Computers and Education: Artificial Intelligence 10 (2026) 100549

Contents lists available at ScienceDirect

Computers and Education: Artificial Intelligence

journal homepage: www.sciencedirect.com/journal/computers-and-education-artificial-intelligence

Conversational AI in children's home literacy learning: effectiveness, advantages, challenges, and family perception

Shuang Quan [a,*], Xintian Tu-Shea [b], Yi Ding [c], Yao Du [d], Qingxiao Zheng [e], Laney E. Gerdich [a]

[a] Department of Education, Juniata College, 1700 Moore Street, Huntingdon, PA, 16652, USA
[b] Institute for Learning Sciences, University at Buffalo, the State University of New York, Buffalo, NY, 14260, USA
[c] Graduate School of Education, Fordham University, New York, NY, 10023, USA
[d] Caruso Department of Otolaryngology-Head and Neck Surgery, University of Southern California, Los Angeles, CA, 90089, USA
[e] Institute for Artificial Intelligence and Data Science, University at Buffalo, the State University of New York, Buffalo, NY, 14260, USA

A B S T R A C T

This study investigates the effectiveness, affordances, limitations, and family perceptions of conversational AI for home literacy learning vs. human. We developed a large language model (LLM)-powered conversational AI system, named Vovo, to teach children vocabulary and co-construct stories using structured literacy pedagogy. The system was tested in home environments over six weeks with 10 families and their children aged 3–7 ($M = 5.4$). Across 150 learning sessions, Vovo delivered structured literacy instruction as effectively as parents, though children achieved higher learning outcomes when learning with parents. Video analysis revealed Vovo's advantages in pedagogical consistency, language modeling, and verbal socioemotional support, while facing challenges in speech recognition, instructional persistence, nonverbal social cues, and phoneme instruction. Parents perceived Vovo as intelligent, useful, and trustworthy, while expecting a multimodal design to improve engagement. Children perceived Vovo smart and fun but still preferred learning with parents due to emotional bonding. As one of the first studies to embed structured literacy pedagogy into home-based conversational AI system, this research contributes empirical insights into the evolving role of AI in home literacy environments. It also underscores the socially responsive AI design in early education and calls for future design that support parent-child-AI triadic interactions to optimize AI in home literacy learning.

1. Introduction

Literacy development in early childhood lays the foundation for later reading comprehension and academic success (Biemiller, 2013; De Jong & Leseman, 2001; National Reading Panel, 2000; Puckett et al., 2009; Scarborough & Dobrich, 1994; Spear-Swerling, 2019). A rich and supportive home literacy environment (HLE) plays a critical role in shaping children's early literacy development and long-term educational outcomes (Burgess et al., 2002; Sénéchal & LeFevre, 2002; Snow et al., 1998; Van Steensel, 2006; Weigel et al., 2006). According to the HLE model (Sénéchal & LeFevre, 2002), home literacy activities can be categorized into two major types: formal HLE activities (e.g., caregivers directly teach children to read and write) and informal HLE activities (e.g., shared reading, storybook exposure, and parent-child conversations) (Sénéchal & LeFevre, 2002). While research demonstrates that both types of activities predict children's later reading performance, formal activities play a particularly critical role in fostering long-term reading outcomes, school readiness, and general academic success (Araújo & Costa, 2015; X. Chen et al., 2010; De Jong & Leseman, 2001; Dickinson & Tabors, 2001; Hood et al., 2008; Leseman & De Jong, 1998; Sénéchal & LeFevre, 2002; Sonnenschein & Munsterman, 2002; S. Z. Zhang et al., 2020). However, formal home literacy practices vary widely across families. Families facing barriers such as lower socioeconomic status, diverse linguistic backgrounds, limited time availability, or lack of educational training may engage in literacy activities less frequently or in ways that are less pedagogically effective. These disparities contribute to early gaps in literacy development (Hammer et al., 2003; Hart et al., 1997; Karpava, 2021; Van Steensel, 2006). Addressing these early disparities remains a persistent challenge in the field, prompting the need for reliable and accessible home literacy support.

With the advancement of artificial intelligence (AI) and large language models (LLMs), new opportunities have emerged to support children's literacy development at home. Recent research has started introducing AI-enhanced tools in literacy development, such as chatbots (Zhang & Huang, 2024), interactive TV and video systems (Tsang et al., 2025; Xu et al., 2024), educational toys and robots (Ho et al., 2025; Kewalramani et al., 2021), augmented and virtual reality systems (Lyu et al., 2024; Zhang, 2025, pp. 1620–1621), and storytelling mobile

* Corresponding author.
E-mail addresses: squ100@juniata.edu (S. Quan), xintiant@buffalo.edu (X. Tu-Shea), yding4@fordham.edu (Y. Ding), yaodu@usc.edu (Y. Du), qingxiao@buffalo.edu (Q. Zheng), gerdile23@juniata.edu (L.E. Gerdich).

https://doi.org/10.1016/j.caeai.2026.100549
Received 9 October 2025; Received in revised form 30 December 2025; Accepted 21 January 2026
Available online 24 January 2026
2666-920X/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

applications (Chen, 2024; Dietz Smith et al., 2024, pp. 408–423; Lee et al., 2025; Zhang et al., 2022, pp. 1–21). These AI-enhanced innovations provide young children with multimodal, personalized, and adaptive learning experiences that positively impact literacy development (Saddhono et al., 2024, pp. 856–861; Solichah & Shofiah, 2024; Su & Yang, 2023b; Wang et al., 2024; Wen et al., 2024; Yang, 2022). LLMs such as ChatGPT and Gemini have enabled new forms of interactive, generative technologies with the potential to augment human teaching and caregiving (Quan et al., 2024, 2025a, 2025b; Xu et al., 2022). These systems can produce fluent, context-sensitive responses, maintain conversational coherence, and adapt in real time to user input. As a result, they attracted growing interest as potential home literacy learning companions, including for storybook reading (Chen, 2024; Zhang et al., 2022, pp. 1–21), learning second languages (Zhang & Huang, 2024), and collaborative problem-solving (Xu et al., 2023, pp. 160–172; Zhang et al., 2022, pp. 1–21). However, most applications have focused on informal literacy activities such as shared reading, storytelling, or dialogic conversation, with limited exploration on formal HLE activities, such as explicit, structured teaching of foundational literacy skills. While conversational AI systems demonstrated promise in generating consistent and developmentally appropriate dialogue, their potential in effective formal HLE instruction remains largely underexplored.

Structured Literacy Pedagogy is an instructional approach that teaches foundational literacy skills through structured, explicit, and systematic instruction (Spear-Swerling, 2019). It has been recognized as an evidence-based effective method for developing early literacy, and has been widely adopted in classrooms and clinical settings across the country. While introducing structured literacy pedagogy into home learning environments, particularly through formal literacy activities, holds strong promise, its adoption in HLE remains limited due to caregivers' limited professional competencies. Conversational AI systems present a promising potential to this challenge. By embedding pre-scripted conversational flow, these systems make it possible to deliver structured literacy pedagogy at home in a consistent, pedagogically grounded manner.

This study addresses this critical gap by developing Vovo, a conversational AI system designed to deliver formal literacy instruction in home environments. Grounded in structured literacy pedagogy, Vovo guides children through a seven-step vocabulary learning sequence to support their engagement in formal home literacy activities. By implementing and evaluating the system in home settings, this study provides empirical evidence on the effectiveness, affordances, and limitations of conversational AI for formal home literacy learning. It also explores families' perceptions of the system, offering insights into how AI can be designed to meet families' needs and better support literacy development in home environments.

2. Related work

2.1. Early literacy development and home literacy environment

Early literacy refers to the foundational set of language, cognitive, and social-emotional skills that children begin to develop from birth and continue to build during the preschool years, before formal schooling begins (Puckett et al., 2009; Snow et al., 1998). According to the National Reading Panel (2000), early literacy encompasses a broad range of emergent skills that are strong predictors of later reading and academic achievement, such as alphabet knowledge, phonological awareness, print knowledge, vocabulary, and narrative skills. Among these, vocabulary knowledge and narrative skills are especially important, as they directly influence children's reading comprehension and long-term academic performance (Biemiller, 2013; Kilpatrick, 2015; Moats, 2020; Scarborough & Dobrich, 1994). However, children of low socio-economic status (SES) or linguistically diverse backgrounds are at a greater risk of falling behind in these areas. Research has shown that by age four, children from high SES families are exposed to approximately 30 million more words than their peers in low SES households (Hart et al., 1997), and children in homes where English is not the primary language often receive less input in English vocabulary and grammar, leading to delayed language proficiency in academic contexts (Hammer et al., 2003). These disparities are largely shaped by differences in home literacy environments.

The Home Literacy Environment (HLE) refers to the range of experiences within the home that influence children's early language development (Roskos & Neuman, 2003; Scarborough & Dobrich, 1994; Zimmerman, 2018). It encompasses literacy-related resources, daily literacy activities, and parent-child interactions about language and print (Bus et al., 1995; De Jong & Leseman, 2001; Zimmerman et al., 2008). According to the HLE model, HLE can be categorized into two broad types: formal activities, where parents directly teach children to read and write (e.g., direct teaching of letters, sounds, or vocabulary), and informal activities, such as storybook exposure and shared storybook reading (Sénéchal et al., 1998; Sénéchal & LeFevre, 2002). Both formal and informal activities contribute significantly to children's language growth and future reading success. However, families from low SES and linguistically diverse backgrounds often face barriers to offering rich HLE experiences to their children. Parents in these contexts may lack the time, skills, and language proficiency necessary to maintain consistent engagement in home literacy learning (Bus et al., 1995; Hammer et al., 2003; Van Steensel, 2006). As a result, children in underserved households may not receive the same level of early literacy support as their more advantaged peers. These disparities are especially evident in formal literacy instruction, which requires sustained parental involvement, pedagogical knowledge, and fluency in the language of instruction (Sénéchal & LeFevre, 2002). These challenges underscore the pressing need for accessible, language-rich, and pedagogically informed HLE tools to support early literacy development at home, particularly for children in under-resourced settings.

2.2. Literacy instruction and structured literacy

Structured literacy is an evidence-based instructional approach that emphasizes the explicit, systematic teaching of language and literacy components, including phonology, morphology, syntax, and semantics (Levey & Polirstok, 2011; Moats, 2020; Spear-Swerling, 2019). This method has been shown to significantly improve foundational literacy outcomes for learners across developmental stages, particularly for children of diverse linguistic and SES backgrounds or at risk of reading difficulties (Kilpatrick, 2015; Moats, 2020). Structured literacy has gained widespread adoption in school-based curricula and teacher preparation programs due to its robust empirical support and its alignment with national and state literacy initiatives (National Reading Panel, 2000). However, its implementation in home literacy environments remains limited. Many parents, particularly those from linguistically diverse or low-SES backgrounds, may lack the pedagogical knowledge or confidence to implement structured approaches at home, relying instead on informal or incidental literacy activities (Hammer et al., 2003; Sénéchal & LeFevre, 2002). This gap creates a missed opportunity to extend high-quality, developmentally appropriate literacy instruction beyond the classroom and into children's everyday environments. Given the importance of early and consistent exposure to structured language input, there is a growing need to design tools and interventions that make structured literacy more accessible and feasible for families, thus improving children's literacy trajectories through integrated school-home partnerships.

2.3. Conversational AI tools in home literacy learning

As LLMs are increasingly integrated into HLE, understanding their potential role in HLE practices has become an increasingly important question for both educational and Human-Computer Interaction (HCI)

researchers. Studies suggest that these AI tools can provide structured instruction (Leech et al., 2025; Siegle et al., 2023), boost motivation and engagement (Xu et al., 2022; Yıldız, 2023), and support early literacy skills (Durski et al., 2020; Han & Cai, 2023, pp. 470–474; Han et al., 2024; Maspul, 2025). Recent advances in AI and LLMs offer new possibilities for supporting children's literacy development, particularly through scholarly work stemming from research in Conversational Agents (CAs) and Intelligent Tutoring Systems. AI-powered CAs can serve as instructional agents and learning partners by providing adaptive feedback (Yaseen et al., 2025), pronunciation support (Xiao et al., 2024), rich lexical output (He et al., 2024; Xu et al., 2021; Yang et al., 2024), reading companion (Durski et al., 2020), and interactive dialogue (Xu et al., 2022, 2023, pp. 160–172; Zhang et al., 2025) to enrich literacy learning experiences. However, it is important to note that providing formal literacy experiences to children requires direct teaching, such as explicitly teaching letters, phonemes, and vocabulary (Burgess et al., 2002; Evans et al., 2000; Hood et al., 2008; Sénéchal et al., 1998; Sénéchal & LeFevre, 2002). This contrasts with implicit or whole-language approaches, which tend to emphasize contextual learning through open-ended conversations, as seen in most AI in HLE research. As a result, a key limitation of many existing CAs lies in their reliance on unstructured dialogue with limited incorporation of pedagogically grounded guidance (Druga et al., 2017, pp. 595–600; Hubbard et al., 2021, pp. 1–10; Pérez-Marín & Pascual-Nieto, 2013). In addition, published research has focused on informal home literacy activities in controlled settings (e.g., lab, school) delivered by researchers or educators (Létourneau et al., 2025; Mousavinasab et al., 2021; Pitura, 2024; Temiz & Kafadar, 2025), with limited evidence on their effectiveness in delivering formal literacy instruction in naturalistic home settings.

Another stream of studies has been focusing on Intelligent Tutoring Systems for their potential to deliver pedagogically grounded educational guidance (Alam, 2023; Mousavinasab et al., 2021; Ni & Cheung, 2023; Schmucker et al., 2024, pp. 75–90; Wang, 2025; Xiao et al., 2025). By leveraging adaptive technologies and domain-specific scaffolding, Intelligent Tutoring Systems can personalize instruction, monitor learner progress, and provide real-time feedback aligned with educational goals (Létourneau et al., 2025; Liu et al., 2025; Mousavinasab et al., 2021; Wang et al., 2023). However, the technical affordances of Intelligent Tutoring Systems are not adequate to support human-like instructional delivery for complex language and literacy skills. For instance, a systematic review by Mousavinasab et al. (2021) found that the most frequently used AI techniques for Intelligent Tutoring Systems were action-condition rule-based reasoning, data mining, and Bayesian networks, which were not adequate to support the complexity of interactive scaffolding necessary for adaptive feedback and personalized learning for literacy instruction at home (Yaseen et al., 2025; Yigitalieva et al., 2024, pp. 347–353). Another systematic review by Letourneau et al. (2025) found positive learning performance of ITS in 28 studies conducted with a total of 4597 K-12 students who used ITS for various subjects. However, many studies used shorter interventions (half lasting less than a week, some as brief as a single-class period) and a small sample size. Additionally, 82 % of the studies focused on STEM subjects, and less attention was given to foundational subjects such as language arts and language learning.

Implementing AI-powered literacy instruction tools to support HLE is a promising future direction, but further research is needed to clarify how it compares directly to traditional human-led instruction in both educational and home settings. In early childhood and special education, previous studies have shown that AI-powered tools can significantly improve literacy outcomes, including vocabulary acquisition and reading comprehension, often performing on par with or slightly better than traditional teacher-led methods in certain contexts (Cha & Daud, 2025; Hsu et al., 2023; Lee et al., 2023; Mostow et al., 2003; J. Wu, 2024). AI tools provided personalized and engaging experiences, but concerns about socio-emotional support and over-reliance on technology were noted (Cha & Daud, 2025; Li et al., 2024). Blended models that combine AI tools with the guidance of a human instructor consistently yielded the most robust outcomes, leveraging the strengths of both approaches (Cha & Daud, 2025; Chiu et al., 2023; Lai, 2025; Z. Wu et al., 2024). These models enhance engagement, support self-regulation, and maintain the critical human element necessary for holistic literacy development. Although many studies show that children are capable of engaging in interactive learning interventions that integrate AI and benefit both cognitive and emotional growth (Chen, 2024; Solichah & Shofiah, 2024), other challenges remain, including lack of parent guidance, limited curriculum design for home settings, and the need for more long-term studies to assess the sustained impact of AI on literacy (Chen, 2024; Solichah & Shofiah, 2024; Su & Yang, 2023b; Yang, 2022).

2.4. Parents' perceptions of AI in home literacy learning

Many parents view AI technologies positively, recognizing their potential to improve children's learning experiences and support early literacy, especially when tools are designed to be engaging and user-friendly (Han & Cai, 2023, pp. 470–474; Su, 2024; Su & Yang, 2023a). For example, Tsang et al. (Tsang et al., 2025) found that parents appreciate AI-powered systems for their ability to personalize and engage children, and these tools can foster collaboration and self-reflection in family literacy activities. When explicitly comparing parent-led vs. AI-guided dialogic reading in an exploratory randomized clinical trial (RCT), Xiao et al. (Xiao et al., 2025) found dialogic reading with CAs or with parents demonstrated comparable benefits for children's vocabulary, comprehension, and retelling in post- and delayed tests in e-book reading. AI-based tools can also influence parental involvement in early literacy development in several ways. In a survey with 74 U.S. parents with children between five and eight years old, Quan et al. (2024) found more than 50 % parents demonstrated positive attitudes towards ChatGPT use for parenting tasks, which further led to encouragement for their children to use ChatGPT. Qualitative interviews with parents with children under 13 years old (Quan et al., 2025a) further indicated that both play-driven and knowledge-driven conversations, especially related to various disciplinary literacy on STEM (e.g., math, biology, astronomy) and language art (e.g., reading, writing). Parents are motivated to adopt AI-based tools when they perceive both practical (utilitarian) and enjoyable (hedonic) benefits, and these motivations are strengthened by the tool's autonomy, human-like interaction, and parents' openness to innovation (Lin et al., 2023). However, parents also report practical concerns, such as children's limited comprehension abilities, uncertainty about the appropriateness of AI content, and time constraints which limit their supervision or support for their children's AI use (Su, 2024). Su (2024) surveyed 215 kindergarten parents and reported perceived acceptance for using AI-powered literacy tools at the kindergarten level, yet their children often experience barriers such as limited comprehension ability with AI-generated learning content, time constraints, and uncertainty regarding the appropriateness of AI learning content. Other researchers also expressed concerns about parents' equitable access, cultural sensitivity, and the need for comprehensive frameworks to guide the integration of AI into home literacy practices (Chen, 2024; Kucirkova & Hiniker, 2024; Su & Yang, 2023b). For instance, Kucirkova and Hiniker (2024) found that Norwegian parents were concerned that conversational AI may hinder family values of literacy education, family bonding, and habit cultivation. Specifically, parents were concerned about CAs' inability to complex, non-standardized conversations with flexible language use in a typical human-to-human conversation.

Therefore, a deeper understanding of the diverse needs, concerns, and expectations of families is critical to the successful adoption and long-term engagement with AI-powered literacy tools.

2.5. Research questions

A growing body of research highlights the potential of AI to support

home literacy learning, with recent advancements in conversational AI and intelligent tutoring systems demonstrating their abilities to provide personalized, adaptive, and engaging educational experiences. However, most existing AI tools focus on informal literacy activities and are typically implemented in controlled and researcher-led settings. There is a limited comprehensive evaluation of these tools in naturalistic home environments, particularly in the context of formal, direct literacy instruction. In addition, little is known about how parents and children perceive these tools or how their usage experiences shape those perceptions. These gaps highlight the need to develop and rigorously evaluate AI-powered systems that provide pedagogically informed, formal literacy support within home environments, particularly for families facing challenges such as limited time, instructional knowledge, or language proficiency.

To address this gap in current research, the present study designed a conversational AI system grounded in the structured literacy pedagogy to deliver formal literacy instruction in home environments. The aim of this study is to investigate; (1) whether Vovo can successfully deliver structured instruction and engage children in meaningful language use as human caregiver, (2) to compare performance across Vovo-led and parent-led sessions, identifying the unique benefits and challenges of AI system compared to human, and (3) what are the parents and children's attitudes and perceptions toward using AI as a home literacy learning tool. By situating the AI system within real home environments and implementing it with natural child interaction, we seek to better understand AI-powered tools as potential partners in early language development. Three research questions guided the current study:

- RQ1: How effective is a conversational AI system in supporting children's literacy learning at home?
- RQ2: What are the advantages and challenges of the conversational AI system in literacy learning at home?
- RQ3: How do children and parents perceive conversational AI as a home literacy learning tool?

3. Methods

3.1. Study design

This study developed an LLM-powered conversational AI system, named Vovo, to support children's vocabulary learning and creative storytelling in home environments. A within-subject design was employed where each child participated in both AI-led and parent-led instructional conditions during each learning session, in order to compare the AI system's performance with human caregivers while minimizing cross-subject variability. A six-week micro-longitudinal study was conducted to capture the consistency of the system over time, facilitating a nuanced understanding of both children's learning experiences and parents' perceptions of the AI system in naturalistic home settings.

3.2. Design of Vovo

Vovo is a web-based conversational AI system designed for home literacy learning, implemented as a single-page application accessible through a standard web browser. The system integrates multiple technical components to support real-time, child-facing interaction. At its core, Vovo is powered by OpenAI's GPT-4o large language model, which generates instructional dialogue grounded in structured literacy pedagogy through carefully engineered prompts. Speech input is processed using automatic speech recognition (ASR), while speech output is delivered via a synthetic text-to-speech (TTS) voice selected for warmth, clarity, and child-friendly prosody. The system supports real-time turn-taking with interruption handling, allowing children to interject at any point during instruction. In parallel, on-screen transcription using Deepgram Nova-4 is displayed during the interaction to provide visual reinforcement of spoken language, supporting printed language exposure. Vovo uses a cartoon skunk avatar as its visual identity to foster playful engagement with young children. Fig. 1 demonstrates the interface of Vovo.

The conversational flow was grounded in the structured literacy pedagogy of vocabulary instruction (Biemiller, 2013; Spear-Swerling, 2019), incorporating a seven-step instructional sequence that targets three key vocabulary learning goals: accurate pronunciation, word meaning, and contextual usage (seven-step instructional sequence is demonstrated in Fig. 1). Target vocabulary was drawn from the Berkeley Unified School District Academic Vocabulary List and categorized into three developmental levels based on children's baseline language proficiency (for more details, see the Measurement section). Each session began with a friendly greeting from Vovo, "*Hey buddy! Can you tell me which team you're in—green, blue, or red?*", prompting the child to select a level corresponding to their assigned vocabulary set. After confirming the level, Vovo randomly selected a word from the corresponding list and asked the child whether they already knew the word. If the child confirmed they were unfamiliar with it, Vovo initiated the structured instructional sequence, guiding them through pronunciation, definition, and contextual use. Vovo's output was prompted to maintain a developmentally appropriate language and engaging tone by incorporating simplified phrasing, active scaffolding, adaptive feedback, and friendly expressions tailored to young children.

The system relies primarily on verbal input/output and does not track gaze, facial expressions, or body movement. The platform operated without requiring login credentials and used minimal data collection. All Vovo sessions were conducted using laptop, tablet, or desktop computer through a web browser. Although the system is technically accessible on mobile devices, smartphones were not used during data collection to ensure consistency in screen size, audio quality, and interaction conditions across participants. As a result, all reported quantitative and qualitative findings reflect children's interactions with Vovo on laptop, tablet or desktop platforms only. Prior to deployment, Vovo underwent in-lab pilot testing to ensure system stability, content safety, and usability.

3.3. Participants

This research is approved by the IRB committee at the first author's institution. After receiving IRB approval, the research team recruited participants through personal and professional networks. All participants consented to be part of the research. Inclusion criteria included: (1) child age between 3 and 7 years, (2) English as the primary home language, and (3) no reported developmental delays. The final participants included 10 children (5 girls, 5 boys), ages 3.7–7.7 years (M = 5.4). Demographic information is shown in Table 1. Descriptive statistics showed that children's average age was 5.8 years (M = 68 months, SD = 15.80). 70 % of children identified as White, 20 % as multiracial, and 10 % as Hispanic. All children's language proficiency fell within the typical developmental range (M = 106.30, SD = 17.08) (Dunn, 2019). All parents reported English as the primary language used at home. In terms of socioeconomic status (SES), the study included families across a broad household income range, from $35,000-$49,999 to $150,000-$199,999, with 50 % of participants declining to report income and 10 % reporting incomes below the national average.

3.4. Measurement

3.4.1. Baseline language proficiency

To accommodate language proficiency levels and ensure developmentally appropriate instruction, children's receptive vocabulary was assessed prior to the study using the Peabody Picture Vocabulary Test, Fifth Edition (PPVT-5) (Internal reliability = .95) (Dunn, 2019). Based on children's standard scores and age-equivalent levels, children were grouped into three instructional tiers that matched their developmental

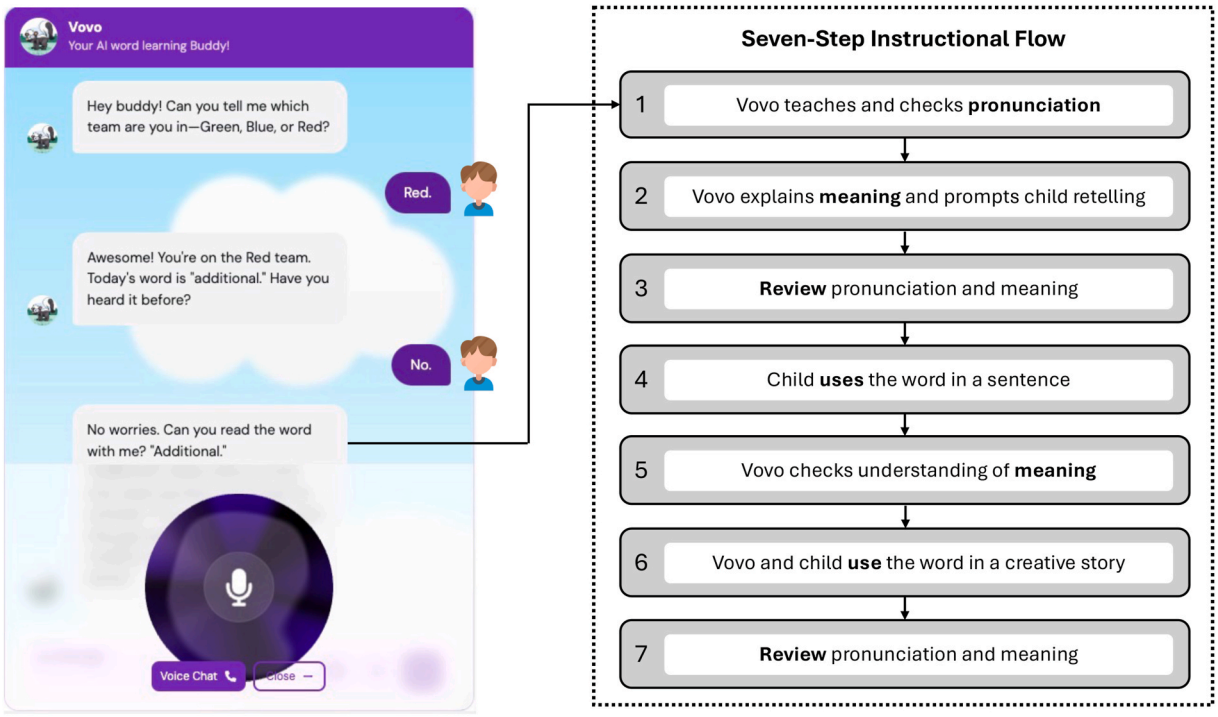


Fig. 1. Vovo's interface and instructional flow.

Table 1
Demographic information of participants.

Child ID	Age (Years: Months)	Gender	Race	Hispanic	Language Proficiency	Parent ID	Caregiver	Age (Years)	Family Income	Prior AI Usage
C1	5:10	Boy	Multiracial	No	106	P1	Mother	40	$35,000–$49,999	Sometimes
C2	5:10	Boy	White	No	94	P2	Father	30	Not Specified	Never
C3	5:01	Girl	White	No	119	P3	Father	30	Not Specified	Never
C4	4:10	Boy	White	No	113	P4	Mother	–	Not Specified	Not Specified
C5	6:01	Boy	White	No	126	P5	Mother	33	Not Specified	Sometimes
C6	5:02	Girl	White	No	102	P6	Grandmother	52	$100,000–$149,999	Never
C7	7:07	Boy	White	No	135	P7	Father	40	$150,000–$199,999	Rarely
C8	3:07	Boy	White	No	92	P8	Father	40	$150,000–$199,999	Rarely
C9	7:05	Girl	White	No	93	P9	Other Caregiver	19	Not Specified	Sometimes
C10	4:03	Boy	Multiracial	Yes	78	P10	Mother	25	$150,000–$199,999	Never

Note. Language Proficiency = Standard Score of Peabody Picture Vocabulary Test – Fifth Edition (PPVT-5).

stage: Green (Grade 1), Blue (Grade 3), and Red (Grade 5). The vocabulary used in this study was purposefully selected to focus on high-utility academic words that is essential for reading comprehension but not often acquired through informal HLE activities and everyday conversation. Research has shown that early exposure to academic vocabulary contributes significantly to school readiness and long-term academic achievement (Beck et al., 2013; Biemiller, 2013). By introducing these words in a structured and supported way at home, our study aimed to build the foundation for children's future academic success. The Berkeley Unified School District (BUSD) Academic Vocabulary List was selected for this study because it draws from a range of well-established academic sources, including the Coxhead Academic Word List, the English for Academic Purposes (EAP) Word List, and Dr. Kate Kinsella's Academic Word Lists. The selection criteria emphasized word frequency, morphological diversity (e.g., roots, prefixes, suffixes), and alignment with the Common Core Literacy Standards, ensuring that the chosen words are both age-appropriate and likely to appear across multiple subject areas and grade levels.

3.4.2. *Instructional delivery*

To evaluate the effectiveness of Vovo in delivering instruction and compare with a human caregiver, instructional delivery was evaluated by scoring how many instructional steps out of seven were successfully delivered during the session. A step was marked as delivered if the AI or parent initiated the specific instructional prompt (e.g., *"Can you read the word with me? Additional."*) regardless of the child's response. Each step was coded using binary code (1 = completed; 0 = not completed), allowing each session to receive a total score ranging from 0 to 7.

3.4.3. *Learning outcome*

Based on language development and structured literacy pedagogy (Beck et al., 2013; Coyne et al., 2007; Levey & Polirstok, 2011; Spear-Swerling, 2019), children's learning outcomes were assessed across three key learning goals: pronunciation, meaning, and contextual

use. Each skill was scored from 2 = accurate response, 1 = partially correct/imprecise response, to 0 = incorrect/no response. A mean score across three areas was calculated to represent the child's overall learning outcome for each word in one learning session.

3.4.4. Family's perception

Semi-structured interviews were conducted after the intervention to explore both parents' and children's perceptions of Vovo. The parent interview protocol was adapted from a developed framework by the authors (Quan et al., 2024), grounded in the Technology Acceptance Model (TAM). It included items addressing perceived usability, instructional value, trustworthiness, and design expectations. The children's interview protocol was developed by the research team based on the Godspeed Questionnaire Series (Bartneck et al., 2009), and focused on children's general perceptions of AI, their learning experiences with Vovo, and their preferences regarding AI versus parent-led instruction. The interview protocols are detailed in Table 2.

3.5. Study procedure

To explore in-depth interaction patterns, we conducted a six-week home-based study in which each child was expected to engage in learning sessions with both Vovo and their parent three times per week. While the goal was consistent engagement over six consecutive weeks, families were given flexibility in scheduling to accommodate individual routines. The study consisted of three phases: (1) Kick-off Zoom Session; researchers conducted a baseline vocabulary assessment to evaluate each child's baseline language proficiency and provided a walkthrough of Vovo's features and instructions to both parents and children. (2) Six-Week Interaction Period; children engaged in learning sessions with Vovo and their parents, separately, at home. The research team monitored the learning process via online Zoom meetings and recorded the interactive process for later analysis. (3) Debrief Zoom Session; both children and parents participated in a semi-structured interview where they shared their perceptions and experiences using Vovo. All sessions were conducted via Zoom and video-recorded for transcription and further analysis. All child interviews were conducted under the supervision of a parent or guardian. Fig. 2 demonstrates the procedure of the study.

3.6. Data analysis

In our data analysis, we employed a mixed-methods approach that combined both quantitative and qualitative analysis. For the quantitative analysis, we assessed Vovo's instructional delivery and children's learning outcomes across both AI-led and parent-led conditions. We then conducted paired-samples t-tests to examine whether there were statistically significant differences between the two conditions. To identify and account for potential confounding variables, we conducted Pearson correlation analyses, which revealed that age and prior language proficiency were correlated with learning outcomes. To mitigate these confounds, we used standardized pre-assessment tools and compared participants using age-normed standard scores, allowing us to interpret learning gains relative to each child's developmental language level.

For the qualitative analysis, we employed inductive thematic analysis (Braun & Clarke, 2006) to analyze both the video-recorded learning sessions and the post-intervention parent interviews. The procedure began with the researcher reviewing all video recordings to identify recurring keywords, salient moments, and illustrative quotes for open coding. These excerpts were used to generate preliminary themes. Next, the researcher conducted another round of systematic video review, this round with detailed note-taking and partial transcription of key episodes. The transcripts and field notes were then organized under the emergent themes, with iterative refinement to ensure alignment with the rubric categories. In qualitative research, establishing validity involves demonstrating that the process used to develop interpretations is credible and well supported by the data (Creswell & Miller, 2000). Therefore, we establish validity through a combination of thick description and triangulation. In this study, the lead author reviewed the data codes and re-watched the data with other researchers to ensure that the cases presented in the findings are representative and are illustrated with rich context. Discrepancies were resolved through discussion to ensure the inter-coder reliability. The video analyses allowed us to complement the quantitative findings with descriptive insights into the different instructional dynamics present in AI versus human-led sessions. Data cleaning, descriptive statistics, and statistical analyses were conducted using R version 4.4.1. Random intercepts were used where appropriate to account for the nested structure of the data (i.e., repeated sessions within children).

Table 2
Interview protocols for parents and children.

Parents' Interview Protocol	Children's Interview Protocol
1. What are your overall impressions of Vovo as a learning tool for your child	1. What do you think Vovo is like? Is Vovo smart, friendly, and good at listening? Why?
2. How helpful do you think Vovo is for sup- porting your child's literacy learning?	2. Did Vovo help you learn something new? What's a cool word Vovo taught you, and what does it mean?
3. How much do you and your child enjoy using Vovo?	3. Was it fun to play and talk with Vovo? Why?
4. Do you trust Vovo to interact with your child? Do you have any concerns about your and/or your child's use of Vovo?	4. When you talk to Vovo, do you feel safe and comfortable, or do you feel nervous? Why?
5. How likely are you to continue using or try other conversational AI tools (or allow your child to use them)?	5. Who do you prefer to learn with, Vovo or your parents? Why?
6. Do you have any suggestions on the future development of Vovo?	6. Let's pretend you're the super designer of Vovo and can make it even cooler! What would you keep the same? What would you change?

4. Findings

In this section, we first present the quantitative findings that show how effective the AI system was in supporting literacy learning at home. Next, we draw on qualitative data to help explain and add context to those results. Finally, we explore how families perceived this experience, offering a comprehensive picture of what learning literacy with the AI system was like at home.

4.1. RQ1: How effective is a conversational AI system in supporting children's home literacy learning?

4.1.1. Descriptive statistics

Over the six weeks, child participants completed a total of 150 instructional sessions, with 84 AI-led and 66 parent-led sessions. Across the sample, children participated in an average of 15 learning sessions ($SD = 7.7$). Participation ranged from 7 to 28 sessions, with a median of 12.5 sessions, indicating moderate variability in engagement across children. The average length of each session is 3 min 40 s ($SD = 1.59$). On average, children completed 4.47 out of 7 instructional steps ($SD = 1.64$) and achieved a mean outcome score of 1.14 out of 3 ($SD = .69$) across both instructional conditions. Correlation analyses revealed that age was significantly associated with language proficiency, instructional delivery, and learning outcomes. Language proficiency was correlated with learning outcomes and instructional delivery. Additionally, learning outcomes were associated with instructional delivery and learning conditions (see Table 3). The Shapiro-Wilk test was employed to examine the assumption of normality of the difference scores between paired conditions. Results indicated that the difference scores for both instructional delivery ($W = .955$, $p = .727$) and learning outcomes ($W = .937$, $p = .517$) did not significantly deviate from normality. Therefore,

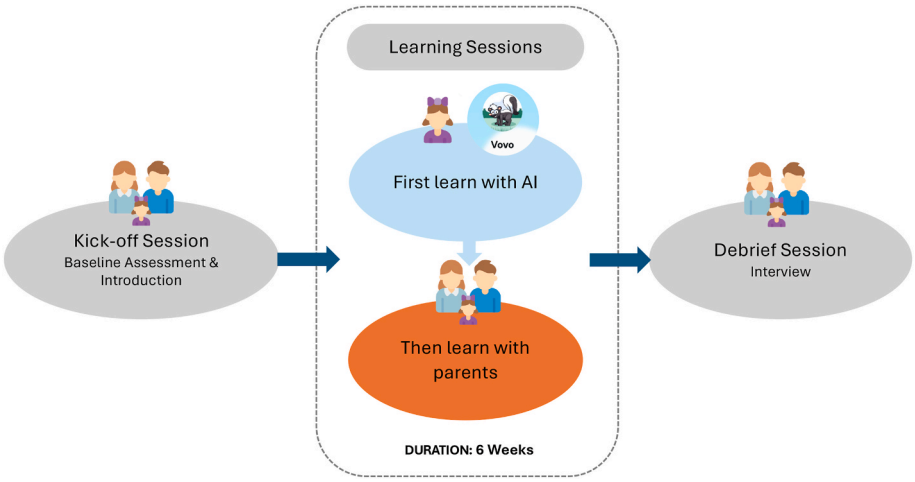


Fig. 2. Study design.

Table 3
Descriptive statistics and correlations among study variables.

Variable	M	SD	Min	Max	Age	Lan.	Ins.	Lea.
Age	68.75	15.80	43	89	1			
Language Proficiency	106.30	17.08	78	135	.34***	1		
Instructional Delivery	4.47	1.64	0	7	.52***	.30***	1	
Learning Outcome	1.14	.69	0	2	.71***	.42***	.66***	1

Note. Correlations are shown below the diagonal. *$p < .05$, **$p < .01$, ***$p < .001$.

the normality assumption for the paired-samples t tests was met.

4.1.2. Instructional delivery

To examine Vovo's effectiveness as a home literacy learning partner, we first compared how well it delivered target learning tasks to children versus parents. Instructional delivery was evaluated by scoring the number of instructional steps out of seven that were successfully delivered by the system or parents, separately. For each child, an average delivery rate was calculated across all sessions over the six-week period for both conditions. Then we conducted a paired-sample *t*-test to examine whether there was a significant difference in instructional delivery between Vovo and human caregivers. Results indicated no statistically significant difference between the two conditions, with $t(9) = 1.70$, $p = .123$, Cohen's $d = .38$. Although the difference was not statistically significant, the observed small-to-moderate effect size ($d = .38$) suggests that Vovo demonstrated a similar instructional effect in comparison to parents. This indicates that Vovo provided consistent instruction across sessions, acting as a consistent instructional partner similar to a human caregiver (Fig. 3).

4.1.3. Learning outcomes

To further evaluate Vovo's effectiveness in supporting literacy learning at home, we conducted a paired-samples *t*-test to compare the learning outcomes between the AI-led and parent-led conditions. Three core vocabulary skills, pronunciation, meaning, and use in context, were assessed to determine whether the AI could effectively support the

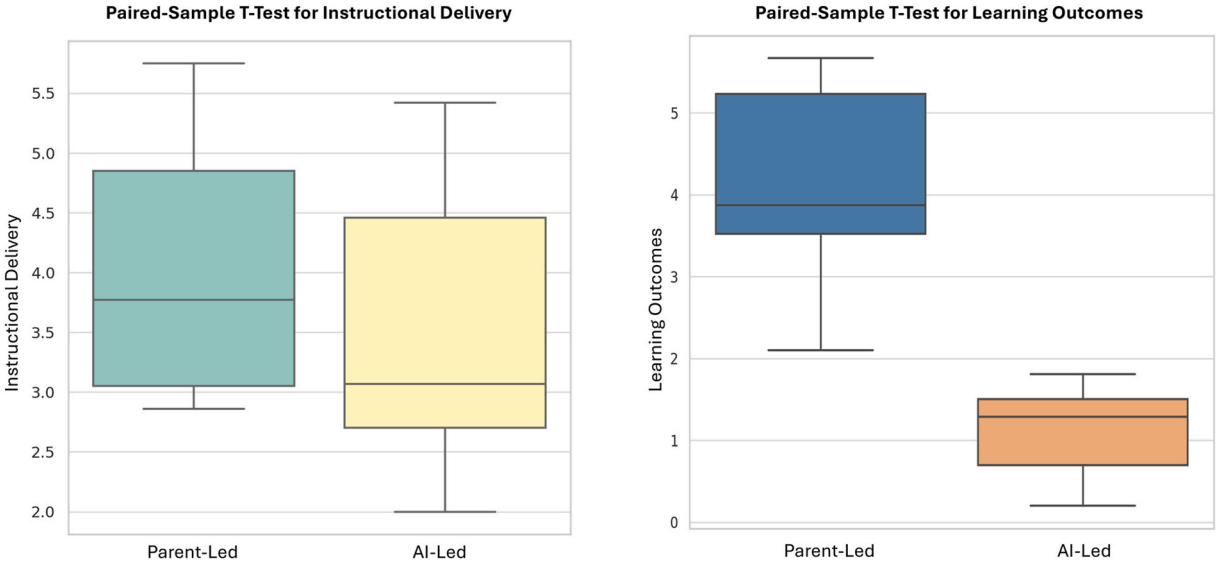


Fig. 3. Paired Sample T-Test for Effectiveness of AI vs. Parent-Led Session.

development of essential early literacy skills. For each child, we calculated the mean learning outcomes value across all sessions with Vovo and with a parent. The results indicated a statistically significant difference between AI-led and parent-led sessions, with learning outcomes in parent-led sessions significantly higher than in AI-led sessions (t (9) = 11.83, $p < .001$, Cohen's d = 3.74). The large effect size ($d = 3.74$) indicates a substantial advantage for parent-led instruction, suggesting the strong influence of parent-led learning within the home context. This finding suggests that children achieved greater learning gains when instruction was led by parents rather than the AI system, even when both conditions followed the same instructional sequence (See Fig. 3).

In summary, while the AI system provided an effective instructional delivery comparable to that of human caregivers, parent-led sessions resulted in more substantial improvements in children's literacy learning outcomes.

4.2. RQ2: What are the observed advantages and challenges of the conversational AI system in home literacy learning?

To better understand the quantitative findings on the AI system's effectiveness, we conducted video analyses to capture more nuanced aspects of the interactions during both AI-led and parent-led sessions. Through qualitative analysis, we identified the system's distinct advantages, which positively contributed to the learning process, as well as challenges that hindered learning and are in need of refinement in future design.

4.2.1. Advantages

Advantage in Pedagogical Advancement. The AI-led sessions demonstrated a high level of pedagogical advancement in instruction by (1) consistently delivering the **structured instructions**, (2) providing **effective scaffolding**, and (3) offering **personalized learning experiences**. First, the system was able to follow the scripted structured literacy instructional sequence with precision. The instructional delivery showed strong consistency, with no notable failures on the pedagogical level. Second, the system was able to provide effective scaffolding by offering clarifying prompts, elaborations, and contextual examples when children showed hesitation or confusion. Parent-led sessions occasionally revealed limitations in instructional scaffolding. Some caregivers (e.g., P1, P5) defaulted to repeating dictionary definitions without simplification or contextualization, while others (e.g., P1, P2) responded to children's disengagement by reiterating instructions without adjustment, such as *"I'll say the definition one more time. Limited means restricted in size, amount, or extent. Can you tell me what it means?"* (P1). Moreover, the system demonstrated capacity for personalized interaction by adapting to each child's individual needs, interests, and engagement levels. It was able to incorporate children's unique ideas, tailor its prompts and responses accordingly, provide elaborations, and integrate their creative contributions into a cohesive instructional dialogue centered on target vocabulary. This capability is similar to what was observed in parents (e.g., P1, P4, P6, P7, P8, P9, P10), who demonstrated a nuanced understanding of their children's preferences and often incorporated familiar routines, personal stories, or shared family experiences into the learning process. This combination of structured pedagogy, responsive feedback, and child-centered personalization contributed to a professional and engaging instructional experience aligned with core principles of early literacy instruction. For illustrated example quotes, see Table 4.

Table 4
Example quotes of pedagogical advancement.

Advantage	Example Quotes
Instructional Consistency	**Vovo:** *"Let's explore the word analyze together. Analyze means to look at something very carefully to understand it better. For example, if you have a puzzle, you might analyze each piece to see where it fits. Do you understand the word analyze? Can you try to explain it in your own words?"* **C7:** *"Okay. I analyze the puzzle, and like so like I like after I think that I finished every piece, I analyze it to make sure that I got every piece in the right spot."* **Vovo:** *"That's fantastic. You used the word analyze perfectly. You looked carefully to make sure every piece is in the right spot. Well done."*
Effective Scaffolding	**Vovo:** *"Let's create a story together. How about we start with the setting of a birthday party. We can say Jamie was so excited that she could anticipate all her friends arriving with gifts and laughter. What happens next in your story?"* **C4:** *"Jamie opens her gifts."* **Vovo:** *"Great job. The story could go something like this; Jamie was so excited that she could anticipate all her friends arriving with gifts and laughter. As the party went on, Jamie opened her presents and her friend Tammy … "*
Personalized Learning	**Vovo:** *"Now, would you like to make up a short story using the word as- sert?"* *"***C7:** *"Like if a Komodo Dragon asserts their feelings."* **Vovo:** *"That's a creative start. Let's continue. A Komodo Dragon asserts its feelings by standing tall and showing its strength to other animals. It confidently lets everyone know who is in charge. Great job using the word assert in your story."*

Advantage in Language Appropriateness. Vovo exhibited a strong advantage in language appropriateness across three dimensions. First, it consistently employed **age-appropriate language**, using clear and developmentally suitable vocabulary and sentence structures tailored to young children aged 3–7. The system avoided overly complex phrasing, instead using simple terms and familiar context to explain the target words and facilitate comprehension. In contrast to the gen-AI system, many parents faced difficulties in accurately defining academic vocabulary and translating it into child-friendly language. 4 out of 10 children encountered situations in which their parents were not familiar with the target word and had to check a dictionary or seek external help. For example, *"I don't know what ameliorate means.What does it mean?.. Um, I just looked it up and it says to make something back or unsatisfactory better. Okay … I need you to repeat that definition in your own words."* (P5). Second, the AI demonstrated a capacity for **creative and lexically diverse language** use. The AI system demonstrated notable strengths in generating creative and imaginative language, particularly when children struggled to use target vocabulary in context. While parents varied in their ability to model contextual usage, Vovo consistently provided rapid, engaging, and age-appropriate prompts. These creative outputs make the storytelling and sentence generation go smoothly while also promoting children's expressive language and narrative skills development. Finally, the system offered immediate and accurate **syntactic scaffolding**. When children produced grammatically incorrect responses, Vovo gently modeled the correct usage while offering constructive suggestions and positive reinforcement, supporting children's syntactic development. In contrast, parent-led sessions revealed limited evidence of grammar correction. Illustrated example quotes are listed in Table 5.

Advantage in Verbal Social-emotional Support**.** Vovo demonstrated the strength in verbal social-emotional support by providing (1) positive feedback and (2) conversational regulation. First, Vovo consistently provided encouraging **positive feedback**, such as praise (e.g., Vovo to C3: *"That's an exciting addition to your story."*), encouragement (e.g., Vovo to C1: *"I'd love to hear more about it."*), affirmations of effort (e.g., Vovo to C4: *"You're getting better each time!"*), and comfort children when they were frustrated (e.g., Vovo to C2: *"That's okay, lets try again."*). Similarly, parents also be generous and emphasize providing positive feedback, but AI and parents demonstrated different patterns in how they provided positive feedback. Statistical analyses revealed that the AI system provided more frequent, elaborated, and lexically rich praise, whereas parental feedback was generally shorter and simpler, though often paired with nonverbal cues such as clapping, smiling, or gestures (e.g., P1: *"That was your first sentence." [Clapping]*). Second, Vovo showed promise in **conversational regulation**, particularly in sustaining instructional flow. It was able to redirect off-task

Table 5
Example quotes of language appropriateness.

Advantage	Example Quotes
Age Appropriateness	**Vovo:** *"Let's explore the word analyze together. Analyze means to look at something very carefully to understand it better. For example, if you have a puzzle, you might analyze each piece to see where it fits. Do you understand the word analyze? Can you try to explain it in your own words?"* **C7:** *"Okay. I analyze the puzzle, and like so like I like after I think that I finished every piece, I analyze it to make sure that I got every piece in the right spot."* **Vovo:** *"That's fantastic. You used the word analyze perfectly. You looked carefully to make sure every piece is in the right spot. Well done."*
Lexical Diversity	**Vovo:** *"Let's make a story."* **C1:** *"What's something you like to create?"* **Vovo:** *"That's a great question. If I could create something I'd love to create a magical garden where all sorts of flowers and animals can talk and help each other. What about you? Is there something special you'd like to create?"* **C1:** *"Yes."*
Syntactic Scaffolding	**C5:** *"I run out for my benefit because you can exercise and keep healthy."* **Vovo:** *"That's great effort. Here's a little tweak. I went outside for my benefit because exercising keeps me healthy. Can you try saying that?"* **C5:** *"I went outside for my benefit, because, as I said, exercising keeps me healthy."* **Vovo:** *"Good try. Let's refine it a bit. I went outside for my benefit, because, as I said, exercising keeps me healthy. Can you try saying it like that?"* **C5:** *"I went outside for my benefit, because, exercise is keeping me healthy."* **Vovo:** *"That's a great sentence. You said, I went outside for my benefit, because, I said exercise is keeping my knee healthy. Well done."*

dialogue and clarify misunderstandings when children became distracted, confused, or introduced unrelated topics. This capability parallels parents' efforts to maintain engagement, typically through social negotiation, repetition, or incentives. Example quotes can be found in Table 6.

Table 6
Example quotes of verbal social-emotional support.

Advantage	Example Quotes
Positive Feedback	**Vovo:** *"Awesome. Can you say the word analyze and tell me what it means in your own words?"* **C5:** *"Like I analyze really like special stuff to make sure I get everything good."* **Vovo:** *"That's a great sentence. You're using analyze to describe looking at special stuff, to make sure everything is good."*
Conversational Regulation	**Vovo:** *"Awesome. Can you say the word emphasize and tell me what it means in your own words?"* **C7:** *"Emphasize means you can't get water up J's (C7's brother) nose."* **Vovo:** *"It sounds like you're thinking about swimming exercises, but let's get back to emphasize. It means to give special importance to something. Can you try explaining it in your own words?"* **C7:** *"It's like I can't fight J."* **Vovo:** *"Okay. Let's try using emphasize in a sentence. You could say the teacher will emphasize the importance of doing homework. Do you want to try making a sentence with emphasize?"* **C7:** *"Yes."* **Vovo:** *"Great. Go ahead and give it a try. What's your sentence with the word emphasize?"* **C7:** *"My teacher emphasizes that you can't go on the playground when it's wet."* **Vovo:** *"That's a fantastic sentence. Well done!"*

4.2.2. Challenges

Challenge in Speech Recognition. Vovo faced speech recognition limitations, which can be divided into three main areas: (1) mishearing inputs, (2) misunderstanding speech, and (3) misinterpreting the pause. First, Vovo sometimes **mishears** background noises and non-relevant sounds from the environment, often misinterpreting them as user input. This led to interruptions in the interaction flow and off-topic responses. In contrast, human caregivers handle auditory input with greater nuance. They actively manage the environment by reducing background noise, asking others to remain quiet, or moving the child to a quieter space. Second, it sometimes **misunderstands** young children's nonstandard pronunciations or phonological variations (a common phenomenon of phoneme deletion, repetition, or substitution in early childhood, e.g., saying "tat" instead of "cat" (Levey & Polirstok, 2011)), resulting in misunderstandings that could cause frustration. Lastly, Vovo had difficulty **interpreting pauses** accurately. Children exhibited pauses frequently due to thinking, confusion, reflection, distraction, or fatigue. Vovo was not always able to distinguish between a thoughtful silence and disengagement, sometimes ending sessions prematurely or providing prompts too quickly. Human caregivers, in contrast, demonstrated greater sensitivity to the nuances of silence. They used wait time effectively, interpreted children's facial expressions or body language, and scaffolded responses through prompts or encouragement. Example quotes are illustrated in Table 7.

Challenge in Instructional Persistence. A notable limitation of Vovo was its tendency to terminate sessions easily when children expressed disinterest; the system often responded with affirmation and promptly disengaged. While this reflects user-centered AI design principles prioritizing human control, it presents challenges for sustaining learning, especially with young children who lack consistent self-regulation and learning routines. Participants such as C4 (4 years 10 months), C8 (3 years 7 months), and C10 (4 years 3 months) exhibited the lowest session completion rates with the AI. In contrast, human caregivers demonstrated greater instructional persistence through emotional encouragement (e.g., P1: *"You can do it. Try it."*), incentives (e.g., P10: *"Finish the sentence and you'll get a cookie."*), and firm restatement of expectations (e.g., P1:*"I need you to sit up nice and straight. Can you say illustrate?"*). Example quotes can be found in Table 8.

Challenge in Nonverbal Social-Emotional Support. As a conversational AI system, another limitation of Vovo is its lack of nonverbal social cues. Vovo relies solely on auditory input and spoken language for

Table 7
Example quotes of speech recognition.

Challenge	Example Quotes
Mishearing	**Vovo:** *"Great job. The rain can cause the grass to grow. You're doing fantas- tic. Would you like to make your own sentence with the word cause?"* **C2:** *"Hold on."* (Participant talked to his sister. This causes Vovo to buffer.) **Vovo:** *"All right. If you have any questions or want to learn another word in the future. You're doing great."*
Misunderstanding	**Vovo:** *"Great job. The story could go something like this; Jamie was so ex- cited that she could anticipate all her friends arriving with gifts and laughter. As the party went on, Jamie opened her presents and her friend Tammy … "* **C4:** *"I need a birthday party."* (The system misinterpreted it as "I need a bath today," according to the transcript.) **Vovo:** *"That's perfectly fine. Let's wrap up for today. Can you remind me what anticipate means?"* (After C4's response, Vovo ended the session.)
Misinterpreting Pause	**Vovo:** *"Now, how about creating a short story using the word ultimate? You can use any characters or setting you like."* **C7:** *"So once there's this ultimate Komodo/King Cobra … "* (Child pauses, thinking.) **Vovo:** *"That's a great twist. Let's continue. One day a King Cobra appeared on the island, challenging Slash for the title of Ultimate Protector. The island animals watched as Slash and the King Cobra faced off."*

Table 8
Example quotes of instructional persistence.

Challenge	Example Quotes
Instructional Persistence	**Vovo:** *"That's a fun way to remember it. Event is something that happened or a special occasion. Would you like to try making a sentence with event?* *"***C2:** *"Not today, maybe next year."***Vovo:** *"Sure thing. If you want to continue another time or have any ques- tions, feel free to reach out. Have a great day."*

communication, without access to other sensory or expressive modalities. As a result, it is facing challenges in sustaining children's attention, regulating behavior, and building emotional connections during instruction. In contrast, parents frequently employed rich nonverbal behaviors while interacting with their children, such as eye contact, smiles, affirming gestures (e.g., thumbs-up, clapping, high-fives), and attention-directing cues (e.g., snapping fingers, gentle shoulder taps). This absence of embodied social presence in Vovo may reduce children's engagement, especially for those who rely on such cues to remain focused and motivated during learning.

Challenge in Phoneme Instruction. Vovo's functionality is limited to phoneme instruction. As a system built on a text-to-speech generative model (ChatGPT-4o), it lacks the ability to isolate and explicitly teach phonemes, syllables, or blending skills that are foundational for early literacy, particularly for children struggling with the decoding or pronunciation of multi-syllabic words. The system cannot identify difficult syllables or break words down into single phonemes, which often results in children continuing to struggle with unfamiliar sounds without receiving effective instruction. In contrast, human caregivers demonstrated significant advantages; they slowed their speech, modeled articulation using visible mouth and tongue movements (e.g., P8: *"No, look at me, look at me. Still."*), and supported phonological segmentation and blending through repetition (e.g., P3: *"Current-ly, currently-ly, currently-ly."*). By comparison, Vovo only delivers whole-word pronunciations without phonological scaffolding, limiting its effectiveness for children developing foundational literacy skills.

In summary, quantitative analyses revealed that Vovo delivered structured literacy instruction consistently and comparably to human caregivers. However, children achieved higher learning outcomes during parent-led sessions. Qualitative findings supported these results by identifying Vovo's strengths in pedagogical consistency, age-appropriate language, and positive verbal socioemotional support, which contributed to the effectiveness of its instructional delivery. At the same time, limitations in speech recognition, instructional persistence, nonverbal social cues, and phoneme instruction may have contributed to the gap in learning outcomes. Together, the findings suggest that while Vovo holds promise as a consistent instructional partner for formal home literacy learning, the involvement of human caregivers remains irreplaceable and contributes significantly to children's learning outcomes.

4.3. RQ3: How do children and parents perceive conversational AI as a home literacy learning tool?

To better understand how families perceive Vovo as a home literacy learning tool, we analyzed interviews with a focus on parents' and children's attitudes toward the AI system based on their experience and their expectations of the future design.

4.3.1. Parents perceptions

Parents perceived Vovo as intelligent, professional, useful, and trustworthy. Parents generally held positive perceptions of Vovo, frequently describing it as smart, professional, and helpful. For example, P5 noted, "*She (Vovo) explains all the words mean. It's smarter than us*," while P1 mentioned, "*Vovo was able to immediately recite the definition. I feel like that's very smart.*" P4 mentioned, "*I like how it revisits the word again and again.*" P6: "*(Vovo) knew the definitions. Knew how to explain it in an easy understandable way.*" Most parents (P1, P4, P5, P6, P7, P10) viewed the tool as useful for language development and expressed interest in continue using it (e.g., P10: *"It's telling him the meaning of every word … Getting exposed in the language is helpful."* P4: *"I would definitely introduce this in my classroom, I see it as promising."*). P7 highlighted its potential to support independent learning at home, stating, *"It would be a way for him to learn without taking up an adult's time."* It highlighted the potential substitute role that AI can play in home literacy learning. Many parents (P1, P2, P3, P5, P6, P7, P9, P10) also observed that their children were indeed learning new words and enjoying the experience. As P1 described, *"I feel excited whenever I know that he's gonna be talking to Vovo … just because I love hearing the words and seeing him learn the words, he uses them when he's even around the house, he'll use the word, so I do feel like he enjoys."* Importantly, all parents reported no privacy or content concerns with Vovo, describing it as safe and trustworthy. For instance, P10: *"I personally don't (have any privacy concerns.) I've never had an issue with (Vovo) like us being online with him and anything."* P1: *"I don't feel like from the experience that I've had so far with Vovo. I don't feel like Vovo would teach my child any words that he shouldn't know or teach him anything that he shouldn't be doing."*

Parents expected enhancement in speech recognition and multimodal interface design. We found that parents expressed expectations for AI tools in two main areas: technical expectations, which addressed issues such as reliability on speech recognition and functionality, and interface design expectations, which emphasized the need for engaging, multimodal, and visually supportive features. Parents highlighted difficulties when the AI system failed to understand children's pronunciation, particularly for those with emerging language skills or speech issues. Several parents (P2, P4, P7, P8, P10) noted moments when miscommunication led to child frustration, which in turn caused parental frustration. For instance, P4: *"People with speech issues, certain sounds you can't pronounce. I think it would be hard (to communicate with Vovo)."*, and P7: *"He doesn't necessarily speak the clearest, but I could hear what he was saying (correctly). And it wasn't picked up the right word (by Vovo). So it was taking the conversation in another direction … It understands you, but it doesn't understand you conversationally."* One parent explained, *"For people with speech issues, I think it would be hard who can't pronounce certain (sounds), it would be hard"* (P4). These accounts suggest that accurate recognition of young children's varied voices remains a critical design need. Another area of expectation concerned noise handling. Parents also noticed that the tool was overly sensitive, sometimes picking up irrelevant sounds: *"Like you breathed too heavy it stopped. So, they're very sensitive"* (P6). Together, these insights underscore the importance of designing robust AI that can filter distractions and sustain attention on the child.

Another expectation from parents is the improvements in interface design, particularly the inclusion of visual and multimodal supports. Parents stressed that children found it difficult to engage with disembodied voice interaction alone. One parent reflected, *"There's nothing to look at when somebody's talking to her … maybe if there was like a digital person … it might be a better interactive"* (P6). Another emphasized the role of visuals in comprehension: *"Seeing words is obviously hard … some of the words, I think would be more helpful as like, here's a picture of this"* (P10). Suggestions extended beyond simple images to playful and age-appropriate characters, as one parent proposed, *"even if it's a dinosaur even or a little monster … if they could just see something talking back to them. I feel like that would help"* (P1). Other recommendations included game-like elements and interactive options, as described by a parent who imagined, *"set up more like not quite a game, but with characters and interactive … making it more exciting"* (P9). Collectively, these expectations underline that AI tools for young learners should not only focus on auditory but also engage children through rich, multimodal, and playful interfaces.

In summary, parents perceive Vovo positively for its intelligence, professionalism, usefulness, and trustworthiness, with expectations on

improved speech recognition and multimodal engaging design.

4.3.2. Children's perceptions

Children perceived Vovo as fun but prefer to learn with parents. In the interviews, children described their learning experiences with AI as fun, smooth, and effective, and characterized Vovo as kind and engaging. For example, C9 (7 years, 5 months) stated, *"She's fun, and she does try to talk and play with me,"* and C5 (6 years) noted, *"Because it's fun."* Several children (C2, C3, C5, C7, C9) also emphasized that they learned with Vovo, with C7 (7 years, 7 months) explaining, *"She teaches me the word properly,"* and C5 remarking, *"I learn a whole bunch of new words."* All 10 children mentioned that Vovo spoke to them in a smooth way and described Vovo as kind. However, some children also highlighted issues with speech recognition. For instance, one child (C5) observed, *"When there's a noise louder than me, she's mostly focused on that. When I'm the loudest noise, she focuses on me,"* suggesting that background noise could interfere with Vovo's responsiveness.

When asked whether Vovo felt more like a human or a computer, all 10 children identified Vovo as a computer, offering explanations such as *"Because it's in the computer"* (C1, 5 years, 5 months). This suggests that while children recognized Vovo's intelligence, they did not attribute human-like qualities to Vovo. Regarding learning preferences, 9 out of 10 children said they enjoyed learning with their parents more. Their reasoning centered on emotional connection, with children explicitly expressing sentiments such as, "*Because she (Vovo) is not my mommy*" (C10, 4 years, 3 months) and *"Because her is not mommy, I love my mom"*(C1, 5 years, 5 months). These findings indicate that while children regarded Vovo as a fun and capable learning companion, they did not perceive her as a substitute for parental involvement.

In summary, parents perceived Vovo as intelligent, useful, and trustworthy, and expressed expectations for system improvement in speech recognition, noise handling, and the inclusion of multimodal, visually supportive features to sustain engagement. Children found it fun and engaging but still preferred learning with parents, emphasizing the importance of social emotional connection. Fig. 4 provides an integrated summary of findings from quantitative data, video observations, and parent-child interviews, highlighting the system's strengths, limitations, and how families perceived its educational value.

5. Discussion

Drawing on home literacy environment (HLE) theory and structured literacy pedagogy, this study demonstrates how AI tools can serve as formal home literacy learning partner by actively engaging children in structured instructional HLE activities. It extends the application of structured literacy pedagogy beyond school settings into informal home learning contexts. By examining families' perceptions of Vovo as a home literacy learning partner, this study advances understanding of families' perspectives and expectations on conversational AI design and implementation, contributing insights to the development of socially responsive and pedagogically grounded AI systems for home literacy learning.

5.1. Effectiveness of conversational AI in home literacy (RQ1)

This study makes a theoretical contribution by applying structured literacy pedagogy to a home learning environment through a conversational AI system. Structured literacy has long been recognized as an effective instructional approach in formal education settings, but its implementation in home environments has been limited due to the professional expertise required of instructors. By embedding this pedagogy into a conversational AI system, the study demonstrates how structured literacy can be delivered consistently at home, addressing a critical gap in home-based literacy support. It also takes a meaningful step toward promoting educational equity by introducing a promising support for children who may have limited formal HLE experiences due to socioeconomic or contextual constraints. Quantitative findings on instructional effectiveness suggest that conversational AI systems like Vovo can successfully deliver pre-scripted structured literacy pedagogy in a consistent manner, supporting the potential of conversational AI as a learning partner for formal home literacy instruction.

However, results also revealed discrepancies in learning outcomes, pointed out its limitation as an AI system. First, as a computational system, Vovo operated based on pre-programmed instructions and lacked the social-emotional bonds with children that naturally arise in child–parent interactions (Formosa et al., 2025). According to sociocultural theory (Vygotsky, 1978), children learn more effectively when interacting with emotionally responsive human partners, as these relationships promote stronger engagement, motivation, and retention of learning. Such social and emotional responsiveness remains difficult for computational agents to fully replicate (Wu, 2024). Second, the AI system's limitations in phoneme-level instruction likely impacted children's learning outcomes. Unlike parents who could visually model mouth movements, isolate and repeat problematic syllables, and tailor feedback, Vovo was unable to provide phoneme-level scaffolding. This scaffolding is particularly critical for children who are still developing

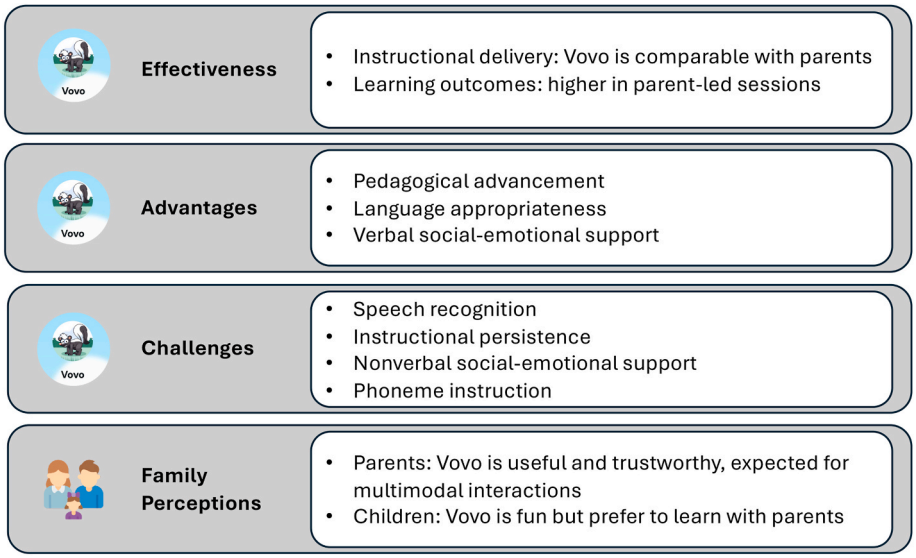


Fig. 4. Summary of findings.

phonemic awareness (Levey & Polirstok, 2011). These findings highlight the importance of designing AI systems with greater social-emotional responsiveness and pedagogical affordances that can better support children's literacy development at home.

5.2. Advantages and challenges of conversational AI in home literacy (RQ2)

This study also extends the theoretical scope of home literacy environment (HLE) frameworks. The design of Vovo extends the traditional HLE activities from parent-child interactions, shared reading activities, and reading resources exposure into an AI-enhanced literacy learning experience. By analyzing Vovo's advantages and challenges, the findings share insights into how conversational AI systems perform as a formal home literacy learning partner compared with human caregivers and offer insights on how to better leverage the strengths of AI to support early literacy learning at home.

Pedagogically, Vovo demonstrates a strong capacity to deliver structured literacy instruction with high fidelity, providing professional-level scaffolding and maintaining a consistent instructional sequence aligned with structured literacy pedagogy. This reflects a key advantage of AI as a formal home literacy learning partner due to its ability to reliably execute pre-designed, pedagogically informed prompts without instructional drift (Schmucker et al., 2024, pp. 75–90; Siegle et al., 2023; Van Steensel, 2006). It overcomes disparities in caregivers' capacity to provide formal home literacy activities, offering children equitable access to pedagogically grounded HLE supports. However, preschoolers and early elementaryers are still developing self-regulation, attention control, and task persistence; as a result, effective instruction at this developmental stage often requires external regulation and support to sustain engagement (Amershi et al., 2019, pp. 1–13; Blair & Raver, 2015; Puckett et al., 2009; Shneiderman, 2020). It highlights the need for future designs to incorporate developmentally appropriate engagement strategies and move beyond child-AI dyadic interactions toward a parent-child-AI triadic learning model, involving parents as a moderator to support learning under the AI learning environments.

Linguistically, Vovo demonstrates the capacity to provide developmentally appropriate language input by modeling vocabulary explanations, syntactic structures, and repeatedly exposing target words within meaningful contexts. This reflects a key educational advantage of conversational AI systems, which is their ability to deliver stable, high-quality linguistic input that aligns with children's language development needs (Beck et al., 2013; Hirsh-Pasek et al., 2015; Justice et al., 2008; Rowe, 2012). From an HLE perspective, such consistent, enriched, and intentional language input represents a core feature of an ideal HLE that supports children's language growth (Sénéchal & LeFevre, 2002). However, limitations in Vovo's speech recognition frequently result in mishearing or misinterpreting children's input, disrupting conversational contingency and potentially constraining opportunities for responsive language learning, highlighting implications for AI language learning systems, particularly the need to strengthen child-directed speech recognition and conversational repair mechanisms.

Social-emotionally, Vovo demonstrated strengths in verbal engagement, including providing positive feedback more frequently and with greater lexical richness than parents, as well as redirecting off-task conversation back to the learning tasks. Such consistent encouragement and predictable responses help support motivation and task persistence in early literacy learning (Hattie & Timperley, 2007; Justice et al., 2018; Yaseen et al., 2025). However, due to the limitations of LLM technologies adopted in this study which primarily support verbal interactions, Vovo's social responsiveness remains constrained by its lack of embodiment and visual input/output. Language acquisition in early childhood involves using language in a social context through pragmatic cues such as facial expressions, gestures, eye contact, turn-taking, and joint attention (Levey & Polirstok, 2011; Puckett et al., 2009). The development of pragmatic and social-emotional skills is critical for early literacy development. This constraint represents a critical gap in Vovo's design and suggests integration of nonverbal, socially and emotionally responsive features in future AI literacy systems for young children.

5.3. Improve conversational AI to meet family's need (RQ3)

This study contributes to emerging discussions on socially responsive AI design by exploring families' perceptions and needs related to children's use of conversational AI in home literacy learning. Through examining parents' and children's use experiences, we found that families expect AI systems that were multimodal, consistent, and socially responsive. Children expressed a preference for learning with caregivers, revealing a future direction for designing socially responsive AI and exploring parent-child-AI triadic learning dynamics that better support emotional engagement and learning.

Useful function identified by parents and children. Across interviews, families repeatedly described Vovo as "smart," "knowledgeable," and "helpful." Such perceptions of the system's intelligence contributed to parents' trust in its educational value. These findings align with prior research indicating that families perceive conversational AI tools as trustworthy educational resources for young children (Quan et al., 2024). This trust in the system's pedagogical credibility reinforces its role as a supportive tool for home literacy learning. Parents also frequently emphasized Vovo's capacity to provide immediate definitions and clear explanations of unfamiliar words. This function not only reduced parents' need to intervene but also created opportunities for children to expand their vocabulary independently. Several parents observed that their children were able to transfer these newly acquired words into daily contexts, using them spontaneously around the house. Parents viewed such transfer as evidence of meaningful literacy learning. Parents highlighted the potential for Vovo to serve as a tool for children's self-directed literacy practice. They valued the possibility that children could engage productively with literacy tasks without requiring constant adult supervision, thereby freeing parents' time while maintaining ongoing educational support. This function was seen as particularly promising for busy households seeking supplemental literacy opportunities.

Improve the speech recognition and multimodal design in AI Literacy Tools. Our findings suggest that technical and design limitations shaped how families experienced Vovo as a home literacy tool. In the study, we found that parents consistently noted that inaccurate speech recognition disrupted the flow of interaction, particularly in noisy home environments or when recognizing young children's voices. This highlights a critical need for more robust speech recognition that is sensitive to developmental and contextual factors, ensuring smoother and more reliable child-AI exchanges. In addition, both parents and children emphasized the importance of multimodal features for sustaining engagement and supporting comprehension. Children expressed a preference for visible, interactive characters, while parents highlighted the need for visual support to enhance children's knowledge retention and engagement. Together, these insights point to the potential of embodied, visually engaging, and multimodal designs to not only maintain young learners' motivation but also to reinforce literacy learning through richer, multi-sensory experiences.

Emphasis on social-emotional bond in the learning process. Our interviews revealed that while children often described Vovo as "fun" or "smart," they still expressed a preference for learning with their parents. This finding emphasizes the irreplaceable role of affective bonds in young children's learning (Puckett et al., 2009; Vygotsky, 1978). Given this, AI-enhanced home literacy tools like Vovo should be designed to complement, rather than replace, parental involvement. The goal should be to support shared learning experiences and extend the educational environment while maintaining opportunities for emotional closeness and co-regulation. Future AI design should consider features that encourage joint engagement, provide emotionally responsive feedback, and facilitate caregiver participation. Designing for emotional

connection as well as instructional content may lead to more meaningful and effective literacy support for young children at home.

These findings point to the family's needs and key design directions for conversational AI in home literacy. Strengthening speech recognition and noise handling can reduce frustration. Integrating multimodal and age-appropriate supports can sustain engagement. Positioning the tool to complement rather than replace parent-child interactions ensures that affective bonds remain central to early literacy learning. By aligning technical reliability with developmental needs and relational contexts, conversational AI can more effectively scaffold young children's literacy learning at home.

6. Implications, limitations, and future directions

This study contributes to ongoing conversations on the role of conversational AI in supporting home literacy learning. It represents a novel educational innovation by integrating structured literacy pedagogy into a home-based, AI-enhanced instructional experience. By embedding structured literacy instruction within a conversational AI system, the study demonstrates how young children interact with AI for formal home literacy learning and how such systems can function as pedagogically grounded and consistent literacy partners at home. The study also captures families' lived experiences, highlighting their expectations for multimodal and socially responsive AI systems design that meets children's developmental needs and home learning contexts.

While the present study offers valuable contributions to home literacy learning, there remain limitations for improvement and further exploration in future research. In terms of **study design**, the within-subjects design involving a small group of families enabled in-depth comparisons between AI-led and parent-led instructional conditions while controlling for cross-subject variability. However, the study's limited sample restricts the generalizability of the findings. Although the study included families from diverse socioeconomic backgrounds, the small sample size limits its generalizability and may not capture the full range of literacy challenges experienced by families. Additionally, the within-subject design was not fully counterbalanced, which may have led to order effects or learning transfer between conditions. Future research should aim to expand the participant pool to include families from a broader range of demographic backgrounds. This would enable a more comprehensive understanding of how factors such as age, socioeconomic status, linguistic background, and prior technology exposure influence children's learning with conversational AI. Moreover, while qualitative analyses followed established procedures of inductive thematic analysis, the small sample size and interpretive nature of qualitative data constrained broader generalizability. Future research should expand both the participant pool and the coding team to enable greater inter-rater agreement and ensure the robustness and replicability of emergent themes across a more diverse range of child-AI interaction patterns.

In terms of **AI system design**, findings from this study suggest that future conversational AI systems for early literacy should move beyond voice-only interaction toward multimodal sensing and feedback. Parents and children consistently highlighted limitations related to speech recognition, attention regulation, and the absence of nonverbal social cues from our findings. Future systems could integrate multimodal sensing capabilities such as gesture recognition, gaze tracking, and facial expression analysis to better infer children's engagement, confusion, or hesitation. For example, detecting gaze aversion or prolonged silence could allow the system to distinguish between thoughtful pauses and disengagement, enabling more developmentally appropriate wait time and adaptive prompting (Abeles & Yuval-Greenberg, 2017; Doherty-Sneddon, Bruce, Bonner, Longbotham, & Doyle, 2002). Similarly, gesture-based input (e.g., pointing, nodding) could support joint attention and reduce reliance on verbal responses alone, which is particularly important for younger children or those with emerging language skills (Nyatsanga et al., 2023, May). Addtionally, children's preference for learning with parents highlights the need to design AI systems that complement rather than replace caregiver involvement. A promising direction is the development of triadic interaction models in which the AI, child, and parent collaboratively participate in literacy activities.

Also, while Vovo demonstrated strong consistency in instructional delivery, it lacked mechanisms for dynamically adjusting instructional persistence and affective support. It also lacked personalization features such as memory across sessions or adaptive difficulty levels based on prior performance. This limited its capacity to deliver differentiated instruction tailored to individual learning trajectories. Future research should explore adaptive scaffolding strategies that adjust instructional pacing, repetition, and difficulty based on children's real-time performance and affective states, incorporating affect-aware feedback, such as detecting frustration, boredom, or excitement through vocal prosody or interaction patterns (Grawemeyer et al., 2017). Future work should also incorporate memory-augmented architecture and real-time learner modeling to support sustained, personalized interaction over time and enable rigorous evaluation of individual learning outcomes.

In addition, prior research suggests that device form factor can shape children's attention, posture, and interaction style during learning activities (Dubé & McEwen, 2017; Crescenzi, Jewitt, & Price, 2014). Differences in perception and engagement when using smartphones versus larger-screen devices represent an important future research direction. Future studies should explicitly compare children's learning experiences and social perceptions of conversational AI across device types, including laptops, tablets, and mobile phones, particularly given the prevalence of smartphone-based learning in home environments.

7. Conclusion

This study investigated the role of conversational AI in supporting formal literacy learning by developing Vovo, a structured literacy grounded AI system, in a six-week home intervention with 10 families and their children aged 3–7. Across 150 learning sessions, Vovo demonstrated consistent instructional delivery comparable to that of parents. While children showed greater learning outcomes in parent-led sessions and expressed preference for learning with parents, they also found Vovo to be kind, fun, and helpful. Video analyses highlighted Vovo's strengths in pedagogical consistency, language appropriateness, and verbal social-emotional support, alongside limitations in speech recognition, nonverbal social cues, instructional persistence, and phoneme instruction. Parents perceived Vovo as useful and trustworthy but also called for improvement in speech recognition and multimodal design. The study broadens the implications of structured literacy pedagogy into home settings and expands the theoretical scope of the home literacy environments by integrating AI-enhanced learning experiences. These findings highlight that conversational AI systems like Vovo can function as consistent and pedagogically grounded instructional partners, while emphasizing the irreplaceable role of human caregivers. The study underscores the need for future AI design to enhance pedagogical affordances and social responsiveness, while also emphasizing the importance of developing triadic learning models that involve parents, children, and AI collaboratively.

CRediT authorship contribution statement

Shuang Quan: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Xintian Tu-Shea:** Data curation, Writing – original draft, Writing – review & editing. **Yi Ding:** Conceptualization, Methodology, Writing – review & editing. **Yao Du:** Writing – original draft, Writing – review & editing. **Qingxiao Zheng:** Software, Writing – review & editing. **Laney E. Gerdich:** Data curation, Formal analysis, Project administration.

Statement on open data and ethics

The datasets used in this study are not publicly available. The authors do not have the authority to share the report drafts, interview transcripts, or recordings; however, other materials are available from the corresponding author upon reasonable request. This study was approved by the Institutional Review Board (IRB) at the first author's institution (ID: 24–2025). Informed consent was obtained from all participants, and participants' privacy and confidentiality were strictly protected.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This study was supported by Juniata College. Special thanks are extended to the student members of the AI in Education Lab (AEL) at Juniata College, supervised by Dr. Shuang Quan, for their contributions to this project. In particular, to undergraduate research assistants Laney Gerdich and Tatum Livelsberger for their dedicated assistance throughout the study, Jayden Strausbaugh for his assistance with system development, and Maddy Detz, Courtney Clippinger, Hannah Johnston, Ava Gummer, Emma Childers, Sarah Bradley, Aaliyah McGee, and Alisyn Wildner for their assistance with participant recruitment and data collection. We are also deeply grateful to the children and parents who generously invested their time and support in this research.

References

Abeles, D., & Yuval-Greenberg, S. (2017). Just look away: Gaze aversions as an overt attentional disengagement mechanism. *Cognition, 168*, 99–109.

Alam, A. (2023). Harnessing the power of ai to create intelligent tutoring systems for enhanced classroom experience and improved learning outcomes. In *Intelligent communication technologies and virtual mobile networks* (pp. 571–591). Springer.

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., et al. (2019). Guidelines for human-ai interaction. *Proceedings of the 2019 chi conference on human factors in computing systems*.

Araújo, L., & Costa, P. (2015). Home book reading and reading achievement in eu countries: The progress in international reading literacy study 2011 (pirls). *Educational Research and Evaluation, 21*(5–6), 422–438.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, like ability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics, 1*(1), 71–81.

Beck, I. L., McKeown, M. G., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.

Biemiller, A. (2013). Vocabulary development and. *Handbook Early Liter. Res., 2*(2), 41.

Blair, C., & Raver, C. C. (2015). School readiness and self-regulation: A developmental psychobiological approach. *Annual Review of Psychology, 66*(1), 711–731.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.

Burgess, S. R., Hecht, S. A., & Lonigan, C. J. (2002). Relations of the home literacy environment (hle) to the development of reading-related abilities: A one-year longitudinal study. *Reading Research Quarterly, 37*(4), 408–426.

Bus, A. G., Van Ijzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research, 65*(1), 1–21.

Cha, W. K., & Daud, P. (2025). Enhancing early education with artificial intelligence: A comparative study of ai-powered learning versus traditional methods. *International Journal of Academic Research in Business and Social Sciences*. https://doi.org/10.6007/ijarbss/v15-i2/24690

Chen, G. (2024). The impacts of ai-driven storytelling applications on language acquisition and literacy development in early childhood education: A systematic review. *Asian Journal of Advanced Research and Reports*. https://doi.org/10.9734/ajarr/2024/v18i11770

Chen, X., Zhou, H., Zhao, J., & Davey, G. (2010). Home literacy experiences and literacy acquisition among children in guangzhou, south China. *Psychological Reports, 107*(2), 354–366.

Chiu, T. K. F., Moorhouse, B. L., Chai, C., & Ismailov, M. (2023). Teacher support and student motivation to learn with artificial intelligence (ai) based chatbot. *Interactive Learning Environments, 32*, 3240–3256. https://doi.org/10.1080/10494820.2023.2172044

Coyne, M. D., McCoach, D. B., & Kapp, S. (2007). Vocabulary intervention for kindergarten students: Comparing extended instruction to embedded instruction and incidental exposure. *Learning Disability Quarterly, 30*(2), 74–88.

Crescenzi, L., Jewitt, C., & Price, S. (2014). The role of touch in preschool children's learning using iPad versus paper interaction. *Australian Journal of Language and Literacy, 37*(2), 86–95.

Creswell, J. W., & Miller, D. L. (2000). Determining validity in qualitative inquiry. *Theory Into Practice, 39*(3), 124–130.

De Jong, P. F., & Leseman, P. P. M. (2001). Lasting effects of home literacy on reading achievement in school. *Journal of School Psychology, 39*(5), 389–414.

Dickinson, D. K., & Tabors, P. O. (2001). *Beginning literacy with language: Young children learning at home and school*. Paul H. Brookes Publishing.

Dietz Smith, G., Prasad, S., Davidson, M. J., Findlater, L., & Shapiro, R. B. (2024). Contextq: Generated questions to support meaningful parent-child dialogue while co-reading. *Proceedings of the 23rd annual ACM interaction design and children conference*.

Doherty-Sneddon, G., Bruce, V., Bonner, L., Longbotham, S., & Doyle, C. (2002). Development of gaze aversion as disengagement from visual information. *Developmental Psychology, 38*(3), 438.

Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017). " hey google is it ok if i eat you?" initial explorations in child-agent interaction. *Proceedings of the 2017 conference on interaction design and children*.

Dubé, A. K., & McEwen, R. N. (2017). Abilities and affordances: Factors influencing successful child–tablet communication. *Educational Technology Research & Development, 65*(4), 889–908.

Dunn, D. M. (2019). *Peabody picture vocabulary test* (5th ed.). Bloomington, MN: NCS Pearson [Measurement instrument].

Durski, S., Müller, W., Rebholz, S., & Massler, U. (2020). Reading fluency training with Amazon alexa. *CSEDU*, (2), 139–146.

Evans, M. A., Shaw, D., & Bell, M. (2000). Home literacy activities and their influence on early literacy skills. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 54*(2), 65.

Formosa, P., Hipólito, I., & Montefiore, T. (2025). Artificial intelligence (Ai) and the relationship between agency, autonomy, and moral patiency. *arXiv preprint arXiv: 2504.08853*.

Grawemeyer, B., Mavrikis, M., Holmes, W., Gutiérrez-Santos, S., Wiedmann, M., & Rummel, N. (2017). Affective learning: Improving engagement and enhancing learning with affect-aware feedback. *User Modeling and User-Adapted Interaction, 27*(1), 119–158.

Hammer, C. S., Miccio, A. W., & Wagstaff, D. A. (2003). Home literacy experiences and their relationship to bilingual preschoolers' developing English literacy abilities. *Language, Speech, and Hearing Services in Schools, 34*(1), 20–30.

Han, A., & Cai, Z. (2023). Design implications of generative ai systems for visual storytelling for young learners. *Proceedings of the 22nd annual ACM interaction design and children conference*.

Han, A., Zhou, X., Cai, Z., Han, S., Ko, R., Corrigan, S., & Peppler, K. A. (2024). Teachers, parents, and students' perspectives on integrating generative ai into elementary literacy education. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–17.

Hart, B., Risley, T. R., & Kirby, J. R. (1997). Meaningful differences in the everyday experience of young American children. *Canadian Journal of Education, 22*(3), 323.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112. https://doi.org/10.3102/003465430298487

He, K., Levine, J., Cervera, K., Ojeda-Ramirez, S., Xu, Y., & Warschauer, M. (2024). A home study of parent-child co-reading with a bilingual conversational agent. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–8.

Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., & Kaufman, J. (2015). Putting education in "educational" apps: Lessons from the science of learning. *Psychological Science in the Public Interest, 16*(1), 3–34. https://doi.org/10.1177/1529100615569721

Ho, H.-R., Kargeti, N., Liu, Z., & Mutlu, B. (2025). Set-paired: Designing for parental involvement in learning with an ai-assisted educational robot. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–20.

Hood, M., Conlon, E., & Andrews, G. (2008). Preschool home literacy practices and children's literacy development: A longitudinal analysis. *Journal of Educational Psychology, 100*(2), 252.

Hsu, T.-C., Chang, C., & Jen, T.-H. (2023). Artificial intelligence image recognition using self-regulation learning strategies: Effects on vocabulary acquisition, learning anxiety, and learning behaviours of English language learners. *Interactive Learning Environments, 32*, 3060–3078. https://doi.org/10.1080/10494820.2023.2165508

Hubbard, L., Ding, S., Le, V., Kim, P., & Yeh, T. (2021). Voice design to support young children's agency in child-agent interaction. *Proceedings of the 3rd conference on conversational user interfaces*.

Justice, L. M., Jiang, H., & Strasser, K. (2018). Linguistic environment of preschool classrooms: What dimensions support children's language growth? *Early Childhood Research Quarterly, 42*, 79–92.

Justice, L. M., Mashburn, A., Pence, K. L., & Wiggins, A. (2008). Experimental evaluation of a preschool language curriculum: Influence on children's expressive language skills. *Journal of Speech, Language, and Hearing Research, 51*(4), 983–1001.

Karpava, S. (2021). Family bilingualism and home literacy environment in the context of Cyprus migrants. *ELIA, 2*(21), 93–138.

Kewalramani, S., Kidman, G., & Palaiologou, I. (2021). Using artificial intelligence (ai)-interfaced robotic toys in early childhood settings: A case for children's inquiry literacy. *European Early Childhood Education Research Journal, 29*, 652–668. https://doi.org/10.1080/1350293X.2021.1968458

Kilpatrick, D. A. (2015). *Essentials of assessing, preventing, and overcoming reading difficulties*. John Wiley & Sons.
Kucirkova, N., & Hiniker, A. (2024). Parents' ontological beliefs regarding the use of conversational agents at home: Resisting the neoliberal discourse. *Learning, Media and Technology, 49*(2), 290–305.
Lai, Z. C.-C. (2025). The impact of ai-assisted blended learning on writing efficacy and resilience. *International Journal of Computer-Assisted Language Learning and Teaching*. https://doi.org/10.4018/ijcallt.377174
Lee, J., Lee, K., Hwang, I., Park, S., & Kim, Y.-H. (2025). Autihero: Leveraging generative ai in social narratives to engage parents in story-driven behavioral guidance for autistic children. *arXiv preprint arXiv:2509.17608*.
Lee, J., Shin, D., & Noh, W.-j. (2023). Artificial intelligence-based content generator technology for young english-as-a-foreign-language learners' reading enjoyment. *RELC Journal, 54*, 508–516. https://doi.org/10.1177/00336882231165060
Leech, K. A., Lin, G., Pedonti, S., Schoenfeld, I., & Bilgin, C. U. (2025). Evaluating the usability of a conversational agent to enhance parent-child shared reading interactions. *International Journal of Child-Computer Interaction, 43*, Article 100720.
Leseman, P. P., & De Jong, P. F. (1998). Home literacy: Opportunity, instruction, cooperation and social-emotional quality predicting early reading achievement. *Reading Research Quarterly, 33*(3), 294–318.
Létourneau, A., Deslandes Martineau, M., Charland, P., Karran, J. A., Boasen, J., & Léger, P. M. (2025). A systematic review of ai-driven intelligent tutoring systems (its) in k-12 education. *npj Science of Learning, 10*(1), 29.
Levey, S., & Polirstok, S. (2011). *Language development: Understanding language diversity in the classroom*. Sage Publications.
Li, H., Huang, W., & Ai, J. (2024). Comparing computer-assisted and teacher-led instruction for teaching emergent literacy skills to Chinese preschoolers with autism. *Remedial and Special Education*. https://doi.org/10.1177/07419325241297279
Lin, G.-Y., Jhang, C.-C., & Wang, Y.-S. (2023). Factors affecting parental intention to use ai-based social robots for children's esl learning. *Education and Information Technologies, 29*, 6059–6086. https://doi.org/10.1007/s10639-023-12023-w
Liu, V., Latif, E., & Zhai, X. (2025). *Advancing education through tutoring systems: A systematic literature review*. arxiv. *arXiv preprint arXiv:2503.09748*.
Lyu, Y., Liu, D., An, P., Tong, X., Zhang, H., Katsuragawa, K., & Zhao, J. (2024). Emooly: Supporting autistic children in collaborative social-emotional learning with caregiver participation through interactive ai-infused and ar activities. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 8*(4), 1–36.
Maspul, K. A. (2025). Chatgpt as an educative and pedagogical tool: Perspectives and prospects in international schools in Indonesia. *Innovative Technologica: Methodical Research Journal*.
Moats, L. C. (2020). *Speech to print: Language essentials for teachers (3rd)*. Brookes Publishing Company.
Mostow, J., Aist, G., Burkhead, P., Corbett, A. T., Cuneo, A., Eitelman, S., Huang, C., Junker, B., Sklar, M. B., & Tobin, B. (2003). Evaluation of an automated reading tutor that listens: Comparison to human tutoring and classroom instruction. *Journal of Educational Computing Research, 29*, 117–161. https://doi.org/10.2190/06AX-QW99-EQ5G-RDCF
Mousavinasab, E., Zarifsanaiey, N., Niakan Kalhori, S. R., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments, 29*(1), 142–163.
National Reading Panel. (2000). Report of the national reading panel: Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction: Reports of the subgroups. *National Institute of Child Health; Human Development*.
Ni, A., & Cheung, A. (2023). Understanding secondary students' continuance intention to adopt ai-powered intelligent tutoring system for English learning. *Education and Information Technologies, 28*(3), 3191–3216.
Nyatsanga, S., Kucherenko, T., Ahuja, C., Henter, G. E., & Neff, M. (2023). A comprehensive review of data-driven co-speech gesture generation. *Computer Graphics Forum, 42*(2), 569–596.
Pérez-Marın, D., & Pascual-Nieto, I. (2013). An exploratory study on how children interact with pedagogic conversational agents. *Behaviour & Information Technology, 32*(9), 955–964.
Pitura, J. (2024). Enhancing advanced vocabulary in efl writing: An ai-assisted intervention for English studies students in Poland. *Journal of China Computer-Assisted Language Learning*, (0).
Puckett, M., Black, J., Wittmer, D., & Peterson, S. (2009). *The young child*. Upper Saddle River, NJ: Pearson.
Quan, S., Du, Y., & Lyu, Y. (2025a). Parents, children, and ChatGPT in home environments: the conversation content and the interaction mode. *In Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (pp. 1-6).
Quan, S., Ding, Y., Zhang, C., & George-Puskar, A. (2025b). *Impacts of Traditional, Digital, and AI-Integrated Home Literacy Environments on Children's Literacy Skills: A Cross-Sectional Study. Early Education and Development* (pp. 1–19). https://doi.org/10.1080/10409289.2025.2573078.
Quan, S., Du, Y., & Ding, Y. (2024). *Young children and ChatGPT: parents' use of ChatGPT in parenting. Extented Abstracts of the CHI Conference on Human Factors n Computing System* (pp. 1–7).
Roskos, K., & Neuman, S. B. (2003). For early literacy teaching and learning. *Handbook of Early Literacy Research, 1*, 281.
Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child Development, 83*(5), 1762–1774.
Saddhono, K., Suhita, R., Istanti, W., Kusmiatun, A., Kusumaningsih, D., & Sukmono, I. K. (2024). Ai-powered language learning: Enhancing literacy in the digital age. *2024 4th international conference on advancement in electronics communication engineering (AECE)*. https://doi.org/10.1109/AECE62803.2024.10911149
Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review, 14*(3), 245–302.
Schmucker, R., Xia, M., Azaria, A., & Mitchell, T. (2024). Ruffle&riley: Insights from designing and evaluating a large language model-based conversational tutoring system. *International conference on artificial intelligence in education*.
Sénéchal, M., & LeFevre, J.-A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73*(2), 445–460.
Sénéchal, M., Lefevre, J.-A., Thomas, E. M., & Daley, K. E. (1998). Differential effects of home literacy experiences on the development of oral and written language. *Reading Research Quarterly, 33*(1), 96–116.
Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction, 36*(6), 495–504.
Siegle, R. F., Schroeder, N. L., Lane, H. C., & Craig, S. D. (2023). Twenty-five years of learning with pedagogical agents: History, barriers, and opportunities. *TechTrends, 67*(5), 851–864.
Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. National Academy Press.
Solichah, N., & Shofiah, N. (2024). Artificial intelligence (ai) literacy in early childhood education: A scoping review. *Psikologika: Jurnal Pemikiran dan Penelitian Psikologi*. https://doi.org/10.20885/jstl.vol29.iss2.art1
Sonnenschein, S., & Munsterman, K. (2002). The influence of home-based reading interactions on 5-year-olds' reading motivations and early literacy development. *Early Childhood Research Quarterly, 17*(3), 318–337.
Spear-Swerling, L. (2019). Structured literacy and typical literacy practices: Understanding differences to create instructional opportunities. *Teaching Exceptional Children, 51*(3), 201–211.
Su, J. (2024). Kindergarten parents' perceptions of the use of ai technologies and ai literacy education: Positive views but practical concerns. *Education and Information Technologies, 30*, 279–295. https://doi.org/10.1007/s10639-024-12673-4
Su, J., & Yang, W. (2023a). Ai literacy curriculum and its relation to children's perceptions of robots and attitudes towards engineering and science: An intervention study in early childhood education. *Journal of Computer Assisted Learning, 40*, 241–253. https://doi.org/10.1111/jcal.12867
Su, J., & Yang, W. (2023b). Artificial intelligence (ai) literacy in early childhood education: An intervention study in Hong Kong. *Interactive Learning Environments, 32*, 5494–5508. https://doi.org/10.1080/10494820.2023.2217864
Temiz, G., & Kafadar, E. N. (2025). Utilization of ai-aided vocabulary teaching in k-12: A case study. *The Journal of Educational Research*, 1–8.
Tsang, S., Miao, R., Xiao, J., & Xiong, H. (2025). Animalte:designing ai-infused cartoon videos to improve preschoolers' language learning with family engagement at home. ArXiv, abs/2502.12526 https://doi.org/10.48550/arXiv.2502.12526.
Van Steensel, R. (2006). Relations between socio-cultural factors, the home literacy environment and children's literacy development in the first years of primary education. *Journal of Research in Reading, 29*(4), 367–382.
Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes, 86*. Harvard university press.
Wang, Q. (2025). Ai-driven autonomous interactive English learning language tutoring system. *Journal of Computational Methods in Sciences and Engineering, 25*(2), 1155–1166.
Wang, H., Tlili, A., Huang, R., Cai, Z., Li, M., Cheng, Z., Yang, D., Li, M., Zhu, X., & Fei, C. (2023). Examining the applications of intelligent tutoring systems in real educational contexts: A systematic literature review from the social experiment perspective. *Education and Information Technologies, 28*(7), 9113–9148.
Wang, Y., Wu, J., Chen, F., Wang, Z., Li, J., & Wang, L. (2024). Empirical assessment of ai-powered tools for vocabulary acquisition in efl instruction. *IEEE Access, 12*, 131892–131905. https://doi.org/10.1109/ACCESS.2024.3446657
Weigel, D. J., Martin, S. S., & Bennett, K. K. (2006). Contributions of the home literacy environment to preschool-aged children's emerging literacy and language skills. *Early Child Development and Care, 176*(3–4), 357–378.
Wen, Y., Chiu, M., Guo, X., & Wang, Z. (2024). Ai-powered vocabulary learning for lower primary school students. *British Journal of Educational Technology, 56*, 734–754. https://doi.org/10.1111/bjet.13537
Wu, J. (2024). Social and ethical impact of emotional ai advancement: The rise of pseudo-intimacy relationships and challenges in human interactions. *Frontiers in Psychology, 15*, Article 1410462.
Wu, Z., Halim, H. A., & Saad, M. R. M. (2024). Artificial intelligence (Ai) and gamification in blended learning: Enhancing language and literacy in Shanxi, China. *Malaysian Journal of Social Sciences and Humanities (MJSSH)*. https://doi.org/10.47405/mjssh.v9i12.3159
Xiao, F., Zhao, P., Sha, H., Yang, D., & Warschauer, M. (2024). Conversational agents in language learning. *Journal of China Computer-Assisted Language Learning, 4*(2), 300–325.
Xiao, F., Zou, E. W., Lin, J., Li, Z., & Yang, D. (2025). Parent-led vs. ai-guided dialogic reading: Evidence from a randomized controlled trial in children's e-book context. *British Journal of Educational Technology, 56*(5), 1784–1813.
Xu, Y., Aubele, J., Vigil, V., Bustamante, A. S., Kim, Y.-S., & Warschauer, M. (2022). Dialogue with a conversational agent promotes children's story comprehension via enhancing engagement. *Child Development, 93*(2), e149–e167. https://doi.org/10.1111/cdev.13733
Xu, Y., He, K., Levine, J., Ritchie, D., Pan, Z., Bustamante, A., & Warschauer, M. (2024). Artificial intelligence enhances children's science learning from television shows.

Journal of Educational Psychology, 116, 1071–1092. https://doi.org/10.1037/edu0000889

Xu, Y., He, K., Vigil, V., Ojeda-Ramirez, S., Liu, X., Levine, J., Cervera, K., & Warschauer, M. (2023). "rosita reads with my family": Developing a bilingual conversational agent to support parent-child shared reading. *Proceedings of the 22nd annual ACM interaction design and children conference*.

Xu, Y., Wang, D., Collins, P., Lee, H., & Warschauer, M. (2021). Same benefits, different communication patterns: Comparing children's reading with a conversational agent vs. a human partner. *Computers & Education, 161*, Article 104059.

Yang, W. (2022). Artificial intelligence education for young children: Why, what, and how in curriculum design and implementation. *Computers and Education: Artificial Intelligence, 3*, Article 100061. https://doi.org/10.1016/j.caeai.2022.100061

Yang, D., Ge, Y., Sun, Y., Collins, P., Jaeggi, S. M., Xu, Y., Shea, Z. M., & Warschauer, M. (2024). Self-regulation and comprehension in shared reading: The moderating effects of verbal interactions and e-book discussion prompts. *Child Development, 95*(6), 1934–1949.

Yaseen, H., Mohammad, A. S., Ashal, N., Abusaimeh, H., Ali, A., & Sharabati, A.-A. A. (2025). The impact of adaptive learning technologies, personalized feedback, and interactive ai tools on student engagement: The moderating role of digital literacy. *Sustainability, 17*(3), 1133.

Yigitalieva, Z., Dehqonov, I., Mamatova, A., Zinatullina, A., Tulanov, S., & Akbarova, S. (2024). The future of literacy education: Artificial intelligence and adaptive learning systems. *2024 international conference on IoT, communication and automation technology (ICICAT)*.

Yıldız, T. A. (2023). The impact of chatgpt on language learners' motivation. *Journal of Teacher Education and Lifelong Learning, 5*(2), 582–597.

Zhang, Q. (2025). Immersive ai-powered language learning experience in virtual reality: A gamified environment for Japanese learning. *2025 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW)*. https://doi.org/10.1109/VRW66409.2025.00455

Zhang, Z., & Huang, X. (2024). The impact of chatbots based on large language models on second language vocabulary acquisition. *Heliyon, 10*, Article e25370. https://doi.org/10.1016/j.heliyon.2024.e25370

Zhang, S. Z., Inoue, T., Shu, H., & Georgiou, G. K. (2020). How does home literacy environment influence reading comprehension in Chinese? Evidence from a 3-year longitudinal study. *Reading and Writing, 33*(7), 1745–1767.

Zhang, L., Jiang, Q., Xiong, W., & Zhao, W. (2025). Effects of chatgpt-based human–computer dialogic interaction programming activities on student engagement. *Journal of Educational Computing Research*, Article 07356331251333874.

Zhang, Z., Xu, Y., Wang, Y., Yao, B., Ritchie, D., Wu, T., Yu, M., Wang, D., & Li, T. J.-J. (2022). Storybuddy: A human-ai collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. *Proceedings of the 2022 CHI conference on human factors in computing systems*.

Zimmerman, M. (2018). Teaching ai: Exploring new frontiers for learning. *International Society for Technology in Education*.

Zimmerman, S. S., Rodriguez, M. C., Rewey, K. L., & Heidemann, S. L. (2008). The impact of an early literacy initiative on the long term academic success of diverse students. *Journal of Education for Students Placed at Risk, 13*(4), 452–481.