A Content Analysis of Voice-based Apps on the Market for Early Literacy Development

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

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

ABSTRACT

Voice-based applications powered by conversational agents can potentially support young children's literacy development in informal settings. Yet, to realize such potential, designers must consider young users' typical communication and learning patterns. In this paper, we developed a framework of seven design dimensions across three aspects (i.e., learning content and goals, interactions and gamifications, and conversation design) that could influence the educational benefits young children receive from voice-based apps. We then used this framework to conduct a content analysis of 535 literacy-focused voice apps on the market to examine whether the prevalent design features of these apps meet the unique needs of young learners. Lastly, we discussed the implications of current design tendencies with the aim of encouraging future voice-based app designers to bridge the gap between research and practice.

Author Keywords

Conversational agents; voice interfaces; apps; young children; content analysis

CCS Concepts

•**Human-centered computing** → **Natural language interfaces; Empirical studies in HCI;** •**Social and professional topics** → **Children;**

INTRODUCTION

In recent years, a broad range of what are often marketed as "conversational" or "voice" technologies—including smart speakers such as Amazon Echo or Google Home—has gained popularity in many homes in the U.S. and around the world. A report shows that the number of voice-based apps on the Google and Amazon platforms has more than doubled in one year [62]. Many of these apps are education oriented and target early literacy development for preschoolers. These apps typically involve preset dialogical flows that provide children with interactive content. Children may practice spelling, play word games, engage in choose-your-own-adventure activities, or ask

the app to tell a story. These apps have the potential to become valuable sources of language exposure at home, providing opportunities for children to practice their communicative and literacy skills through fun learning experiences [21, 31, 56, 59]. These educational resources may be especially valuable for preschool-aged children, with a typical age range from 3 to 6 years. Young children in this age group are not yet able to read and write fluently, thus amplifying the value of voice-based interactions. Children in this age group also typically have sufficient oral language skills, thus benefiting from their productive communication with intelligent systems. Yet to actualize the promise of voice interfaces for young children, app designers must consider young users' typical communication and learning patterns [47].

Past research from the fields of child development, education, and human-computer interaction has identified general strategies that can facilitate young children's engagement and learning in face-to-face settings and when using interactive digital technology such as tablets and computers. For a review, see [3, 8, 30, 37, 46]. Another line of research focusing on children's interaction patterns with voice-based technologies also provides useful prior knowledge for voice-based app designs [9, 33, 43]. These separate but overlapping strands of literature should be synthesized to outline the facilitative elements that voice-based app designers should consider incorporating in their products. However, no research has yet synthesized these disparate elements into one comprehensive framework, nor has any research used such evidence-based framework to evaluate the apps on the market. To fill this research gap, we explore the following question in this paper: *What are the common and missing educational design features of literacy-focused voice-based apps on the market targeting young children aged 3 to 6 years?*

To answer this question, we developed a framework of 7 design dimensions in three focal aspects (i.e., learning content and goals, interactions and gamifications, conversation design) that could influence the educational benefits that young children receive from voice-based apps. Based on the framework, we conducted a content analysis of the existing literacy-focused voice apps on the market to examine whether the prevalent design features of these apps meet the unique communicative abilities and learning needs of young learners. For this task, we systematically analyzed 535 available literacy apps on the Google Assistant and Amazon Alexa platforms. We then discussed the implications of current design tendencies with

the aim of encouraging future voice-based app designers to bridge the gap between research and practice.

This paper makes at least two contributions to the field. First, the framework we develop may be of use to parents, educators, and developers for their app selection or development. Second, the app analysis may provide a holistic picture of whether current apps are designed to support preschool-aged communication and learning. Although we focus on literacy-focused apps in this paper, we believe some of our findings could also be generalized to other voice-based interactive designs for young children's informal learning.

RELATED WORK

Voice Interfaces in Early Literacy Development

Children's language is developed not only by exposure to words, but also through dialogic interactions with a more experienced conversation partner [20,40]. Studies have consistently found that parents vary in their conversation styles at home, thus resulting in differential language and literacy skills among young children [23, 24, 53]. Some researchers proposed that conversational technologies could, to some extent, simulate a human language partner in a child's learning processes [59], and many voice apps targeting literacy skills have thus been developed for children. These newly available apps can be deployed on smart phones or smart speakers that users already own, thus making these apps easily accessible to a wide range of users.

A small body of prior research has documented how children utilize voice-based interfaces on smart speakers for literacy-related activities. For example, Sciuto and colleagues conducted parent interviews and found that children had a variety of conversations with Alexa, including asking the smart speaker questions about homework and requesting Alexa to tell a story [43]. Similar findings were reported in Lovato and colleagues' study: Children engaged in conversation related to language learning with the smart speakers, including vocabulary, spelling, and translations [33]. In another study conducted by Lovato and colleagues, the authors analyzed YouTube videos on children's use of smart speakers and suggested that voice interfaces support information seeking and knowledge base development [32]. Children's active use of voice-based apps documented in these studies points to the feasibility of refining conversational technologies to better approximate a language learning partner. However, these studies did not detail the features of the voice interfaces children interacted with. This ambiguity makes it difficult to connect the app designs to their affordances.

Another related line of research also attempted to leverage voice interfaces to mimic children's conversational partners through embodied screen-based agents or robots [11,17,44,54]. This line of research usually reports design features of the systems and also assesses children's learning outcomes, demonstrating how systems based on learning principles support early literacy development. For example, Kory and Breazeal developed a robotic learning companion for preschool children's oral language development [29]. The robot was designed to tell children stories at different difficulty levels while introducing new vocabulary words and modelling good story narration styles (e.g., including a beginning, middle, and end of a story). Children then told their own story, and the robot used a set of scripted prompts to lead children through their storytelling. The study found that children learned new words from the robot. In another study, Freed developed a robot that played a food-sharing game for English-speaking preschoolers to learn French vocabulary [16]. This robot was programmed to introduce itself and name foods in French. While these two studies point to well-designed voice-interfaces' capacity to promote literacy development, many of the learning affordances of such systems may result from the robot's embodiment and its capacity to communicate through multiple modalities. Thus, the insights from the studies above do not necessarily tell us about the optimal design traits of a speech-only conversational interface. Moreover, these robots target a very narrow consumer base and may not be accessible to most children.

Evaluation of Literacy Apps Targeting Young Children

As noted in the previous section, many voice-based apps targeting early literacy skills have already been developed, and children frequently utilize them. Young children differ from older individuals in that the former are still developing their cognitive skills [22] and their mental schema for interacting with intelligent artifacts [2, 9]. As such, it is important to develop apps that meet the communication and learning needs of young children. However, to our knowledge, there exists no evaluation of or rubric for evaluating voice-based educational apps tailored to preschoolers. This lack of scholarly focus may stem from the fact that voice-based products have only emerged in recent years.

Over the past decade, however, there have been numerous frameworks created to evaluate the appropriateness of computer- or tablet-based apps targeting young children [3, 8, 30, 37, 46]. Although these frameworks are not directly capable of addressing the specific functionalities afforded by voice interfaces, they can be used to establish broad guidelines for evaluating voice-based apps. For example, Soni and colleagues' TIDRC (Touchscreen Interaction Design Recommendations for Children) framework suggests that educational apps should meet children's cognitive, physical, and social-emotional needs [46]. The cognitive category within this framework focuses on an app's multimedia and interactive features. The physical category focuses on haptic and gestural features associated with operating the app. The social-emotional category focuses on contextual features that incorporate social interactions and gamification to promote the enjoyment of app use. Papadakis and colleagues included these same categories in their REVEAC (Rubric for the EValuation of Educational Apps for preschool Children) framework, with a slightly different label for each category [37].

Researchers have used these rubrics to evaluate apps on the market and identified striking research-practice gaps [8, 26, 38, 49]. For example, Vaala and colleagues pointed out most of the apps did not provide high-quality feedback; these apps either did not include any responsive feedback or included feedback that lacked exploratory information to guide

young learners to understand the educational content being taught [49]. In another preschool app evaluation, Callaghan and Reich evaluated the scaffolding structure of apps and suggested that many preschool apps did not include leveling that strategically scaffolds young children's learning to gradually increase in difficulty as preschoolers progressively increase understanding of the material [8]. These two studies came to a similar conclusion: The design of most preschool tablet-based apps is not optimal in light of what research has found helps preschool-aged children learn and develop. It would not be surprising if similar limitations are found in voice-based apps.

Current Guidelines for Developing Voice Interfaces

While few research studies have focused on frameworks that guide the development of voice-based apps, some companies that market popular voice interfaces, including Google, Amazon, Apple, and Microsoft have published such guidelines. Branham and Roy reviewed these commercial guidelines and pointed out their common assumption that voice interfaces should be developed to carry out efficient and concise conversation from the user's perspective [5]. More specifically, these guidelines aim to provide users with the information they require without demanding unnecessary time and cognitive effort. These guidelines, in general, support productive language exchange between task-oriented adult users and the voice interface. However, these commercial guidelines are not specifically tailored for use by young children. Commercial voice-based apps are primarily designed to accomplish transactional tasks (e.g., placing an order on Amazon or checking the weather), and these tasks benefit from conversational efficiency. However, when developing apps specifically for young children's learning, a variety of dimensions other than efficiency need to be taken into consideration. Because young children are still developing their language proficiency, they would benefit from additional support and scaffolding provided by the voice interfaces they interact with. Thus, a clearer understanding of the dimensions which impact children's engagement with and learning from voice interfaces is needed.

FRAMEWORK DEVELOPMENT

Establishing Dimensions

To develop our framework, we first synthesized several lines of literature including child development, psychology, education, and informatics to identify key evaluative dimensions that may influence children's engagement with and learning from voice-based apps. The section below details how each of the dimensions we focus on in the current study is informed by the scholarly literature. We grouped these dimensions into three aspects: i) learning content and goals, ii) interactions and gamifications, and iii) conversation design.

Learning Content and Goals

Learning activity. Young children develop their literacy skills in everyday informal, play-based literacy activities [58, 61]. These activities, such as having conversations with family members, listening to and telling stories, and playing games, create opportunities for situated learning by providing immersive and motivating contexts [36]. Researchers also argued that learning activities that only quiz children on pre-existing knowledge through drill-and-practice tasks or heavily rely on explicit instructions may not be as suitable for young learners [51].

Goal clarity. As preschool-aged learners are just beginning to build their cognitive processing and literacy skills [18], apps with clear and simple goals may be more effective in supporting young children's literacy development. Learning that is guided by a specific goal reduces children's cognitive load by focusing children's attention on the current task [57]. Research has also suggested that apps with clear learning goals may decrease learner frustrations during use [1]. Moreover, voice-based apps do not support multi-media presentations of the learning content that could facilitate learners to understand the learning goal. As such, using language to present the goal clearly may be especially important for these apps.

Interactions and Gamifications

Interactivity. Interactivity, in general, may promote active participation in the learning process [13]. This finding should hold true when applying to the case of voice-based interfaces; if opportunities for children to interact with the content are present, children may be more attentive and more "minds-on," thus leading to better learning outcomes. This is evidenced in a study with a robot tutor, suggesting that children better acquired novel vocabularies if they had the opportunities to retell the story rather than solely listening to the story [29].

Gamification. Gamification is the use of game design elements and game mechanics in learning contexts, which have been viewed as an effective means to motivate children to learn and develop intrinsic interests for learning in the long run [35]. Researchers suggested that gamification increases the playfulness of learning [12], or makes learning a "fun and pleasurable experience" [19]. The gamification elements commonly utilized in studies include points, rewards, leveling, competition, and customization [15, 28]. While there has been little research evidence, it is fair to conjecture that gamification elements may promote children's engagement and learning in voice-based apps in a similar way.

Conversation Design

Conversational prompts. The design of conversational prompts is crucial in voice-based apps where language exchanges constitute the majority of the learning experiences. A good conversational prompt should invite thoughtful responses from the child and move conversation forward [45]. In general, a trade-off exists among different prompting strategies. On the one hand, posing open-ended questions (e.g., What is snow? Why does it snow?) encourages children to put in cognitive effort and engage in deep-level processing [25]. Yet this may increase the programming and interaction issues in the context of automated conversations due to the unbounded user responses. On the other hand, posing closed questions (e.g., Is this snow?) reduces the challenges in dialogue design and programming, yet may be less beneficial for learning [55].

Feedback. In this paper, feedback was broadly defined and refers to how the app followed up after children's responses, but such feedback was not necessarily used to assess right or wrong responses. Responsive feedback that temporally and

semantically continues the conversation is fundamental for promoting children's learning [50]. Furthermore, a wealth of literature indicates that elaborative feedback that provides more instruction is more beneficial for children. Elaborative feedback may contain multiple components, including praise, extension of children's responses, explanation of the question, or a further question that provokes deeper thinking [42]. This feedback strategy has consistently been proven to engage children in more cognitively active roles in face-to-face dialogic interactions [27].

Scaffolding. Scaffolding refers to the steps taken to reduce the degrees of freedom in carrying out some task, so that the child can concentrate on the difficult skill he is in the process of acquiring. Adults may pose follow-up questions that are adaptive based on children's response to a previous conversational prompt [6]. Many studies have suggested scaffolding is especially beneficial for children who are just beginning to learn a certain topic [4]. Scaffolding also plays an important role in supporting children's successful communications with voice-based apps. Children may encounter particular problems probably due to CAs' imperfect ability to understand children's speech and children's less sophisticated schema regarding interaction with CAs [9]. As such, scaffolding features from the CAs themselves should be provided to help children avoid communication breakdowns resulting from CAs' misinterpretation of their speech.

Developing Codes
After establishing the seven dimensions, we then further developed the codes for each dimension using a combination of deductive and inductive approaches, which will be detailed below.

We first developed basic codes using a deductive approach. Codes were drawn from design heuristics that are implied from prior literature on children's interaction and conceptualization of voice-based apps [14, 32, 48]. To evaluate educational affordances, we included codes adapted from two coding schemes developed for content analyses of educational preschool tablet-based apps [7, 49]. We then used an inductive approach to modify the codes based on our preliminary play of voice-based apps. Once a basic coding framework was created, two instructional technology professionals (not authors) were asked to evaluate 20 different apps using the basic framework. Their feedback was used to improve the clarity and comprehensiveness of the framework elements.

Modifications were then made to the framework based upon the instructional technology professionals' suggestions (i.e., change of wording, modification of codes). The research team with two researchers tested the modified framework by coding the same 20 apps, and any discrepancies in coding were negotiated until mutual agreement was achieved. By doing so, we further collaboratively refined codes to better capture the various design features used within the apps.

CONTENT ANALYSIS
App Selection
A list of voice-based apps was collected for both the Amazon Alexa and Google Assistant platforms available in January 2019. Because there is not a standardized indicator of which apps were developed with a focus on early literacy, we first conducted a key term search to identify literacy-related apps and then sought to identify those presumably designed for young children. For Amazon, our key term search included apps (termed as Alexa Skills) that contain "story," "literacy," "language," "book," "read," or "vocabulary" in the title or description and are in the "Kids" category. For Google, we used the same search terms as we did on the Amazon platform and included apps (termed as Google Actions) under the "Kids and Families" category. We soon noticed there were relevant apps that might be suitable for young children but that were not included in the "Kids" or "Kids and Families" categories on either platforms. We therefore expanded our search to Google's "Education and References" and "Games and Fun" categories as well as Amazon's "Education and References" and "Games" categories. This preliminary search yielded a total of 1613 literacy-related apps from the two platforms.

Since our evaluation focuses specifically on voice apps for preschool-aged children, we looked through the description of each app and excluded those with an explicit target age range over six years. For those apps that did not provide a target age range, we read the description, tried the app, and then determined whether or not the apps were age-appropriate for young children. We only excluded those apps that were obviously unintended for use by young children. Nine hundred and seventy-five apps fell within this category, owing to their requirement of advanced language ($n = 320$), their difficult or inappropriate content ($n = 452$), or their assumed prior knowledge or skills that young children typically do not possess ($n = 203$). One hundred and three apps were also excluded because of technical issues (e.g., could not be activated, terminated in the middle of play). The total number of eligible apps for evaluation was 535, with 323 Alexa skills and 212 Google actions. A complete list of apps is displayed as supplementary materials .

Apps developed by third party developers must be reviewed and certified before they are made public to users. The review process involves a functional test, which ensures the app's proper functionality; a policy check, which ensures that the content does not violate legal and ethical standards; and a security check, which protects user privacy.

The Coding Procedure
Two coders coded the 535 eligible apps. The coders played through each app by providing three types of responses: 1) correct responses, 2) incorrect responses, and 3) random (i.e., semantically irrelevant responses) in order to understand how the app handles edge cases. The random responses that coders provided were answers that were clearly out of the context of the app content. In addition, coders sat idle during each task for at least one minute to see if the app provided additional directions when users stopped interacting with it. Depending on the apps' complexity, the coders spent 2 to 15 minutes to play through each app.

Establishing reliability between the two coders was done over the course of five weeks. Pilot coding of 10 apps was first done together, discussing where certain game features would fall

within the coding framework. Then, additional pilot coding of another 10 apps was done separately, eventually comparing each other's codes and discussing discrepancies. The two coders' inter-rater reliability (Cohen's Kappa) reached 0.80 to 0.85 across categories before they eventually started the formal coding. Every two weeks, coders calibrated their coding by separately coding a new app, comparing their codes, and calculating the Cohen's Kappa coefficient. The Cohen's Kappa remained above 0.80 for all categories throughout the coding process.

RESULTS

Learning Content and Goals

Learning activity
As shown in Table 1, audio stories and trivia games were the two most common activity types, in total constituting almost three quarters of the apps we reviewed. Specifically, audio stories, almost half of all apps (48.2%, $n = 258$), narrate stories from a wide range of genres, including fantasy, informational, and religious. The second most common type of activity (25.2%, $n = 135$) was trivia games that quiz children's existing knowledge, where children were prompted to respond to a series of questions with a wrong or right answer. For example, an app solicited answers about baby animal names (e.g., "*What is the baby animal name for Pig?*") and another app asked children to name animals that rhyme with specific words (e.g., "*What animal rhymes with box?*"). Lessons were in 89 apps (16.6%). These apps contained explicit instructions, showing children new vocabulary words or teaching the definition of a word. Some of these lessons also included brief quizzes on the material taught.

Story-generator and questioning-and-answering apps were less common. Story-generator apps (3.9%, $n = 21$) engaged children in story creation. These apps typically prompted children to provide story elements such as the name, gender, and characteristics of the main character, or events in the story. The app then generated a complete story (or sentence) with children's input incorporated. Questioning-and-answering apps (6.0%, $n = 32$) allowed children to ask questions in a specific domain such as opposite words and spelling.

Code value	Value definition	n	%
Story generator	Users provide story elements for the app to automatically assemble the elements and create a story	21	3.9%
Questioning and answering	Users ask questions and receive answers in a specific domain	32	6.0%
Lesson	Users listen to a lesson on a specific topic	89	16.6%
Trivia game	Users answer a set of questions on a specific topic (e.g., animal facts)	135	25.2%
Audio story	Users engage in storybook listening	258	48.2%

Table 1. Learning activity ($n = 535$)

Goal clarity
Less than half of the apps (41.5%, $n = 222$) presented clear goals to users. These apps usually articulated the goal at the beginning of the play (e.g., "*Let's learn some opposite words today!*", "*In this game, I will ask you some questions about letters.*").

Interactions and Gamifications

Interactivity
We categorized the apps into three levels: non-interactive, lowly interactive, and interactive (see Table 2). This three-level categorization is in line with other works that review tablet-based apps [8], and our own play of the apps revealed that they could be intuitively categorized into one of the three interactivity levels. About a quarter of the apps (24.8%, $n = 133$) were categorized as non-interactive. Some of these apps merely played a randomly selected story or taught a word-of-the-day after children used a speech command to start the app. Another quarter of the apps (24.5%, $n = 131$) were categorized as lowly interactive. These apps provided a list of content children could choose from after the app was activated, but children had no control over the content once a selection was made. The remaining half of the apps (50.7%, $n = 271$) were categorized as interactive since children could interact with the agent by responding to conversational prompts throughout their experience with the app.

The interactivity levels differed among apps involving different learning activities. Apps designed to engage children with story generating, questioning and answering, and trivia games were almost always categorized as interactive. Given the nature of these activities, it is not surprising that these apps would require a higher level of participation. There was a wider variation in interactivity among the apps containing lessons or audio stories. A large majority of the 89 lesson apps were non-interactive (n = 65, 73%); children simply listened to some information on a particular topic and did not have any opportunities to interact. The remaining 24 (27.0%) were interactive since children were given a short quiz after the lesson. Among the 258 audio story apps, 91 (35%) of the apps were categorized as non-interactive since children could not navigate the content once the app was activated, and 83 were categorized as lowly interactive since children only chose the story they wanted to listen to. The remaining 84 audio story apps were categorized as interactive. These were often "choose-your-own-adventure" stories, where children's choices determined how the stories progressed. Others either asked children questions throughout the story or required children to say specific words (e.g., in response to a "*please repeat after me*" command) for the story to continue.

Code value	Value definition	n	%
Non-interactive	Users have no control over the content after app activation	133	24.8%
Lowly interactive	Users may choose from content list but no control over the content once selection is made	131	24.5%
Interactive	Users may interact with content throughout the app	271	50.7%

Table 2. Interactivity levels ($n = 535$)

Gamification
None of the non-interactive and lowly interactive apps contained gamification elements. As such, we excluded these apps from the coding results and only examined the 271 interactive apps.

Among the 271 apps that provided a high level of interactivity, almost half employed one or multiple gamification techniques

(see Table 3). Customization was the most commonly seen technique; over a quarter of the apps allowed children to decide how they would like the stories to progress (26.6%, *n* = 72), and ten apps allowed users to customize the character names and use the names throughout the app (3.7%, *n* = 10). About 20% of the apps (*n* = 58) utilized a reward system where children earned points or were praised for answering questions correctly. Ten percent of apps contained leveling, in which children may unlock a more difficult challenge by accumulating a certain amount of points (*n* = 28). Multiple-player games that allowed children to engage simultaneously and compete with each other were less common; only 6% of the apps utilized such strategy (*n* = 17). Interestingly, in cases when users indicated that they would play the game alone, these apps usually generated a virtual opponent.

Code value	Value definition	n	%
Not present		129	47.6%
Customization	Users may alter learning progress (e.g., choose-your-own adventure) or configuration (e.g., character name)	82	30.3%
Rewarding	Users may get points for answering questions correctly	58	21.4%
Leveling	Users may progress to more challenging levels	28	10.3%
Multi-players	Users may play the apps with nearby users or with virtual players	17	6.2%

Table 3. Gamifications (*n* = 271, non- and lowly-interactive apps excluded). Multiple gamification elements may be used in one single app.

Conversation Design

As defined above, apps with low or no interactivity merely allow users to use voice command to activate or select content from the app; they do not provide opportunities for users to "converse" with the agent about the content. Therefore, conversation design was coded for apps that were categorized as interactive (271 apps).

Conversational prompts

For the 271 apps that provided high interactivity, we coded the nature of the conversational prompts (see Table 4). Multiple-choice questions that listed two to four options were the most common type of prompts, which were used in 169 apps (62.3%). These prompts were either for children to navigate a choose-your-own-adventure story (e.g., "*Where would you like to explore first? Inside the house or outside the house?*"), to answer a quiz (e.g., "*What color is a polar bear? Is it white or is it red?*"), or to answer yes-or-no questions about whether a statement related to a story is correct. The second most prevalent type of prompts asked children to provide a single-word response in a specific field (e.g., an animal, color, number, letter, or a sight word they have just learned), which was incorporated in 67 apps (24.7%). These prompts were mostly used in word guessing games or vocabulary learning apps. Another 19 apps (7.0%) prompted users to freely express their thoughts, usually used for the story creation apps. For example, the agent prompted children by saying "*Imagine this is a sunny day. What would happen?*" Fourteen of the high-interactivity apps (5.2%) incorporated repetition prompts that ask children to repeat exactly what the agent has said earlier. Open-ended questions (e.g., wh- questions) that encouraged

multiple-word responses in a specific context were the least common; only 12 apps (4.4%) utilized such questions.

There was an interplay between the apps' learning activities and the types of conversational prompts included in the apps. The most common type of prompts among questioning and answering apps were free responses, while the most common prompt type for trivia games and story generating apps was one-word responses. Audio stories and lessons most frequently utilized multiple-choice prompts. Repetition and open-ended prompts were only rarely included in apps for any of the learning activities.

Code value	Value definition	n	%
Multiple-choice	Users select options from a list	169	62.3%
One-word response	Users provide a single word in a certain field (e.g., color, food)	67	24.7%
Free response	Users can respond anything	19	7.0%
Repetition	Users repeat what the app says	14	5.2%
Open-ended prompt	Users answer a "what," "when," "where," "how," "why" question	12	4.4%

Table 4. Conversational prompts (*n* = 271, non- and lowly-interactive apps excluded). Multiple conversation prompting strategies may be used in one single app.

Feedback

We first coded feedback for valid responses (see Table 5). Among the 271 interactive apps we coded, 188 apps asked children fact-based questions that had correct or incorrect answers, and the other 83 apps did not. Among the 188 apps asking questions that had correct or incorrect answers, 20 apps (7.4%) did not provide responsive feedback; these 20 apps responded the same way regardless of users' voice input. For example, an app asked children to recall a story fact based on a story they had just heard, but did not acknowledge whether the child's recall was correct. Another app prompted children to provide a noun in a specific group (e.g., "*tell me an animal name*"), but in the cases when users offered a word not from the required group (e.g., children responded a color name), the agent did not point this out. In addition to the 20 apps, the rest of 168 apps with correct or incorrect answers provided one form or a combination of different forms of responsive feedback. Praise was the most common type of feedback (*n* = 116, 42.8%), sound effects was the second most common type (*n* = 90, 33.2%), and corrective was the third most common type (*n* = 64, 23.6%). Less apps provided elaborative feedback that contained explanations or further instructions (*n* = 44, 16.2%). There were also 12 apps (4.4%) that repeated the child's answer before assessing it. Among the apps that were not intended to assess correct or incorrect responses (*n* = 83), all responded to children responsively. Sixty-five apps (23.9%) continued the story or conversation based on children's choices if child utterances were recognizable. Ten apps (3.7%) responded to child's questions, and 8 apps (3.0%) repeated exactly what the child said.

Apps also responded to unrecognizable input differently (see Table 6). When a user gave an edge response, some apps apologized (17.7%, *n* = 48) or/and acknowledged their inability to understand the user's input or inability to perform the task a user commands (14.4%, *n* = 39), with most using language and tones that did not seem friendly to a child (e.g., "*Sorry. I*

Code value	Value definition	n (crt)	n (non-crt)	%
Praise	e.g., "Great job!"	116	0	42.8%
Sound effects	e.g., rising sound or claps for correct answers	90	0	33.2%
Continuation of narration	Narration proceeds according to user responses	0	65	24.0%
Corrective	Evaluating whether a response is correct or not. (e.g., "That's right!")	64	0	23.6%
Explanation	Explains why user's response is correct or incorrect	44	0	16.2%
Repetition of user input	Repeating what a user says	12	8	7.4%
Answers to user questions	Answering the questions asked by users	0	10	3.7%
Non-responsive	Apps do not respond based on user input	20	0	7.4%

Table 5. Feedback to valid responses (*n* = 271, non- and lowly-interactive apps excluded).crt = apps that ask questions with correct or incorrect answers; non-crt = apps that ask questions without correct or incorrect answers.
.

don't understand." "I didn't catch that."). About half of the apps repeated the same question (48.3%, *n* = 131) or preceded the questions by encouraging the child to attempt a second time (e.g., "*Let's try again!*"; 45.4%, *n* = 123). There was a considerable proportion of apps (22.9%, *n* = 62) that treated edge responses as incorrect answers.

Code value	Value definition	n	%
Apology	e.g., "Sorry"	48	17.7%
Acknowledgement	e.g., "I don't understand." "I can't help with that."	39	14.4%
Repetition of prompts	Repeating the same question with same or slightly different wording	131	48.3%
Encouragement	Encouraging a user to try a second time (e.g., "try again!")	123	45.4%
Incorrect	Treating edge responses as incorrect response	62	22.9%

Table 6. Feedback to unrecognizable user input (*n* = 271, non- and lowly-interactive apps excluded). Multiple feedback strategies may be used in one single app.

For the cases when children did not respond when a prompt was asked (see Table 7), almost all of the apps (*n* = 223, 82.3 %) repeated the exact same prompt, and 56 apps (20.7%) encouraged children to respond by saying, for example, "*I really want to know what you think!*" A considerable portion of apps treated no voice input as if unrecognizable input; the apps apologized for (*n* = 10, 3.7%) and acknowledged (*n* = 12, 4.4%) their failure to interpret the user response. A few apps discontinued without a voice input after users' nonresponse.

Scaffolding
Table 9). Indeed, these two forms of scaffolding were rarely incorporated in the apps we analyzed. Most of the apps (*n* = 233, 86.0%) did not provide learning scaffolding where children were allowed to ask for hints, contextual information, or definitions of words. Conversation scaffolding was also uncommon, 226 apps did not include such scaffolding (16.6%). When children produced unrecognizable answers, 37 apps (13.7%) rephrased the wording of questions to make them

Code value	Value definition	n	%
Apology	e.g., "Sorry"	10	3.7%
Acknowledgement	e.g., "I don't understand." "I can't help with that."	12	4.4%
Repetition of prompts	Repeating the same question with same or slightly different wording	223	82.3%
Encouragement	Encouraging a user to try a second time (e.g., "try again!")	56	20.7%
App discontinued	The app does not follow-up or re-prompt and crashes silently after a period of time	9	3.3%
Incorrect	Treating edge responses as incorrect response	62	22.9%

Table 7. Feedback to no user input (*n* = 271, non- and lowly-interactive apps excluded). Multiple feedback strategies may be used in one single app.

more accessible for children to answer. For example, an app first asked, "*Would you like a muffin or banana cake?*" If the child produced an unrecognizable answer, the app rephrased the question as "*What would you like? Please say muffin or banana cake.*" Another app used a similar strategy but rephrased a multiple-choice question into a yes-or-no question. In addition, 10 apps (3.7%) provided options that children could choose from to help them respond to the prompt. Another question-and-answer app provided an example question for children if they did not ask a question the app could recognize.

Code value	Value definition	n	%
Not present		233	86.0%
Word definition	Users may request definition of a word	22	8.1%
Cues	Apps provide hints when a user fails to respond appropriately	10	3.7%
Contextual information	Apps provide additional background information	6	2.2%

Table 8. Learning scaffolding (*n* = 271, non- and lowly-interactive apps excluded). Multiple scaffolding strategies may be used in one single app.

Code value	Value definition	n	%
Not present		226	83.4%
Rephrase	Apps rephrase a question in simpler language	37	13.7%
Options	Apps provide multiple options for a user to choose from	10	3.7%
Examples	Apps provide example input phrases a user may model after	1	0.0%

Table 9. Conversation scaffolding (*n* = 271, non- and lowly-interactive apps excluded). Multiple scaffolding strategies may be used in one single app.)

DISCUSSION
Given that voice-based apps are relatively new, guidelines for designing educational voice-based apps for young children have not yet been established. It is thus necessary to create such guidelines by leveraging prior knowledge in the fields of child development, education, and human-computer interaction. In this paper, we first developed an evaluative framework along seven dimensions in three focal aspects (i.e., learning content and goals, interactions and gamifications, and conversation design). We then used this framework to evaluate how well currently available voice-based apps are designed to meet young children's learning needs.

The first focal aspect we focused on was the apps' learning content and goals. Regarding learning activities, the current apps covered a wide range of learning activities. Given that story reading is a vital source of literacy exposure in early childhood, it is encouraging that the majority of apps emphasized this activity [10]. In particular, a considerable number of apps allowed children to engage in "choose-your-own-adventure" stories or create their own stories by choosing story elements. These activities encouraged children's active participation and creativity [41]. However, we also found that many apps merely quizzed children on their prior knowledge and did not extend children's knowledge base. Also, we found that explicit instruction was heavily used in the apps that involved vocabulary and foreign language lessons. The apps typically taught children a new word and provided its definition without situating it in any context. While such explicit instructions may be effective in teaching children some skills in the short term, this kind of app may not be attractive for young children and may be less effective in promoting children's long-term interest in literacy [10]. Regarding goal clarity, we found that almost half of the apps explicitly presented a clear objective to the users, making it clear what to expect in the play. Nevertheless, many apps simply involve playing an audio story that does not solicit user input, for which it may not be necessary to present users with the over-arching goals.

The second focal aspect we focused on was interactions and gamifications. We noted that almost half of the voice apps we studied lacked on-going interactivity. This is disappointing because young children can more readily engage with spoken prompts than with haptic interactions or text-based prompts. Spoken prompts may require fewer cognitive resources from young children as compared to the latter types which may require sophisticated fine motor or literacy skills [60]. Among the voice-based apps that did build in interactivity, many of them also incorporated gamification to enhance the playfulness of the apps. These gamification techniques were similar to those typically used in tablet-based or computer-based apps [8, 38].

The third aspect we focused on was conversational design, a feature essentially unique to voice-based apps. Regarding conversational prompts, most currently available voice-based apps utilized multiple-choice prompts rather than open-ended prompts. Multiple-choice prompts constrained children's responses to a small range of pre-selected options, thus minimizing the programming and interaction issues accompanying unbounded user responses. However, open-ended prompts are more ideal for learning and interaction [52]. As such, despite the potential programming and design challenges, we believe allowing open-ended responses is vital for maximizing the educational benefits of an app. In terms of the feedback features, many apps were designed to follow up according to children's responses, but these apps varied in the quality of feedback they provided. Very few apps offered the type of elaborative explanation that has been proven to help children clarify their confusion or solidify their understanding [34, 39]. In addition, very few apps provided scaffolding to ease children's communication and learning, and an equally small number adopted strategies to prevent conversation breakdown in cases where apps consistently fail to understand children's speech. This kind of scaffolding is especially important for facilitating children's interactions with apps that only entail a voice interface [9].

In summary, through these analyses, we identified the common and missing educational design features of voice-based apps available on the market. Apps were designed to carry out a wide variety of literacy activities that mirror the ones children usually engage in at home or at school, yet many of the apps lack interactivity that encourages children's verbal engagement. Moreover, it appeared that the conversation needs to be developed through an educational lens, with questions that better stimulate thinking and scaffolding that keeps the conversation within children's developmental levels.

Based on the results discussed above, we propose several recommendations for future app development:

- **Clarify the target age range of potential users**. We noticed that most of the apps did not specify the intended age range of users. This may indicate that many developers do not initially determine a target age range, which makes it difficult to develop apps that meet users' specific needs.

- **Leverage the conversation technologies to promote interactive learning**. Many apps we reviewed were not categorized as interactive. We recommend more apps incorporate interactivity to provide children with more opportunities to engage in conversation. Such interactions should aim to provide children with more enjoyable and playful user experiences.

- **Encourage less-restricted verbal expressions**. We found that prompts that restrict children's voice responses into single words were heavily used among the apps we reviewed. We recommend that more apps seek to promote a higher level of verbal engagement by incorporating open-ended prompts. While this may introduce more challenges on the programming side due to the decreased predictability of children's responses, the learning advantages are worth the effort.

- **Provide elaborative feedback**. We found that many apps did not provide children with the kind of elaborative feedback that could guide children's future learning. We encourage more apps to incorporate elaborative feedback, which provides additional information to children.

- **Provide scaffolding to support conversation**. Children may encounter challenges when interacting with voice interfaces which may undermine the productivity of the apps. We propose that future apps provide conversational scaffolding, which may adjust the conversation based on the child's responses. For example, an app could rephrase the prompt with more accessible language if a child has difficulty responding to the prompt.

- **Introduce a break-down prevention mechanism**. We noticed that many apps crashed silently or discontinued if a child consistently provided responses that the app could not understand. We recommend that apps be designed to be

defensive against breakdown. This could be achieved by either incorporating scaffolding features mentioned above, or in the cases where children fail to respond to the scaffolding prompts, the apps could provide a default fallback response to keep the app going.

CONCLUSION
In this study, we developed a framework to evaluate the educational design features of voice-based apps and then used this framework to evaluate existing literacy-focused voice-based apps. This study presents a case of how to evaluate and design voice apps in light of the extant knowledge of how young children communicate and learn. The design dimensions evaluated in this study, including app learning activities, goal clarity, interactivity, gamifications, conversational prompts, feedback, and scaffolding, may be used as a guideline for future app evaluation and development. Given the emerging prevalence of voice apps for young children, this study may be an important step in actualizing the potential of voice technologies in enriching children's entertainment and educational experiences.

SELECTION AND PARTICIPATION OF CHILDREN
No children participated in this work.

REFERENCES
[1] Lúcia Blondet Baruque and Rubens Nascimento Melo. 2004. Learning theory and instruction design using learning objects. *Journal of Educational Multimedia and Hypermedia* 13, 4 (2004), 343–370.

[2] 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*. ACM, 243.

[3] Laura Benton, Asimina Vasalou, Kay Berkling, Wolmet Barendregt, and Manolis Mavrikis. 2018. A critical examination of feedback in early reading games. In *proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, 373.

[4] 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.

[5] 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.

[6] Sonia Q Cabell, Laura M Justice, Anita S McGinty, Jamie DeCoster, and Lindsay D Forston. 2015. Teacher–child conversations in preschool classrooms: Contributions to children's vocabulary development. *Early Childhood Research Quarterly* 30 (2015), 80–92.

[7] Melissa N Callaghan. 2018. *Connecting Learning and Developmental Sciences to Educational Preschool Apps: Analyzing App Design Features and Testing Their Effectiveness*. Ph.D. Dissertation. UC Irvine.

[8] Melissa N Callaghan and Stephanie M Reich. 2018. Are educational preschool apps designed to teach? An analysis of the app market. *Learning, Media and Technology* 43, 3 (2018), 280–293.

[9] 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*. ACM, 337–348.

[10] Carol McDonald Connor, Frederick J Morrison, and Lisa Slominski. 2006. Preschool instruction and children's emergent literacy growth. *Journal of Educational Psychology* 98, 4 (2006), 665.

[11] Daniela Conti, Alessandro Di Nuovo, Carla Cirasa, and Santo Di Nuovo. 2017. A comparison of kindergarten storytelling by human and humanoid robot with different social behavior. In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 97–98.

[12] Sebastian Deterding, Dan Dixon, Rilla Khaled, and Lennart Nacke. 2011. From game design elements to gamefulness: defining" gamification". In *Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments*. 9–15.

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

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

[15] Sonia Fizek. 2014. Why fun matters: in search of emergent playful experiences. In *Rethinking gamification*. meson press, 273–287.

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

[17] Marina Fridin. 2014. Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education. *Computers & education* 70 (2014), 53–64.

[18] Nancy Garon, Susan E Bryson, and Isabel M Smith. 2008. Executive function in preschoolers: a review using an integrative framework. *Psychological bulletin* 134, 1 (2008), 31.

[19] William W Gaver, John Bowers, Andrew Boucher, Hans Gellerson, Sarah Pennington, Albrecht Schmidt, Anthony Steed, Nicholas Villars, and Brendan Walker. 2004. The drift table: designing for ludic engagement. In *CHI'04 extended abstracts on Human factors in computing systems*. 885–900.

[20] 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.

[21] Arthur Graesser and Bethany McDaniel. 2017. Conversational agents can provide formative assessment, constructive learning, and adaptive instruction. In *The future of assessment*. Routledge, 85–112.

[22] Prajna Das Gupta and Victor Lee. 1995. *Children's cognitive and language development*. Open University Press.

[23] Betty Hart and Todd R Risley. 1992. American parenting of language-learning children: Persisting differences in family-child interactions observed in natural home environments. *Developmental psychology* 28, 6 (1992), 1096.

[24] Erika Hoff. 2006. How social contexts support and shape language development. *Developmental review* 26, 1 (2006), 55–88.

[25] Laura M Justice, Andrew J Mashburn, Bridget K Hamre, and Robert C Pianta. 2008. Quality of language and literacy instruction in preschool classrooms serving at-risk pupils. *Early childhood research quarterly* 23, 1 (2008), 51–68.

[26] Michail Kalogiannakis and Stamatios Papadakis. 2017. An evaluation of Greek educational Android apps for preschoolers. In *proceedings of the 12th Conference of the European Science Education Research Association (ESERA), Research, Practice and Collaboration in Science Education, Dublin City University and the University of Limerick, Dublin, Ireland*. 21–25.

[27] Aisha Kawalkar and Jyotsna Vijapurkar. 2013. Scaffolding Science Talk: The role of teachers' questions in the inquiry classroom. *International Journal of Science Education* 35, 12 (2013), 2004–2027.

[28] Gabriela Kiryakova, Nadezhda Angelova, and Lina Yordanova. 2014. Gamification in education. Proceedings of 9th International Balkan Education and Science Conference.

[29] Jacqueline Kory and Cynthia Breazeal. 2014. Storytelling with robots: Learning companions for preschool children's language development. In *The 23rd IEEE international symposium on robot and human interactive communication*. IEEE, 643–648.

[30] Tracey L Leacock and John C Nesbit. 2007. A framework for evaluating the quality of multimedia learning resources. *Journal of Educational Technology & Society* 10, 2 (2007), 44–59.

[31] Iolanda Leite, Marissa McCoy, Monika Lohani, Daniel Ullman, Nicole Salomons, Charlene Stokes, Susan Rivers, and Brian Scassellati. 2015. Emotional storytelling in the classroom: Individual versus group interaction between children and robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 75–82.

[32] Silvia Lovato and Anne Marie Piper. 2015. Siri, is this you?: Understanding young children's interactions with voice input systems. In *Proceedings of the 14th International Conference on Interaction Design and Children*. ACM, 335–338.

[33] 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*. ACM, 301–313.

[34] Krista R Muis, John Ranellucci, Gregory Trevors, and Melissa C Duffy. 2015. The effects of technology-mediated immediate feedback on kindergarten students' attitudes, emotions, engagement and learning outcomes during literacy skills development. *Learning and Instruction* 38 (2015), 1–13.

[35] Jason Nolan and Melanie McBride. 2014. Beyond gamification: reconceptualizing game-based learning in early childhood environments. *Information, Communication & Society* 17, 5 (2014), 594–608.

[36] Ron Owston, Herb Wideman, Natalia Sinitskaya Ronda, and Christine Brown. 2009. Computer game development as a literacy activity. *Computers & Education* 53, 3 (2009), 977–989.

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

[38] Stamatios Papadakis, Michail Kalogiannakis, and Nicholas Zaranis. 2018. Educational apps from the Android Google Play for Greek preschoolers: A systematic review. *Computers & Education* 116 (2018), 139–160.

[39] Melissa M Patchan and Cynthia S Puranik. 2016. Using tablet computers to teach preschool children to write letters: Exploring the impact of extrinsic and intrinsic feedback. *Computers & education* 102 (2016), 128–137.

[40] Lynn K Perry, Emily B Prince, Adriana M Valtierra, Camila Rivero-Fernandez, Mary Anne Ullery, 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.

[41] Louise Phillips. 2000. Storytelling: The seeds of children's creativity. *Australasian Journal of Early Childhood* 25, 3 (2000), 1–5.

[42] Helena Rasku-Puttonen, Marja-Kristiina Lerkkanen, Anna-Maija Poikkeus, and Martti Siekkinen. 2012. Dialogical patterns of interaction in pre-school classrooms. *International Journal of Educational Research* 53 (2012), 138–149.

[43] Alex Sciuto, Arnita Saini, Jodi Forlizzi, and Jason I Hong. 2018. Hey Alexa, What's Up?: A mixed-methods studies of in-home conversational agent usage. In *Proceedings of the 2018 Designing Interactive Systems Conference*. ACM, 857–868.

[44] Chao-Fen Shih, Chih-Wei Chang, and Gwo-Dong Chen. 2007. Robot as a storytelling partner in the english classroom-Preliminary discussion. In *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*. IEEE, 678–682.

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

[46] Nikita Soni, Aishat Aloba, Kristen S Morga, Pamela J Wisniewski, and Lisa Anthony. 2019. A Framework of Touchscreen Interaction Design Recommendations for Children (TIDRC): Characterizing the Gap between Research Evidence and Design Practice. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. ACM, 419–431.

[47] Silvia Tamayo-Moreno and Diana Pérez-Marín. 2016. Adapting the design and the use methodology of a Pedagogical Conversational Agent of Secondary Education to Childhood Education. In *2016 International Symposium on Computers in Education (SIIE)*. IEEE, 1–6.

[48] Anuj Tewari and John Canny. 2014. What did spot hide?: a question-answering game for preschool children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1807–1816.

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

[50] Fabienne M Van der Kleij, Remco CW Feskens, and Theo JHM Eggen. 2015. Effects of feedback in a computer-based learning environment on students' learning outcomes: A meta-analysis. *Review of educational research* 85, 4 (2015), 475–511.

[51] Stella Vosniadou. 2003. How children learn. *Successful schooling* 16 (2003).

[52] Barbara A Wasik and Mary Alice Bond. 2001. Beyond the pages of a book: interactive book reading and language development in preschool classrooms. *Journal of educational psychology* 93, 2 (2001), 243.

[53] Adriana Weisleder and Anne Fernald. 2013. Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological science* 24, 11 (2013), 2143–2152.

[54] Jacqueline Kory Westlund and Cynthia Breazeal. 2015. The interplay of robot language level with children's language learning during storytelling. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction extended abstracts*. ACM, 65–66.

[55] Jennifer Whorrall and Sonia Q Cabell. 2016. Supporting children's oral language development in the preschool classroom. *Early Childhood Education Journal* 44, 4 (2016), 335–341.

[56] Preben Wik and Anna Hjalmarsson. 2009. Embodied conversational agents in computer assisted language learning. *Speech communication* 51, 10 (2009), 1024–1037.

[57] Joachim Wirth, Josef Künsting, and Detlev Leutner. 2009. The impact of goal specificity and goal type on learning outcome and cognitive load. *Computers in Human Behavior* 25, 2 (2009), 299–305.

[58] Karen Wohlwend and Kylie Peppler. 2015. All rigor and no play is no way to improve learning. *Phi Delta Kappan* 96, 8 (2015), 22–26.

[59] 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*. ACM, CS10.

[60] Ying Xu, Joanna C Yau, and Stephanie M Reich. 2019. The Added Challenge of Digital Reading: Exploring Young Children's Page Turning Behaviors. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children*. ACM, 432–438.

[61] Nicola Yelland. 2011. Reconceptualising play and learning in the lives of young children. *Australasian Journal of Early Childhood* 36, 2 (2011), 4–12.

[62] David B Yoffie, Liang Wu, Jodie Sweitzer, Denzil Eden, and Karan Ahuja. 2018. Voice War: Hey Google vs. Alexa vs. Siri. (2018).