

# Are Current Voice Interfaces Designed to Support Children’s Language Development?

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

Stacy M. Branham  
sbranham@uci.edu  
University of California, Irvine  
Irvine, California

Xinwei Deng  
xinwed1@uci.edu  
University of California, Irvine  
Irvine, California

Penelope Collins  
p.collins@uci.edu  
University of California, Irvine  
Irvine, California

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

## ABSTRACT

With the rapid development of artificial intelligence, voice user interfaces (VUIs) capable of speech-based interaction are poised to support children’s language development by serving as their language partners. This paper reports an analytic evaluation of the currently available voice-based apps targeting young children to examine whether and how they incorporate evidence-based dialogue strategies that effectively support children’s learning. We found that, despite the fact that the current apps support a variety of language activities, most fail to carry out open-ended dialogue and provide extended back-and-forth opportunities, thus limiting their ability to encourage children’s language output and increase children’s language exposure. We discuss four design implications for developing VUIs that initiate dialogue and provide feedback in ways that better facilitate children’s language learning.

## CCS CONCEPTS

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

## KEYWORDS

Voice user interfaces, language development, young children, scaffolding

### ACM Reference Format:

Ying Xu, Stacy M. Branham, Xinwei Deng, Penelope Collins, and Mark Warschauer. 2021. Are Current Voice Interfaces Designed to Support Children’s Language Development?. In *CHI Conference on Human Factors in Computing Systems (CHI ’21)*, May 8–13, 2021, Yokohama, Japan. ACM, New York, NY, USA, 12 pages. <https://doi.org/10.1145/3411764.3445271>

## 1 INTRODUCTION

Young children develop their language skills through everyday communication with their parents, teachers, and other familiar

adults [39]. Decades of research has shown that extensive back-and-forth, contingent conversation is the fuel that primes children’s language learning (for reviews, see [26, 69]). With the rapid development of artificial intelligence, voice user interfaces (VUIs) are now capable of simulating speech-based, interpersonal interactions [49], and researchers believe these VUIs could play an important role in children’s language development by serving as their language partners [6, 68].

Smart speakers are one common form of VUIs that many children already interact with. These technological gadgets are becoming increasingly popular within hundreds of millions of households around the world. Having recognized the great potential of leveraging VUIs’ capacity to support young children’s language development, developers have made thousands of such voice-based apps available to children [55, 67]. These apps carry out dialogue with children based on pre-designed dialogical flows and engage children in focused activities with interactive speech-based content. A variety of activities are available through these apps, including playing a trivia word game, practicing spelling, and engaging in a choose-your-own adventure story [67]. These educational resources are especially valuable for preschool-aged children, who have not yet learned to read or write and thus primarily rely on oral communication [34].

Just like children’s interactions with parents or teachers, the language interactions carried out by voice-based apps should be thoughtfully designed in order to realize their intended educational and developmental goals [67]. While many HCI studies have proposed that VUIs’ conversation design should generally be grounded in human-to-human communication [15, 33, 44], these suggestions are often quite vague and not tailored for specific user groups [8]. Given children’s developing cognitive and language abilities, it is especially important that effective adult-child communication strategies be utilized when designing VUIs intended for use by young children [4, 9]. Nevertheless, little research has been devoted specifically to the communication strategies that ought to be incorporated in such apps.

In this paper, we seek to analytically evaluate how well the conversation design of voice-based apps appearing to target young children actually support their language development. We develop our evaluation framework based on the adult-child communication literature that identifies effective dialogue strategies adults can use to support children’s learning (for a review, see [57]). We then use



This work is licensed under a Creative Commons Attribution International 4.0 License.

*CHI ’21, May 8–13, 2021, Yokohama, Japan*

© 2021 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-8096-6/21/05.

<https://doi.org/10.1145/3411764.3445271>

this framework to evaluate 143 voice-based apps, all focusing on language development and available on the Google Assistant and Amazon Alexa platforms. Our first question examines the conversation design of the apps: *Do these apps exhibit evidence-based dialogue strategies?* Our second question seeks to understand the impact these apps' conversation design has for learning: *How effectively do the apps' dialogue strategies support children's language learning?*

## 2 RELATED WORK

### 2.1 Patterns of High-quality Adult-child Conversation

Decades of research has explored how adults support children's language development by engaging them in conversation during a variety of scenarios, such as storybook reading, game play, or simple day-to-day interaction. This body of scholarship has suggested that high-quality adult-child conversation is typically characterized by the adult assuming the role of a "language guide" who scaffolds the child's participation in the conversation. This kind of guided adult-child conversation generally follows a systematic structure referred to as the Initiation-Response-Feedback cycle. A typical conversation involves multiple cycles within which the adult *initiates* a topic of discussion and encourages the child's verbal expression. After the child *responds* to the prompt, the adult then provides *feedback* that builds upon the child's response. Sometimes the feedback completing one cycle leads into the initiation of the next cycle. In this way, the adult is responsible for moving the conversation toward a learning goal and helping shape the child's conversation experience. Many researchers have examined how adults' use of particular initiation and feedback strategies can promote children's language comprehension, vocabulary learning, and narrative skills (e.g., [17, 47, 59]).

**2.1.1 Initiation.** Initiation involves the adult asking the child a question that prompts active cognitive engagement and language production. Prompting strategies can be categorized based on the level of cognitive demand and the pattern of verbal response they elicit from the child. From this perspective, some prompts demand much from children in terms of cognition and language, and these prompts are typically open-ended in nature [41, 53]. Such prompts may invite children to recall an event (e.g., "Tell me an interesting thing that happened at school today!") or they may include a question beginning with "who," "what," "why," "where," "when," or "how". These types of open-ended prompts encourage children to practice expressive language to convey their thoughts [48]. Furthermore, the quantity, content, and form of language production resulting from such open-ended questions is frequently close to or at the upper limit of the child's competence [30]. For example, answering the question "what do you know about snow?" requires children to engage in information retrieval and processing before formulating a response in their own words, which is usually more linguistically complex than responses consisting of multiple words. More restrictive questions, on the other hand, do not demand as much from children in terms of cognition and language [7]. Such questions either pose options for children to choose from or logically and grammatically constrain children's responses. For example, a question asking "Is snow white or blue?" directs children to make a

choice between the two options. Similarly, asking children to "use one word to describe snow" limits their response to a single word. In general, the literature on adult-child conversation suggests that restrictive prompts are less beneficial for children's learning than are open-ended prompts [7, 14, 70], but also advises that adults use these two types of prompts in combination depending on individual children's language competence or topical knowledge [71].

**2.1.2 Feedback.** Feedback involves the adult verbally commenting on the child's response to the prompt. The literature describes four dimensions that can be used to categorize the adult's feedback.

*Contingency* refers to whether the adult's feedback is semantically tailored to the child's response. From the child's perspective, the adult's contingent feedback shows the adult's involvement in the conversation, their acknowledgement of the child's response, and their desire that the child reflect on that response (e.g., affirmation, correction) [25, 45, 63]. In these ways, contingent feedback supports the cognitive aspects of children's learning while also heightening their engagement in the conversation [25, 63].

*Extension* involves the adult deepening the conversation based on the child's response. Adults may elaborate on the topic by providing further explanation or new pieces of information. The education literature sometimes refers to this kind of feedback as "elaborative feedback", which contrasts it to feedback that merely verifies the child's response [38, 58]. A meta-analysis suggests that adults' use of elaborative feedback frequently results in the child speaking further on the topic at hand, thus leading to a more in-depth conversation [24].

*Encouragement* calls for children to complete a language task to their full potential. Some examples include encouraging children to express themselves verbally, pronounce words correctly rather than use baby talk, provide accurate or more complex answers, and clarify or elaborate on their responses [57]. It could also involve encouraging children to avoid distractions and stay focused on the language exercise [18].

*Adjustment* refers to the adult adapting their feedback to the child's developmental levels and the cognitive and linguistic demands the child faces during the conversation. Some strategies include the adult making their language more accessible to the child [16] or providing contextual or linguistic hints [1]. Many studies have suggested that adjustment strategies are especially beneficial for children who are still developing their language skills [31, 42].

In summary, the traditional literature on adult-child conversation has laid out a useful and precise framework for understanding how conversation partners acting as a "language guide" could effectively utilize initiation and feedback strategies to support children's language development. Our analysis of the conversation design within voice-based apps targeting young children's language learning is thus built upon this framework.

### 2.2 Children's Conversation with Voice Interfaces

The rapid development of AI-enabled voice interfaces is redefining what it means to be a language partner. Language partners are no longer restricted to adults but also expanded to VUIs with agents that are designed to comprehend complex speech input [34, 68].

In fact, many children already interact regularly with the smart speakers in their own homes, and researchers see this child-agent conversation as a fruitful addition to children’s everyday language experiences [5, 34]. An emerging body of literature using interviews, observations, diary instruments, and in-home audio recordings has documented two types of interactions children commonly have with VUIs: open-domain conversations with general assistant tools (e.g., Apple Siri, Amazon Alexa, Google Assistant) and agent-led context-specific conversations specifically designed for children (i.e., voice-based apps) [3, 5, 34, 35, 40, 61]. Of these two types of child-VUI interactions, the latter most closely resembles the dialogue structures of effective adult-child guided conversation we discussed in the above section.

In agent-led context-specific conversation, the VUI guides children along a previously designed dialogic roadmap for a conversation on a particular topic. [28, 67, 68]. Although such pre-programmed content offers less flexibility compared to general assistant tools, carefully designed agent-directed conversations can lead to a variety of child-focused educational experiences [23]. A handful of research projects have developed experimental VUIs that incorporate effective guided conversation strategies as identified in the traditional literature, in particular, the prompt-response-scaffolding cycle [66, 68]. For example, one study developed a storytelling app that asks children open-ended questions, provides feedback, and follows up on children’s inappropriate responses with helpful hints [68]. Children in this study were found to actively engage in the storybook reading, remain attentive to the agent’s explanation, and improve their answers based on the hints provided. Moreover, this agent-led dialogue during storybook reading was found to be as supportive as an adult reading partner in promoting children’s comprehension [65]. Another study implemented a phone-based VUI to teach French phonological awareness to children in a rural community. Each lesson began with explaining one phonological concept and then prompted children to respond to a relevant question. Children then received feedback depending on whether they answered that question correctly [37]. In addition to research-based voice-based applications still in the experimental stage, there are also a large number of consumer-facing apps already available to children. These apps can tell stories, teach vocabularies, engage children in trivia games, and talk about children’s daily experiences. While these apps have attracted millions of young users, whether they effectively incorporate good conversation design in pursuit of children’s language learning is still unclear.

Good conversation design is especially important if agent-led interactions are to support young children’s learning. On the one hand, young children are still developing their cognitive abilities, communicative skills, and mental schema to interact with a digital interlocutor [13, 62]. On the other hand, because a VUI’s “understanding” of what a child expresses is based predominately on the pre-designed dialogic roadmap, VUIs are not able to adjust the conversation flow as flexibly as a human language partner [64]. These two factors lead to child-agent conversations being prone to breakdown. Most studies documenting such conversation breakdowns have focused on how children attempt to adapt their communication strategies to avoid potential conversation breakdowns [4, 13].

However, very few studies have explored how VUIs could be proactively designed to provide richly scaffolded conversation that is also age appropriate.

Indeed, little research has examined the conversation principles guiding the development of such interfaces. To our knowledge, most existing guidelines are developed by major companies, such as Google, Amazon, Apple, and Microsoft, that market popular voice interfaces. Such guidelines are intended to support the development of VUIs targeting general users and scenarios. According to one analysis, existing guidelines collectively emphasize the importance of designing VUIs to replicate human-to-human conversation [8, 54]. However, one thing missing from this broad guiding principle is the understanding that, depending on the context of the conversation, different humans may have different needs. This lack of consideration for the needs of children may limit the usefulness of these guidelines for designing child-oriented VUIs.

### 3 METHOD

#### 3.1 App Selection

This paper evaluates voice-based apps targeting preschool-aged children’s language development. We identified these apps using the following procedure. First, we conducted a keyword search on the Amazon Alexa and Google Assistant platforms. Apps were initially included if their titles or descriptions contained any of these keywords related to language development: “story,” “language,” “literacy,” “book,” “reading,” “vocabulary,” and “spelling.” This preliminary search yielded a total of 1613 apps from the two platforms. Second, we excluded apps not intended for young children. Apps whose description suggested a target age range exclusively over six years of age were excluded first (n = 52). For the remaining apps, the lead author and a non-author licensed preschool teacher read the app’s description and evaluated the apps’ content to determine whether the apps were indeed intended for use by preschool-aged children. We thus excluded an additional 1174 apps due to their inclusion of advanced language skills, such as difficult vocabularies or complex sentence structure (n = 300), their difficult or age-inappropriate content (n = 452), or their assumption of prior knowledge or skills, such as the ability to spell multi-syllabic words, a skill young children do not typically possess (n = 370). Third, we excluded apps that did not initiate any dialogue with children (n = 252). Finally, we excluded apps that had technical issues, including being unable to open or crashing during use (n = 44). Our final analytic sample included 143 apps, with 42 available on the Google platform and 101 on the Amazon platform. The app selection occurred between December 2019 and January 2020. On average, apps on Amazon received 71 reviews, and those on Google received 31 reviews. Figure 1 illustrates the app selection process.

#### 3.2 Coding Protocol

**3.2.1 Dialogue Design Coding.** Our protocol for coding the apps’ dialogue designs was initially constructed around the initiation and feedback strategies established in the literature on adult-child communication, and then slightly modified based on our actual testing of the apps. The specific coding protocol is described below.

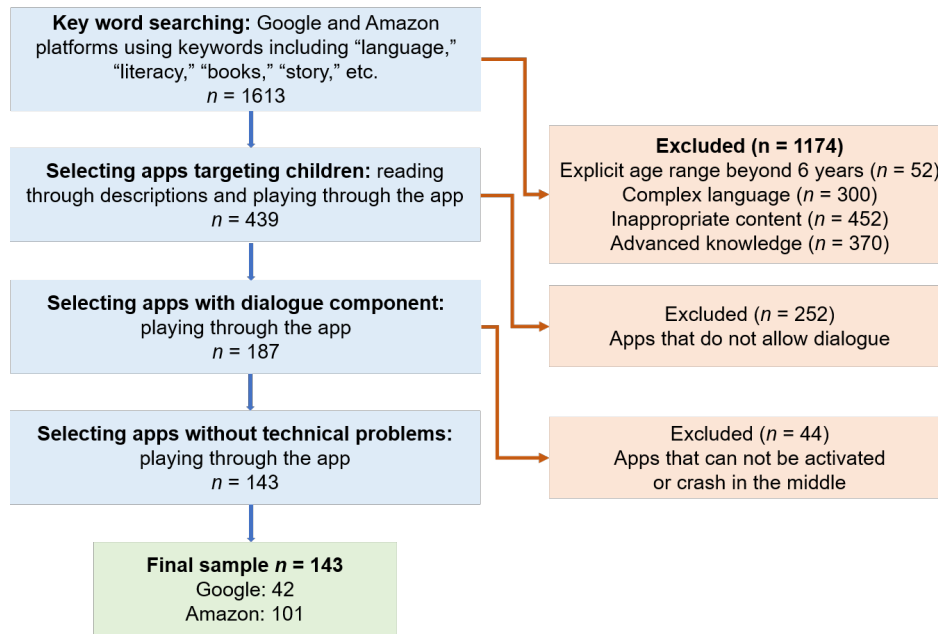


Figure 1: App Selection Process

We first coded the voice apps' use of prompting to **initiate** conversation. Two types of open-ended prompts were coded: *wh*-questions and free response prompts. *Wh*-prompts begin with who, what, why, where, when, and how and typically aim to elicit children's responses on some specific topic; *free-response prompts* invite children's completely free expression without any directives. Three types of restrictive prompts were also coded: repetition, one-word response, and multiple-choice prompts. *Repetition prompts* ask children to repeat a particular word or sentence. *Multiple-choice prompts* ask children to choose an option from a list of two to three choices. *One-word response prompts* ask children to provide one word within a particular domain (e.g., one color). We primarily relied on the categories found in the literature on parents' and teachers' question types for preschoolers [59, 73]. In addition to these predefined categories (i.e., *wh*-prompts, free-response prompts, multiple-choice prompts, repetition prompts), we added "one-word response prompts" to cover a type of prompts that emerged from our actual coding but did not seem appropriate to the predefined categories.

We then coded whether the apps' **feedback** to the child's response included the following features: contingency, extension, encouragement, and adjustment. *Contingency* was coded as present if coders determined that the feedback specifically addressed the child's response in a logical way instead of merely providing generic or unrelated feedback that was not based on the child's response. *Extension* was coded as present if the feedback contained additional information instead of merely verifying the child's response. *Encouragement* was coded as present if the feedback following a child's non-response encouraged the child to respond to the prompt. *Adjustment* was coded as present if the feedback scaffolded a child's off-topic or otherwise inappropriate or incomplete response. These

four binary feedback dimensions (i.e., whether contingency, extension, encouragement, and adjustment is present or not) are predefined based on the literature described in the Related Work section.

**3.2.2 Language Activity Coding.** To examine how well the app's agent-led dialogue supports children's language learning, we also categorized the types of activities the apps could carry out. This protocol was based on the common language activities children engage in with their parents or teachers identified in the literature [26] and modified based on our own trial of the apps. We coded four types of activities: *story listening activities* that include narrative-rich elements and story sequence, *trivia games* that ask children questions on a specific topic, *lessons* that involve explicit instruction aiming to relay knowledge, and *story creation activities* that prompt children to provide specific story elements in order to form a customized story.

**3.2.3 Coding Procedure.** Two non-author coders were involved in the coding process. The coders played through each app, recording how the app initiates dialogue and provides feedback to three different types of responses that sufficiently covered how an app would typically categorize a child's response:

- appropriate responses, either correct or incorrect, that are semantically relevant to the prompt,
- random responses that are clearly off-topic, and
- silence if the user does not provide any verbal response to prompt.

The adult coders used this same set of structured response to test all apps. This allowed us to comprehensively examine the apps' dialogue flow and also enabled cross comparison among the apps [32]. Inter-coder reliability was established by first jointly

conducting a pilot coding sequence, discussing where certain app features would fall within the coding framework. Then, each coder conducted additional pilot coding individually before eventually comparing that coding and resolving any discrepancies. Finally, the two coders periodically coded a new app together to further calibrate their coding. This allowed the inter-coder reliability to reach 0.80 to 0.85 across categories.

In addition to coding based on the calibrated protocol, the coders also transcribed their interactions with each app and recorded qualitative notes to provide contextual evidence for understanding the conversation design of each app. The coders periodically compared notes to ensure that the notes accurately reflected their interactions with the apps.

The coding was carried out during four months from January 2020 to April 2020.

## 4 RESULTS

In this section, we first detail how the apps utilized initiation and feedback strategies and then describe the dialogic patterns observed across the voice-based apps, discussing the impact each dialogic pattern has on children’s language learning. Table 1 displays the result of each coding category.

### 4.1 Initiation and Feedback Strategies

Coding item	Count	Percentage
Initiation		
Wh- prompts	16	11.2%
Free-response prompts	11	7.7%
Multiple-choice prompts	85	59.4%
Repetition prompts	10	7.0%
One-word-response prompts	21	14.7%
Feedback		
Contingency	110	76.9%
Extension	75	52.5%
Encouragement	88	61.5%
Adjustment	45	31.5%

**Table 1: Summary of Coding Results**

**4.1.1 Initiation.** We found that the overwhelming majority of apps initiated dialogue cycles using prompts that restrict either the length or scope of children’s responses (i.e., multiple-choice prompts, repetition prompts, and one word response prompts). Altogether, 116 of the 143 apps (81.1%) utilized these restrictive prompts. Specifically, **multiple-choice prompts** were found within 85 apps (59.4%). For example, the app “Kids Quiz” assessed children’s vocabulary associated with plants by asking “*The leaves on a tree are attached to which part of the tree? You can answer branches, roots, or the trunk.*” Another app “My adventure book” paused at a particular point during a story and asked children a multiple choice question “*Do you choose to search inside the apartment or look for evidence outside*” to allow them to decide what happens next. **One-word-response prompts** that explicitly required children to limit their response to one word or to a very short phrase were found within

21 apps (14.7%). This type of prompts were commonly used among apps that focus on vocabulary. For example, an app “Opposite” asked children to provide the antonym of a given word (e.g., “*What is the opposite word of new?*”), and another app “Hangman Game” resembled the classic game that allows children to guess a word letter by letter by saying “*Try a letter*”. One-word-response prompts were also used to help children identify parts of speech, such as “*give me a noun*” or “*give me an adjective*”. **Repetition prompts** were found within 10 apps (7.0%). These prompts, which asked children to repeat a sentence or phrase, can be used to confirm children’s pronunciation (e.g., “*In this game, you must repeat what I say. Talk clearly. Here is your first question: school, spoon, food*” in the app “English One”). Repetition prompts were also used to help children reinforce a piece of information or highlight a particular entertaining part of the dialogue. For example, an app “Oliver’s Adventure” invites the children to participate using requested phrase to help the main character Oliver. In a moment when Oliver is in danger in the story, the narrator asked children to repeat “*Sound the alarm*” loudly so that Oliver can be rescued. After children repeated the phrase, the narrator replied “*Olivia really appreciates your help!*” and continued the story.

By contrast, relatively few apps (18.9%) implemented open-ended prompts (i.e., wh- prompts and free response prompts) that did not present any response options a child may choose from. **Wh-prompts** were only found in 16 of the apps (11.2%), and the overwhelming majority of them (n = 13) including wh- questions that asked for a matter-of-fact response with one single correct answer and could be sufficiently answered with a short phrase. For example, an app “Super Animals” played the sound of different animals (e.g., an elephant squeak) and asked children, “*what is the animal?*” Another app “ABC Bee” asked children “*how to spell*” different words. There are only three apps utilizing wh- prompts to elicit a more personal, and potentially more complex, response that without a correct or incorrect answer. For example, one app “Imagine Christmas” narrated a Christmas story and asked children how they would like to welcome Santa (“*What do we need to do to get ready for Santa?*”). Another app “Conversation Buddy” engaged children in discussion of their hobbies by asking “*What is your hobby?*” and “*What makes you like your hobby?*” **Free response prompts** were only found in 11 apps (7.7%). This group of apps typically invited children to ask a question or say whatever they want. For example, some apps allowed children to request translation from English to a different language by saying a word, a phrase, or a sentence (e.g., “*Say a word or a phrase you’d like to translate*” in the app “Pig Latin Translator”). Other apps invited children to say any sentence and then repeat that sentence with a different tone (e.g., “*I can repeat your input. What do you want me to say?*” in the app “Repeat After Me”).

**4.1.2 Feedback.** In terms of **contingency**, the majority of apps (n = 110, 76.9%) provided some form of feedback that was contingent upon children’s responses. More specifically, these apps first performed semantic-based language understanding, then classified children’s responses based on a predefined intent, and finally selected feedback corresponding to that specific intent. Contingent feedback was sometimes used to assess the child’s response. For example, an app may praise a child for giving the correct answer to

a spelling question by saying "*You are very good at spelling!*" Other contingent feedback repeated the exact word or phrase a child had just said. Such repetition can be used to verify a child's response or demonstrate that the app correctly registered the response. Contingent feedback was also associated with prompts that helped control the dialogue flow. For example, some apps intermittently paused the narration and asked children whether they were ready for new information or what information they would like to receive. Then the narration moves forward contingent upon children's response.

The apps whose feedback lacked contingency did not contain the semantic-based language understanding component necessary to provide contingent feedback. Instead, these apps provided feedback from a pool of pre-set utterances that were usually designed to be more general than specific since general feedback helps minimize any disconnect within the dialogue. Apps that asked open-ended questions (i.e., wh- prompts and free-response prompts) are less likely to be able to provide contingent feedback than those asking restrictive questions. This suggests that the decrease in predictability of children's responses to such prompts increased the challenges of providing contingent feedback. For example, one app which asked children what their hobby was provided the exact same feedback ("*Really, that sounds like a lot of fun!*") regardless of the child's reply. Similarly, another app mimicking the voice of Pikachu, a well-known cartoon character, replied to every response with the word "*Pikachu*" in randomly varying intonations<sup>1</sup>.

In terms of **extension**, about half (n = 75, 52.5%) of the apps provide feedback that contains additional information expanding on the child's responses. For example, one trivia app asked children, "*What is the sixth planet from the Sun that features an extensive ring system?*" After children answer "Saturn," the app not only acknowledged the children's correct answer but also told the children, "*Saturn has the most spectacular ring system that can easily be seen from earth with a good telescope.*" Sometimes, yet quite rarely, feedback with extension not only references the child's most recent response but also carries the conversation forward. For example, one app asked children if they had ever been to a zoo. Regardless of whether the child answered yes or no, the app repeated their response and told children that most zoos have peacocks. The app then asked children if they knew what a peacock's tail looked like. In this way, the feedback created a dialogue cycle that led to more extended discussion about the topic related to zoos and animals.

In terms of **encouragement**, 88 of the apps (61.5%) in our sample provided some kind of encouragement that attempted to draw out verbal expression from hesitant children. Of the remaining apps in our sample, 51 (35.6%) simply waited for children to respond and then automatically shut down after several minutes. Another 4 apps (2.8%) waited a few moments before moving on to a new cycle without addressing children's lack of response. For the 88 apps that did provide encouragement, one common type was simply nudging children to say something. For example, some apps said "*Try again*" or while others said "*Answer the question*" and then told children that any response was acceptable. Most of these nudges are generic and not connected to the specific context of the app, with the exception found in an app called "Smallfoot". This app

featured an expressive monster who roared loudly, and if a child did not respond, the monster would encourage children to respond verbally by saying, "*Woohoo, but your voice is kinda hard to hear. You are like 'meee meeee mee meeee meeee...'*".

In terms of **adjustment**, fewer than one third of the apps (n = 45, 31.5%) incorporated feedback that modified or reworded the follow-up questions when children failed to give an appropriate response. The remaining apps simply repeated the question with the exact same wording. Feedback with adjustment may be helpful for children who have difficulty answering a question as originally stated. Such adjustment typically involves providing hints before re-asking the same question in a slightly different way. For example, one choose-your-own-adventure story app asked children "*Should we walk to the boat, follow the eagle, or go towards the music?*" If the child did not provide an answer that contains any of the three options, the app explained what each option entails (i.e., "*We could walk to the boat to find the tropical island, follow the eagle to get the north fruit, or follow the music to find the haunted lighthouse.*") and then again asked the child to make a selection.

## 4.2 Dialogue in Language Activities

We now turn to an examination of how the Initiation-Response-Feedback cycles are woven into different types of language activities and how these dialogic patterns function to support children's language learning. Within our sample, the four most common types of language activity were story listening (n = 58, 40.6%), trivia games (n = 40, 28.0%), lessons (n = 25, 17.5%), and story creation (n = 8, 5.6%). The ten remaining apps represented a mixture of miscellaneous activities that could not be grouped into these categories.

Representative examples of each type of apps type are presented. To increase the practical implications of our findings, we selected dialogue typical for each type of activity from the apps that had more user reviews because those apps are likely to be more frequently used.

### 4.2.1 Story listening: Allowing autonomy in narrative sequence. An excerpt from "Animal Farm P.I.", a story listening app

- App: It was a bright and chipper morning at the farm. Old Man MacDonald was starting up the tractor for some field work.
- App: Would you like to go to the barn, the pond, or the kitchen?
- User: Kitchen
- App: You snuffle your way into the kitchen and find a cell phone on the counter and a refrigerator that looks like a very promising place to find some sauerkraut.
- App: Would you like to call somebody or look in the fridge?
- User: Look in the fridge
- App: You open the refrigerator and see a plastic container that looks like it's got some sort of fermented cabbage in it. That fermented cabbage may very well be the sauerkraut you're looking for.

<sup>1</sup>Until very recently, the real Pikachu character did not speak either but only used non-verbal sounds, 'Pika', and 'Pikachu'. This non-contingent feedback is actually relevant to the character

App: Would you like to try the cabbage or close the fridge?  
 User: I wanna try something else.  
 App: I don't recognize that response.  
 User: Something else.  
 App: You open the container of fermented cabbage and hear the reassuring sound of gas hissing out. Only one way to make sure this is sauerkraut. You take a nice healthy bite. Yowzer! Okay that's definitely not sauerkraut. It tastes more like kimchee. You find it delicious even though it gives you gas. You should probably look somewhere else.

The rich narrative elements of story listening apps distinguish them from the other apps. Dialogue moments are included throughout the narrative with the primary goal of evoking children's interest in the story. These apps typically use dialogue moments that allow children to influence the story's trajectory, mirroring the popular choose-your-own-adventure story format.

In an excerpt taken from "Animal Farm P.I." (available on Google), the app tells a story set on Old Man MacDonald's farm. The children are asked multiple-choice questions that will determine how the story proceeds. The feedback to the child's response is used to carry the story forward, and future questions are logically connected with the child's previous responses. However, all of this takes place within a pre-determined set of storyline branches that children must navigate through. As shown in the last Initiation-Response-Feedback cycle in the excerpt, the app is unable to extend the story if the child's response deviates from the pre-determined storyline. Instead, the app randomly selects one of the pre-determined storyline branches to move forward the story. Despite children's responses being limited by the available storyline branches, the dialogue moments provide children with some level of autonomy during their story listening experiences.

**4.2.2 Trivia games: Providing opportunities to demonstrate existing knowledge.** Trivia apps' initiation typically elicits a direct response to a question involving only one correct answer. Such apps usually provides contingent feedback that simply assesses whether a child's answer is correct or incorrect.

**An excerpt from "Lucky Girl", a trivia app**

App: In this game, I'm going to ask you questions about animals! What is the baby animal name for a pig?  
 User: Piglet.  
 App: Outstanding!  
 App: Let me give you another one. What is the baby animal name for a cat?  
 User: Kitten.  
 App: Exactly! Great job!  
 App: Let me give you another one. What is the baby animal name for a chicken?  
 User: Duckling.  
 App: False. The answer is chick.  
 App: Let me give you another one. What is the baby animal name for ants?

User: I don't know.  
 App: This is not the answer. Try again.  
 User: I really don't know.  
 App: This is not the answer. The answer is antling.

An excerpt from a trivia app "Lucky Girl" (available on Google) shows four Initiation-Response-Feedback cycles. Each cycle begins with the app asking "what" question having a clear directive constraining the scope of the child's answer. After each user response, the app explicitly verifies whether that answer is correct or incorrect; only rarely did a trivia app extend the dialogue with elaborative feedback. The final cycle in the excerpt above is representative of the typical trivia app's adjustment and encouragement strategies. The app simply interprets "I don't know" as an incorrect response and replies, "this is not the answer" before asking the children to try again. Such an approach would not be helpful for encouraging learning among children who do not already have that specific knowledge.

Trivia apps also do not generally adjust their difficulty level according to the child's competence. The questions neither increase in difficulty as children successfully answer the questions nor decrease in difficulty if children struggle to answer them. Furthermore, the feedback within one cycle rarely connects to the next cycle in a conversational way. Cycles are typically organized as a series of independent, disjointed questioning moments with little or no sequential order. What's more, trivia apps rarely contain anything other than these cycles. Taking together, these dialogue structures suggest trivia apps simply assess what children already know and are not particularly focused on teaching them new information.

**4.2.3 Lessons: Teaching new knowledge through reinforcement.** The dialogue cycles of lesson apps resemble those of trivia apps in that the apps initiate a series of prompts to elicit children's responses and then provide feedback assessing whether children's responses were correct or incorrect. One major difference in lesson apps is that they assume children do not already have the prior knowledge needed to correctly answer the questions. Some lesson apps teach children particular bits of knowledge and then quiz them on what they have just heard, while other lesson apps simply instruct children to repeat whatever the app has said and then evaluate whether the child's repetition was accurate.

**An excerpt from "Spanish", a lesson app**

App: The new word for today is "pequeña", it means "small". Say "small" in Spanish.  
 User: pequeña  
 App: Well done!  
 App: Now I will ask you about five Spanish words, one from today and four from previous days.  
 App: Buena suerte  
 User: Good luck  
 App: Good job!  
 App: pequeña  
 User: I don't know  
 App: pequeña means small. Say small in Spanish  
 User: pequeña

This particular app ("Spanish" available on Amazon) consisted of two distinct groups of Initiation-Response-Feedback cycles, an instruction component that teaches children the content (i.e., the meaning of "pequeña") and a test component that assess children's memorization of the content taught previously. The instruction component of this app, like other lesson apps, involves direct, didactic instruction where the agent explicitly provides children with new information. The testing component functions similarly to the trivia apps, except that it provides direct instruction if a child fails to answer a question about content the app previously taught.

**An excerpt from "Language Tutor", a lesson app**

App: I will ask you to say something. Just repeat what you hear.  
 App: The doctor was at the hospital.  
 User: The doctor was at the hospital.  
 App: Well done!  
 App: There were many cars in the street.  
 User: Cars  
 App: There were many cars in the street.  
 User: Cars in the street.  
 App: There were many cars in the street.  
 User: There were many cars in the street.  
 App: Perfect!

This is another example that shows how apps were designed to leverage VUI's speech recognition functionality to provide children with pronunciation training. This excerpt was taken from "Language Tutor" available on Google. The initiation of each cycle is a sentence or phrase children should model their response on, and after children utter the sentence, the app tells them whether their pronunciation was accurate. When a child's repetition fails to match the original sentence, the app repeats the sentence without providing any explanation until the child successfully utters the sentence.

*4.2.4 Story creation: Allowing space for a child-initiated storytelling.* The dialogue moments in story creation apps allow children to create a silly story, much like a "madlib" game, by contributing key elements of the narrative. After asking the child to provide a series of words, the apps use those words to construct the story.

**An excerpt from "Kids Mad Libs", a story creation app**

App: I will ask for some words and I will make a story. Put on your thinking cap. This story uses 16 words. First, an adjective.  
 User: Beautiful  
 App: (Ding, and laughing) 15 words to go.  
 App: Say an animal  
 User: Monkey  
 App: (Ding) You're funny.  
 App: I need a noun.  
 User: Red.  
 App: (Ding) Wow, you are great with words.  
 App: I need an adjective.  
 User: ...  
 App: Any adjective will work.  
 User: ...

App: Okay, let's say funny.

App: Okay, here we go. The title of this is "Driver's checklist". Before you follow the open road to a **beautiful** adventure, don't forget to have a good **monkey** to give you a **red** of a **funny** tuneup.

As shown in the excerpt below, each cycle begins with a request for a word that fits into a particular category. The excerpt was taken from "Kids Mad Libs" available on Amazon. The word categories are usually broad and include grammatical categories (e.g., noun, adjective) and domain categories (e.g., animal, number). If a child provides any verbal response, the app makes a positive sound and praises the child using one of several pre-determined generic phrases. Such positive feedback is given even when the child provides an inappropriate response (e.g., providing an adjective instead of a noun). If the child does not provide a verbal response, the app nudges the child to respond but does not clarify the request. If the child fails to respond a second time, the app chooses a response for the child. Taken together, it is clear that the purpose of the app's feedback is to garner participation rather than address the child's actual response or any potential confusion they may have.

After the Initiation-Response-Feedback cycles are complete, the app inserts the child's responses into a story template. Given the broad range of possible words children may provide for a particular request, the story created is usually entertaining because of how unrealistic it is. However, because the app merely inserts whatever the child has said into the story, the story could also contain grammatically errors, exposing the child to poorly constructed language.

## 5 DISCUSSION

### 5.1 How Can VUIs Be Better Designed?

Four design implications arise out of our analysis of how voice-based apps are designed to initiate dialogue with children and provide feedback to their responses.

First, when designing voice-based apps for children, **it is valuable to incorporate open-ended questions while also preserving the ability to provide meaningfully contingent feedback.** In examining the apps' initiation strategies, we found that the overwhelming majority of apps utilized restrictive rather than open-ended prompts. Relying too heavily on restrictive prompts fails to elicit children's responses at the upper limit of their language competence and may result in responses that are semantically and/or syntactically simplistic[30]. For example, multiple-choice prompts that provide a child with two or three options limit the child's opportunity for free recall. Similarly, prompts that call for a single word response and prompts that ask the child to repeat a phrase already included in the prompt restrict the child's flexibility in generating their own response [41]. However, the fact that open-ended prompts increase the unpredictability of children's responses makes it more challenging to design a VUI capable of interpreting possible user responses and providing contingent feedback [43]. Indeed, in our analysis, we found a clear trade-off between the open-endedness of a prompt and the contingency of the app's feedback: as the the question's level of open-endedness increases, the degree of feedback contingency decreases. To help mitigate this



issue, developers should carry out thorough field-testing of their dialogue script with target users to gather a wide range of possible response variations [46]. They should then develop and refine their dialogue tree to incorporate feedback that is specific to the range of responses actually provided by their target users. However, even with this effort, we should still expect that children who ultimately interact with the VUI may provide responses that the VUI is not prepared to respond to. As such, it is actually valuable to include a restrictive question as a follow-up prompt for when the child responds to an open-ended question in an unanticipated way [2]. Indeed, a study has confirmed that follow-up restricted prompts resulted in the VUI's higher rate of intent detection than open-ended prompts, largely due to the fact that more restrictive questions reduced the likelihood of a child providing unanticipated answers [68].

Second, voice apps designers should **incorporate feedback that extends a child's utterance to enable more in-depth multi-turn dialogue on a single topic**. Half of the apps we examined did not reply to children in a way that extended on their response and thereby missed a key opportunity to deepen the dialogue [22]. One consequence of the lack of extension in the apps' feedback is that the dialogue in a previous Initiation-Response-Feedback cycle is not connected to the following cycle. These types of disjointed cycles usually suggest dialogue that lacks the "progressivity" discussed by Fischer et al. and thereby limits the system's capacity to move the dialogue forward along a meaningful path [21]. To provide children with opportunities for in-depth conversation, VUI developers may want to consider strategies for extended dialogue that have been established in the research on intelligent tutoring systems. One useful model is Graesser et al's two-part feedback frame consisting of a short comment on the child's preceding turn followed by an additional "dialogue advancer" [50]. These dialogue advancers moves the conversation forward with helpful hints, assertions with accurate information, corrections of children's misconceptions, or answers to children's questions [27]. While many of the app we examined already implement the initial short comments, few of them included any form of dialogue advancer. The purpose of such dialogue advancer is to invite the child's response by ending the app's turn with a question or other cue so that the child is guided into specific further participation.

Third, it is important to **utilize design strategies that encourage children's language production, in particular, for those children who are hesitant or unwilling to respond**. Past research on child-VUI interaction suggests that children frequently avoid responding to conversational prompts that they perceive to be too challenging [66, 68]. Children also have a tendency to rely on non-verbal expressions (e.g., gesture [68]). Therefore, app designers should carefully consider the most appropriate way to respond to a child's lack of verbal input. Almost half of the apps we examined simply terminated the dialogue if a child did not respond, despite this type of breakdown being easily prevented by providing the child with additional opportunities to respond (e.g., repeating the question, nudging like "Try again"). However, many of these approaches may not be ideal for child users. Nudges like "Try again" or "Answer the question" are demands which may dampen children's enjoyment from interacting with the VUI or intimate an already hesitant child [68]. Similarly, simply repeating the question

may have limited usefulness if a child did not respond previously because they did not know how to answer. As such, it is important to recognize that the purpose of encouragement is not simply to solicit a child's response. Encouragement should provide sufficient scaffolding to draw children's attention to the task at hand in an enjoyable way and help children participate in the dialogue. For example, apps could encourage a child to respond by using a more friendly tone (e.g., "I am so curious about what your answer is" instead of "Answer the question."). Also, more apps could leverage humor that is connected to the dialogue's theme (as in the app "Smallfoot" we discussed prior). And it may be helpful to rephrase a question to make it easier for the child to understand. Indeed, a prior study has found that using an easier, rephrased question as a follow-up effectively invited responses from those children who did not respond the first time [68].

Fourth, developers should **recognize the need for more nuanced scaffolding that adjusts to children's differing response patterns**. In general, the apps we examined were only capable of detecting one correct answer for any given prompt. All other verbal responses were then lumped into one category (i.e., failure to answer correctly) and received the same feedback. However, children may answer a question incorrectly in many distinct ways, each of which may reflect a differing level of language competence and understanding of the topic [72]. Moreover, children who answer "I don't know" may have a better meta-understanding of their own knowledge than those who provided an incorrect answer [52]. As such, children would benefit from more specific, tailored scaffolding depending on the type of incorrect responses they give. To achieve this goal, developers first need to generate a fine-grained dialogue tree (as discussed in our first point) so that children's incorrect responses can be classified precisely. Based on this classification, scaffolding messages specific to each type of incorrect response can then be developed. In one study on children's interaction with a storytelling VUI, the authors categorized children's most common incorrect answers into multiple categories and also distinguished children's "I don't know" response from other incorrect responses, enabling the VUI to provide tailored scaffolding (e.g., providing hints when children acknowledge their inability to answer a question) that promoted children's comprehension of the story [68]. However, the way incorrect answers were classified depended on the specific question being asked. This suggests that tailoring of VUIs' scaffolding could be a challenging endeavor that requires much time, effort, and expertise.

## 5.2 True Conversation or Pseudo-Interaction?

Many of the design challenges discussed above are reflective of the general limitations inherent to VUIs. For instance, a VUI's dialogic interactions are necessarily pre-defined, and the success of a child's interaction with that VUI depends partially on the child responding in one of the ways predicted by the VUIs designers. If children respond in an unpredicted manner, the VUI is not capable of providing coherent feedback. As such, we must ask whether such interaction can really qualify as "conversation." [29] The literature on inter-personal communication clearly defines conversation as involving multiple parties "mutually orienting to, and collaborating in order to achieve, orderly and meaningful communication" [19]. This

definition emphasizes the necessity of collaboration between the conversation partners. In other words, actual conversation should involve all parties simultaneously influencing and being influenced by each other as well as by the context [56]. In contrast, a child's conversation a VUI is devoid of mutual collaboration in the sense that the VUI's predefined roadmap restricts the possibilities for the conversation while also requiring children to unknowingly calibrate their responses to that roadmap [4, 13]. Because of this, it might be more reasonable to refer to children's interactions with VUIs as "pseudo-conversation", a conversation during which a coherent dialogue may be achieved if the child conforms their own verbal responses to the VUI designers' pre-determined dialogue flows. Children who unknowingly conform to the VUI's pre-determined dialogue flows may believe they are engaging in mutual, meaningful conversation with the VUI (i.e., an "illusion of a natural conversation" [60]), while children who intentionally calibrate their language to meet the VUI's apparent technological limitations would understand the interaction to be a one-sided verbal exercise. Nevertheless, if the VUI is design properly, the kind of pseudo-conversation enabled by the VUI would indeed provide children with valuable opportunities for oral language development. Thus, high-quality VUIs should incorporate complex and sophisticated dialogue flows that accommodate the various communication patterns exhibited by young children [10, 12, 36]. Incorporating the design strategies discussed above could help achieve this goal.

### 5.3 Limitations and Future Directions

Two limitations of our study suggest further research on using VUIs to support children's learning development. First, although we did not intend to restrict our analysis to certain cultural contexts, conducting our literature review and study in English and on English-based apps may have limited the generalizability of our findings across cultural contexts. Indeed, even among English speaking countries, there are variations in children's language development [11] and acceptance of smart technologies [51]. As such, developers wanting to target specific cultural groups may consider tailoring our general framework to the needs of those groups. For example, creating an agent persona that is consistent with the target users' cultural background (e.g., speaking the same dialect) may help build rapport with children and thus improve the quality of interaction [20]. Second, our evaluation was conducted by adult researchers using the apps and following a scripted protocol based on children's possible response patterns. Using this approach allows us to more effectively examine how the apps react to different types of responses. Future studies may want to compare our findings to children's actual use of the voice-based apps we studied. This would provide insight into how child-specific speech characteristics (e.g., imperfect grammar, mispronunciations) might influence the apps' performance.

## 6 CONCLUSION

While many children interact with voice-based apps, it is unclear whether such apps actually support children's language development by serving as their language partners. To evaluate this possibility, we analyzed currently available voice-based apps targeting young children to determine whether those apps incorporate

evidence-based dialogue strategies and how effectively the apps' dialogue structures support children's learning. We found that these apps fell short on engaging children in high-quality conversation as the children's "language guide." Specifically, the current apps predominately initiated dialogue in a restrictive way that may limit children's language production and flexibility. The apps' feedback design did not always come with extensions or encouragement nor did it adjust the dialogue based on children's responses, thereby failing to provide children with effective guidance during the interaction. Moreover, despite the variety of language activities covered by the apps we studied, the apps rarely afforded the opportunity for the type of back-and-forth dialogue that can encourage children's language output and increase their language exposure. Although such limitations may be partially due to the VUIs' reliance on overly simplistic, pre-defined dialogue flows, there are design strategies that developers should consider in order to produce high-quality VUIs that are capable of providing children with valuable language learning support.

## ACKNOWLEDGMENTS

The authors would like to thank the dedicated research assistants in the Digital Learning Lab at the University of California, Irvine for assisting with the search and coding of the apps. This research is in part supported by the National Science Foundation (Grant No.1906321).

## REFERENCES

- [1] Muneeb Imtiaz Ahmad, Omar Mubin, Suleman Shahid, and Joanne Orlando. 2019. Robot's adaptive emotional feedback sustains children's social engagement and promotes their vocabulary learning: a long-term child-robot interaction study. *Adaptive Behavior* 27, 4 (2019), 243–266.
- [2] Zahra Ashktorab, Mohit Jain, Q Vera Liao, and Justin D Weisz. 2019. Resilient chatbots: repair strategy preferences for conversational breakdowns. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.
- [3] Erin Beneteau, Ashley Boone, Yuxing Wu, Julie A Kientz, Jason Yip, and Alexis Hiniker. 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*. 1–13.
- [4] Erin Beneteau, Olivia K Richards, Mingrui Zhang, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication breakdowns between families and Alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [5] Frank Bentley, Chris Luvogt, Max Silverman, Rushani Wirasinghe, Brooke White, and Danielle Lottridge. 2018. Understanding the long-term use of smart speaker assistants. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3 (2018), 1–24.
- [6] Cezary Biele, Anna Jaskulska, Wieslaw Kopec, Jaroslaw Kowalski, Kinga Skorpupska, and Aldona Zdrodowska. 2019. How Might Voice Assistants Raise Our Children?. In *International Conference on Intelligent Human Systems Integration*. Springer, 162–167.
- [7] Pamela Blewitt, Keiran M Rump, Stephanie E Shealy, and Samantha A Cook. 2009. Shared book reading: When and how questions affect young children's word learning. *Journal of Educational Psychology* 101, 2 (2009), 294.
- [8] Stacy M Branham and Antony Rishin Mukkath Roy. 2019. Reading between the guidelines: How commercial voice assistant guidelines hinder accessibility for blind users. In *The 21st International ACM SIGACCESS Conference on Computers and Accessibility*. 446–458.
- [9] Chris Brebner, Jessie Jovanovic, Angela Lawless, and Jessica Young. 2016. Early childhood educators' understanding of early communication: Application to their work with young children. *Child Language Teaching and Therapy* 32, 3 (2016), 277–292.
- [10] Graham Button. 1990. Going up a blind alley: Conflating conversation analysis and computational modelling. In *Computers and conversation*. Elsevier, 67–90.
- [11] Lena G Caesar and Marie Kerins. 2020. Language and Literacy Predictors of Dialect Density Among School-Age African American Children From Two Geographic Regions. *Language, Speech, and Hearing Services in Schools* (2020), 1–14.

- [12] Heloisa Candello, Mauro Pichiliani, Claudio Pinhanez, Sara Vidon, and Mairieli Wessel. 2020. Co-designing a conversational interactive exhibit for children. In *Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts*. 326–331.
- [13] Yi Cheng, Kate Yen, Yeqi Chen, Sijin Chen, and Alexis Hiniker. 2018. Why doesn't it work? voice-driven interfaces and young children's communication repair strategies. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. 337–348.
- [14] Bruce H Choppin and Alan C Purves. 1969. A comparison of open-ended and multiple-choice items dealing with literary understanding. *Research in the Teaching of English* 3, 1 (1969), 15–24.
- [15] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, 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*. 1–12.
- [16] Vanessa Paz Dennen. 2004. Cognitive apprenticeship in educational practice: Research on scaffolding, modeling, mentoring, and coaching as instructional strategies. *Handbook of research on educational communications and technology 2*, 2004 (2004), 813–828.
- [17] Antonella Devescovi and Emma Baumgartner. 1993. Joint-reading a picture book: Verbal interaction and narrative skills. *Cognition and Instruction* 11, 3-4 (1993), 299–323.
- [18] Don C Dinkmeyer and Rudolf Dreikurs. 2000. *Encouraging children to learn*. Psychology Press.
- [19] Amaechi Uneke Enyi. 2015. Language and Interactional Discourse: Deconstructing the Talk-Generating Machinery in Natural Conversation. *Advances in Language and Literary Studies* 6, 4 (2015), 171–178.
- [20] Samantha Finkelstein, Evelyn Yarzebinski, Callie Vaughn, Amy Ogan, and Justine Cassell. 2013. The effects of culturally congruent educational technologies on student achievement. In *International Conference on Artificial Intelligence in Education*. Springer, 493–502.
- [21] Joel E Fischer, Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. Progressivity for voice interface design. In *Proceedings of the 1st International Conference on Conversational User Interfaces*. 1–8.
- [22] Robyn Fivush, Catherine A Haden, and Elaine Reese. 2006. Elaborating on elaborations: Role of maternal reminiscing style in cognitive and socioemotional development. *Child development* 77, 6 (2006), 1568–1588.
- [23] Radhika Garg and Subhasree Sengupta. 2020. Conversational Technologies for In-home Learning: Using Co-Design to Understand Children's and Parents' Perspectives. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [24] Rosemary L Getsie, Philip Langer, and Gene V Glass. 1985. Meta-analysis of the effects of type and combination of feedback on children's discrimination learning. *Review of Educational Research* 55, 1 (1985), 9–22.
- [25] Luigi Girolametto and Elaine Weitzman. 2002. Responsiveness of child care providers in interactions with toddlers and preschoolers. *Language, Speech, and Hearing Services in Schools* (2002).
- [26] Roberta Michnick Golinkoff, Erika Hoff, Meredith L Rowe, Catherine S Tamis-LeMonda, and Kathy Hirsh-Pasek. 2019. Language matters: Denying the existence of the 30-million-word gap has serious consequences. *Child development* 90, 3 (2019), 985–992.
- [27] Arthur C Graesser, Zhiqiang Cai, Brent Morgan, and Lijia Wang. 2017. Assessment with computer agents that engage in conversational dialogues and dialogues with learners. *Computers in Human Behavior* 76 (2017), 607–616.
- [28] Arthur C Graesser, Katja Wiemer-Hastings, Peter Wiemer-Hastings, Roger Kreuz, Tutoring Research Group, et al. 1999. AutoTutor: A simulation of a human tutor. *Cognitive Systems Research* 1, 1 (1999), 35–51.
- [29] Jennifer Hill, W Randolph Ford, and Ingrid F Ferreras. 2015. Real conversations with artificial intelligence: A comparison between human-human online conversations and human-chatbot conversations. *Computers in human behavior* 49 (2015), 245–250.
- [30] Rachel Jean-Baptiste, Harriet B Klein, Danielle Brates, and Nelson Moses. 2018. What's happening? And other questions obligating complete sentences as responses. *Child Language Teaching and Therapy* 34, 2 (2018), 191–202.
- [31] Aidan Jones and Ginevra Castellano. 2018. Adaptive robotic tutors that support self-regulated learning: A longer-term investigation with primary school children. *International Journal of Social Robotics* 10, 3 (2018), 357–370.
- [32] Darryn Lavery, Gilbert Cockton, and Malcolm P Atkinson. 1997. Comparison of evaluation methods using structured usability problem reports. *Behaviour & Information Technology* 16, 4-5 (1997), 246–266.
- [33] Henry Lee. 2018. *Voice User Interface Projects: Build voice-enabled applications using Dialogflow for Google Home and Alexa Skills Kit for Amazon Echo*. Packt Publishing Ltd.
- [34] Silvia B Lovato and Anne Marie Piper. 2019. Young Children and Voice Search: What We Know From Human-Computer Interaction Research. *Frontiers in psychology* 10 (2019), 8.
- [35] Silvia B Lovato, Anne Marie Piper, and Ellen A Wartella. 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*. 301–313.
- [36] Naja A Mack, Dekita G Moon Rembert, Robert Cummings, and Juan E Gilbert. 2019. Co-Designing an Intelligent Conversational History Tutor with Children. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. 482–487.
- [37] Michael A Madaio, Evelyn Yarzebinski, Vikram Kamath, Benjamin D Zinszer, Joelle Hannon-Cropp, Fabrice Tanoh, Yapo Hermann Akpe, Axel Blahoua Seri, Kaja K Jasińska, and Amy Ogan. 2020. Collective Support and Independent Learning with a Voice-Based Literacy Technology in Rural Communities. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [38] Samantha R Mattheiss, Edward J Alexander, and William W Graves. 2018. Elaborative feedback: Engaging reward and task-relevant brain regions promotes learning in pseudoword reading aloud. *Cognitive, Affective, & Behavioral Neuroscience* 18, 1 (2018), 68–87.
- [39] Stewart M McCauley and Morten H Christiansen. 2019. Language learning as language use: A cross-linguistic model of child language development. *Psychological review* 126, 1 (2019), 1.
- [40] Graeme McLean and Kofi Osei-Frimpong. 2019. Hey Alexa... examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior* 99 (2019), 28–37.
- [41] JOYCE H McNEILL and Susan A Fowler. 1999. Let's talk: Encouraging mother-child conversations during story reading. *Journal of Early Intervention* 22, 1 (1999), 51–69.
- [42] Zemira R Mevarech and Sigal Ben-Artzi. 1987. Effects of CAI with fixed and adaptive feedback on children's mathematics anxiety and achievement. *The Journal of Experimental Education* 56, 1 (1987), 42–46.
- [43] Ivonne Monarca, Franceli L Cibrian, Angel Mendoza, Gillian Hayes, and Monica Tentori. 2020. Why doesn't the conversational agent understand me? a language analysis of children speech. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers*. 90–93.
- [44] Christine Murad and Cosmin Munteanu. 2020. Alexa, How Do I Build a VUI Curriculum?. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. 1–3.
- [45] Philip Murphy. 2007. Reading comprehension exercises online: The effects of feedback, proficiency and interaction. *Language Learning & Technology* 11, 3 (2007), 107–129.
- [46] Cathy Pearl. 2016. *Designing voice user interfaces: principles of conversational experiences*. O'Reilly Media, Inc.
- [47] Lynn K Perry, Emily B Prince, Adriana M Valtierra, Camila Rivero-Fernandez, Mary Anne Ulery, Lynne F Katz, Brett Laursen, and Daniel S Messinger. 2018. A year in words: The dynamics and consequences of language experiences in an intervention classroom. *PLoS one* 13, 7 (2018), e0199893.
- [48] Carole Peterson, Craig Dowden, and Jennifer Tobin. 1999. Interviewing preschoolers: Comparisons of yes/no and wh-questions. *Law and Human Behavior* 23, 5 (1999), 539–555.
- [49] Martin Porcheron, Joel E Fischer, Stuart Reeves, and Sarah Sharples. 2018. Voice interfaces in everyday life. In *proceedings of the 2018 CHI conference on human factors in computing systems*. 1–12.
- [50] Sonya Rajan, Scotty D Craig, Barry Gholson, Natalie K Person, Arthur C Graesser, Tutoring Research Group, et al. 2001. AutoTutor: Incorporating back-channel feedback and other human-like conversational behaviors into an intelligent tutoring system. *International Journal of Speech Technology* 4, 2 (2001), 117–126.
- [51] V Rideout. 2020. *The Common Sense Census: Media Use by Kids Age Zero to Eight*. San Francisco, CA: Common Sense Media; 2017.
- [52] Michael Rohwer, Daniela Kloof, and Josef Perner. 2012. Escape from metaignorance: How children develop an understanding of their own lack of knowledge. *Child development* 83, 6 (2012), 1869–1883.
- [53] Hilary P Ruston and Paula J Schwanenflugel. 2010. Effects of a conversation intervention on the expressive vocabulary development of prekindergarten children. *Language, speech, and hearing services in schools* (2010).
- [54] Sashank Santhanam, Alireza Karduni, and Samira Shaikh. 2020. Studying the Effects of Cognitive Biases in Evaluation of Conversational Agents. (2020). arXiv:2002.07927
- [55] Subhasree Sengupta and Radhika Garg. 2019. Impact of Voice-based Interaction on Learning Practices and Behavior of Children.. In *IUI Workshops*.
- [56] Kevin Shockley, Marie-Vee Santana, and Carol A Fowler. 2003. Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance* 29, 2 (2003), 326.
- [57] Catherine Snow. 1983. Literacy and language: Relationships during the preschool years. *Harvard educational review* 53, 2 (1983), 165–189.
- [58] Mele Taumoepeau and Elaine Reese. 2013. Maternal reminiscing, elaborative talk, and children's theory of mind: An intervention study. *First Language* 33, 4 (2013), 388–410.

- [59] Georgene L Troseth, Gabrielle A Strouse, Israel Flores, Zachary D Stuckelman, and Colleen Russo Johnson. 2020. An enhanced eBook facilitates parent-child talk during shared reading by families of low socioeconomic status. *Early childhood research quarterly* 50 (2020), 45–58.
- [60] J Velkovska and M Zouinar. 2018. The illusion of natural conversation: interacting with smart assistants in home settings. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI 18)*.
- [61] Jinping Wang, Hyun Yang, Ruosi Shao, Saeed Abdullah, and S Shyam Sundar. 2020. Alexa as Coach: Leveraging Smart Speakers to Build Social Agents that Reduce Public Speaking Anxiety. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [62] Justin D Weisz, Mohit Jain, Narendra Nath Joshi, James Johnson, and Ingrid Lange. 2019. BigBlueBot: teaching strategies for successful human-agent interactions. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 448–459.
- [63] Elaine Weitzman, Luigi Girolametto, and Janice Greenberg. 2006. Adult responsiveness as a critical intervention mechanism for emergent literacy: Strategies for preschool educators. *Clinical approaches to emergent literacy intervention* (2006), 127–178.
- [64] Yunhan Wu, Daniel Rough, Anna Bleakley, Justin Edwards, Orla Cooney, Philip R Doyle, Leigh Clark, and Benjamin R Cowan. 2020. See what I'm saying? Comparing Intelligent Personal Assistant use for Native and Non-Native Language Speakers. (2020). arXiv:2006.06328
- [65] Ying Xu, Dakuo Wang, Penelope Collins, Hyelim Lee, and Mark Warschauer. [n. d.]. Same benefits, different communication patterns: Comparing Children's reading with a conversational agent vs. a human partner. *Computers & Education* 161 (n. d.), 104059.
- [66] Ying Xu and Mark Warschauer. 2019. Young children's reading and learning with conversational agents. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–8.
- [67] Ying Xu and Mark Warschauer. 2020. A content analysis of voice-based apps on the market for early literacy development. In *Proceedings of the Interaction Design and Children Conference*. 361–371.
- [68] Ying Xu and Mark Warschauer. 2020. Exploring young children's engagement in joint reading with a conversational agent. In *Proceedings of the Interaction Design and Children Conference*. 216–228.
- [69] Lauren Head Zauche, Taylor A Thul, Ashley E Darcy Mahoney, and Jennifer L Stapel-Wax. 2016. Influence of language nutrition on children's language and cognitive development: An integrated review. *Early Childhood Research Quarterly* 36 (2016), 318–333.
- [70] Andrea A Zevenbergen and Grover J Whitehurst. 2003. Dialogic reading: A shared picture book reading intervention for preschoolers. *On reading books to children: Parents and teachers* (2003), 177–200.
- [71] Ninger Zhou and Aman Yadav. 2017. Effects of multimedia story reading and questioning on preschoolers' vocabulary learning, story comprehension and reading engagement. *Educational Technology Research and Development* 65, 6 (2017), 1523–1545.
- [72] Frederick J Zimmerman, Jill Gilkerson, Jeffrey A Richards, Dimitri A Christakis, Dongxin Xu, Sharmistha Gray, and Umit Yapanel. 2009. Teaching by listening: The importance of adult-child conversations to language development. *Pediatrics* 124, 1 (2009), 342–349.
- [73] Tricia A Zucker, Laura M Justice, Shayne B Piasta, and Joan N Kaderavek. 2010. Preschool teachers' literal and inferential questions and children's responses during whole-class shared reading. *Early Childhood Research Quarterly* 25, 1 (2010), 65–83.