well as for other voice-based apps for young children.

- Our observations suggest that the current accuracy of speech recognition and intent detection for young children is satisfactory, especially for children four years of age or older. This may render a CA effective as a reading partner for children, with the goal of enhancing engagement, comprehension, and oral language skills.
- A CA that provides scaffolding with more accessible follow-up questions appears to work well to avoid communication breakdowns and optimize the co-reading experience.
- Specific feedback that shows CAs have accurately understood children's utterances appears to be helpful to engage children.

- Future development of CA reading partners may pay special attention to children under the age of four, as interaction with CAs was especially challenging for this age group in our case study.
- Developers may want to consider ways to relax turn-taking restrictions by allowing a slight overlap between CAs' and children's speech.
- Allowing children to sometimes ask CAs questions may amplify the enjoyment of reading.
- Designers may want to build an embodiment of CAs (such as in a stuffed animal) to encourage greater child-CA empathy.

## REFERENCES

[1] Nian-Shing Chen, Daniel Chia-En Teng, and Cheng-Han Lee. 2011. Augmenting paper-based reading activity with direct access to digital materials and scaffolded questioning. *Computers & Education 57*, 2 (Sep, 2011), 1705-1715.

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

[3] Lisbeth A. Dixon-Krauss. 1995. Partner reading and writing: Peer social dialogue and the zone of proximal development. *Journal of Reading Behavior* 27.1 (Mar, 1995), 45-63. https://doi.org/10.1080/10862969509547868

[4] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 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*, 595-600.

[5] Levinson, Stephen C. and Francisco Torreira. 2015. Timing in turn-taking and its implications for processing models of language. *Frontiers in psychology*, *6*, 731 (Jun, 2015).

[6] Anuj Tewari and John Canny. 2014. What did spot hide?: a question-answering game for preschool children. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, 1807-1816.

[7] Julia Woodward, Zari McFadden, Nicole Shiver, Amir Ben-hayon, Jason C. Yip, and Lisa Anthony. 2018. Using co-design to examine how children conceptualize intelligent interfaces. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*.

[8] Svetlana Yarosh, Stryker Thompson, Kathleen Watson, Alice Chase, Ashwin Senthilkumar, Ye Yuan, and Bernheim Brush. 2018. Children asking questions: speech interface reformulations and personification preferences. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 300-312.