

# Enhancing Children’s Self-Reporting in Chatbot Diaries through Rhyming Style

Shanshan Chen

Department of Industrial Design  
Eindhoven University of Technology  
Eindhoven, North Brabant  
Netherlands  
s.chen1@tue.nl

Jun Hu

Department of Industrial Design  
Eindhoven University of Technology  
Eindhoven, North Brabant  
Netherlands  
j.hu@tue.nl

Gubing Wang

TS Social and Behavioral Sciences  
Tilburg University  
Tilburg, North Brabant, Netherlands  
Medical and Clinical Psychology  
Tilburg University  
Tilburg, North Brabant, Netherlands  
g.wang\_2@tilburguniversity.edu

Jing Li

Department of Industrial Design  
Eindhoven University of Technology  
Eindhoven, North Brabant  
Netherlands  
j.li2@tue.nl

Tzu-Hui Wu

Department of Industrial Design  
Eindhoven University of Technology  
Eindhoven, North Brabant  
Netherlands  
t.wu2@tue.nl

Panos Markopoulos

Department of Industrial Design  
Eindhoven University of Technology  
Eindhoven, North Brabant  
Netherlands  
P.Markopoulos@tue.nl

## Abstract

Children’s self-report is essential for research, education, and healthcare, yet existing methods such as surveys and diaries can be experienced as tedious and so lead to disengagement and low-quality responses. Chatbots have been suggested as a way to support children through conversational interaction, using age-appropriate language and an empathetic tone. Here we explore what could be suitable conversational styles for such chatbots. Specifically, we explore rhyme as a child-centered conversational style. We first conducted a co-design workshop with 35 children, which revealed preferences for short, playful, and soothing conversational patterns. Building on these insights, we designed a voice-based sleep diary in rhyming style and compared it to a prose style in a within-subjects study involving 40 children aged 8-12. Results show that rhyming prompts significantly improved response quality across question types and age groups, while maintaining high engagement even among children who preferred the prose style. We contribute proof-of-concept empirical evidence and design insights demonstrating how phonological scaffolding exemplified through rhyme extends the design space of capability-adapted chatbots beyond semantic simplification alone. While limited to short-term, lab-based sessions, this work provides initial evidence that conversational style can function as a design lever.

## CCS Concepts

• **Human-centered computing** → **Empirical studies in HCI**.

## Keywords

voice-based chatbot, conversational style, rhyme, children, self-report, sleep diary

### ACM Reference Format:

Shanshan Chen, Jun Hu, Gubing Wang, Jing Li, Tzu-Hui Wu, and Panos Markopoulos. 2026. Enhancing Children’s Self-Reporting in Chatbot Diaries through Rhyming Style. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI ’26)*, April 13–17, 2026, Barcelona, Spain. ACM, New York, NY, USA, 22 pages. <https://doi.org/10.1145/3772318.3791083>

## 1 Introduction

Children’s self-reporting plays a crucial role in research, education, and healthcare [78]. Self-reporting tools such as diaries, questionnaires, and surveys capture experiences that cannot be directly observed, such as emotions, behaviors, symptoms, or daily habits, and provide essential data to understand child development and well-being [74]. However, self-reporting remains challenging for children [105]. Even when surveys are designed with simplified wording or visual supports [1, 36, 65], their fixed question-and-answer format can still feel restrictive [94]. Many children perceive these tasks not as opportunities to share their experiences but as formal examinations or homework [62]. As a result, they may engage in satisficing [62]: offering the shortest possible responses, remaining silent when uncertain, or avoiding details for fear of being wrong or saying something unnecessary. These behaviors lead to disengagement, ultimately reducing both the consistency and richness of collected data.

Conversational agents supported by large language models (LLMs) open up new possibilities for interactive survey instruments. A key feature of these systems is their ability to simulate the flow of natural conversation, using tone, phrasing, and timing, making them feel more like a person than a scripted program [3]. This conversational flexibility allows them to transform rigid surveys and question lists into more natural dialogues, offering adaptability and a more supportive interaction experience [55, 101, 104, 115].



This work is licensed under a Creative Commons Attribution 4.0 International License. *CHI ’26, Barcelona, Spain*

© 2026 Copyright held by the owner/author(s).  
ACM ISBN 979-8-4007-2278-3/26/04  
<https://doi.org/10.1145/3772318.3791083>

These qualities are especially relevant for children, who often enjoy engaging with conversational agents in playful and learning contexts [19, 58, 85, 114]. However, most current approaches focus on adapting the dialogue to children’s capabilities and emotions. This includes simplifying vocabulary, shortening responses, or adjusting prompts to match their reading level and attention span, and providing empathic feedback designed to encourage self-disclosure [29, 59, 73, 83, 87]. We refer to these as capability-adapted chatbots. Yet an open question remains: is adapting to capabilities alone sufficient to sustain children’s engagement and elicit rich self-reports?

We argue that it is insufficient. First, prompting strategies aim to simplify dialogue for children, but LLMs do not always stay within those constraints. They often default to adult-like phrasing, reflecting the patterns emphasized in their training data. In other words, “adaptation” is the design goal, but “adult-centered drift” is the practical outcome, which produces responses that are still long, complex, or abstract for children to follow [19, 47, 90]. This mismatch not only reduces accessibility but also makes it difficult for children to provide detailed, high-quality responses, undermining the very goal of self-reporting. Second, capability-adapted chatbots typically focus on one-size-fits-all semantic alignment, such as rephrasing complex utterances, simplifying vocabulary, and tailoring content for children’s comprehension (e.g., [24, 111]). However, this adaptation often excludes richer interaction factors such as expressive prosody, rhythmic cadence, or playful voice modulation [61]. Because these systems usually apply one consistent style throughout the interaction, children’s diverse preferences and engagement needs may not be supported equally. The lack of expressive or phonological cues can make communication feel monotonous or flat, reducing attention, enjoyment, and thereby limiting children’s engagement [20, 69]. Together, these limitations suggest that while capability-adapted prompting offers improvements over rigid surveys, they remain insufficient to support child self-reporting. To be more effective, chatbots should incorporate child-centered conversational styles that resonate with the ways children naturally enjoy expressing themselves.

We propose designing chatbots that employ rhyme as a conversational style for children’s self-reporting. Rooted in nursery rhymes, chants, and songs, rhyme is a familiar and pleasant part of children’s everyday lives [49]. Its rhythmic and patterned form makes interactions more engaging, memorable, and fun to repeat [50, 80]. Importantly, rhyme introduces specific functions that directly address the shortcomings of capability-adapted chatbots for children. First, rhymes break language into short, rhythmic units, reducing cognitive load and making prompts more accessible across ages [70]. By avoiding long, complex, adult-centered phrasing, rhyming prompts can potentially help children better understand questions and provide higher-quality responses. Second, rhyme enriches conversational interaction through phonological play, such as patterns of rhythm and repetition that are enjoyable to say and easy to notice. Integrating such playful phonological cues into chatbots could potentially increase children’s enjoyment and engagement [49] supporting their willingness to participate over time. In short, rhyme directly addresses adult-centered complexity by simplifying language, while simultaneously extending conversational design beyond semantic phrasing toward phonological engagement aligned with children’s natural language development [40]. Despite these

potential benefits, rhyming style has not yet been explored as a conversational style to enhance children’s self-reporting in surveys.

We explore the potential of such a conversational strategy in the context of a chatbot-based sleep diary for children. In this tool, children record bedtimes, routines, nocturnal disturbances, and daytime functioning [12, 23, 92]. Sleep diaries are widely used and clinically important, but also notoriously difficult for children to complete consistently. Their static question list often leads to disengagement, resulting in missing or low-quality data [20]. Recent work has explored voice-based chatbots as alternatives to static questions, showing that conversational formats can make self-reporting more acceptable to children [21, 25, 60, 82]. Building on this foundation, we investigate whether integrating rhyme as a phonological element into a capability-adapted chatbot (rhyming style) can enhance children’s reporting performance and experience compared to a prose-based style. To examine the effects of this conversational design, we formulated two research questions:

- RQ1: How does a rhyming conversational style affect children’s response performance in a voice-based sleep diary in terms of the detail and quality of their responses, compared to a prose (non-rhyming) style?
- RQ2: How do children perceive rhyming chatbots in the context of self-reporting diaries, particularly in terms of engagement, preferences, and what are the reasons underlying their choices?

To address these questions, our research proceeded in two steps. First, we conducted a co-design study in which children acted as chatbot designers, generating ideas that we distilled into three guidelines for implementing a rhyming-style, LLM-powered voice chatbot. Second, we carried out a within-subjects study ( $1 \times 2$  design) with 40 children aged 8 to 12, comparing the rhyming style to the prose style. This study aimed to assess the effect of rhyming style independent of contextual factors such as daily routine, or adaptive switching that may impact the reporting behaviour of children in real-life. Our findings revealed that rhyme enhanced engagement and response performance across both descriptive and non-descriptive questions. Preferences varied by age: rhyming was especially appealing to younger children, while older children valued the naturalness of prose speech. Nevertheless, the advantage of rhyming style for response quality was maintained between age groups. Our research contributes in three levels:

- We extend the design space of child-agent conversational interaction by integrating a phonological cue (rhyme) as a complement to semantic-only capability adaptation. This positions phonological scaffolding as an underexplored and developmentally grounded design dimension for supporting children’s comprehension and engagement.
- We offer a reproducible implementation method informed by children’s participatory input, including LLM prompting and prosody control, design rationales, and text-to-speech configuration for generating rhyming conversational style in voice-based diaries.
- We provide proof-of-concept empirical evidence demonstrating that phonological cues in conversational user interfaces can improve children’s self-reporting performance under controlled conditions. These findings suggest the potential

of semantic-phonological styles and motivate future longitudinal field studies.

The present work focuses on understanding how phonological cues can be designed into children’s voice-based diaries, and on examining the functional contribution of conversational rhyme under controlled conditions. Our laboratory evaluation was purposefully structured to isolate this factor, rather than to observe long-term or in-the-wild self-reporting behavior of children. Our findings suggest the potential of phonological adaptation for children, and motivate future studies that will examine how semantic-phonological adaptation is experienced and how effective it can be in children’s everyday environments.

## 2 Related work

This section reviews previous work in three parts. We first discuss sleep diaries, outlining the shift from static forms to chatbot-based diaries, and the need to consider how these systems communicate with children. We then examine conversational style in LLM-powered chatbots for children, where capability-adapted dialogue has shown benefits, but also important limitations. Finally, we turn to rhyme as a child-centered conversational strategy. Building on its cognitive and affective functions, we highlight the potential of rhyme to enhance engagement and response quality in voice-based diaries.

### 2.1 Self-Reporting Chatbots for Sleep Diaries

Sleep diaries are widely used to monitor sleep patterns and behaviors in both clinical and research settings. They typically take two forms. The first is static in diaries that do not involve conversation. These include paper-based diaries and digital tools such as the Consensus Sleep Diary [17], a graphic diary [103], and EP-Sleepy [100]. Although static forms are simple to administer and support standardized reporting, their rigid form-like structure can make the task feel tedious and lead to survey fatigue [6, 76], particularly among children. To address these challenges, researchers have explored the second form: chatbot-based diaries [1, 27, 51], with a growing emphasis on voice-based conversational systems [21, 38]. Compared to static forms, voice-based chatbots align more closely with how children naturally communicate through speech. This form may improve the quality of the report by making the task feel less like filling out a test [6, 13, 76, 106].

Previous work has shown that voice-based tools can make self-reporting feel more natural to children [34, 38, 110]. However, their effectiveness depends not only on how children report (e.g., by speech instead of text), but also on how the system communicates with children [98]. Most research has examined the reporting modality (e.g., voice input versus text input), while much less attention has been paid to the system’s own communication strategies. We argue that this is a critical gap: the tone, rhythm, and linguistic framing of a chatbot can directly influence children’s willingness to respond, the accuracy of their recall, and the overall quality of reported data [98]. As LLM-powered chatbots for children become more widespread, exploring these features is technically easier and addressing this gap opens new design possibilities that can make reporting feel more accessible and engaging than is the case with conventional conversational interaction.

### 2.2 Conversational Styles in Child-Agent Interaction

Although sleep diary research has focused mainly on reporting formats, research highlights the importance of conversational style, such as tone, rhythm, and framing, in shaping user engagement and responses [18]. For example, warm or expressive interviewers tend to elicit more complete and engaged responses [45, 48]. Within the HCI community, researchers have examined how variations in conversational style affect user engagement and trust toward conversational agents, including high participation vs. high respect [7], warm vs. competent [79], and formal vs. casual (informal) [54, 57]. However, most of these studies focused on adults and text-based conversational agents, and have not yet explored how conversational and especially phonological adaptation can affect children’s spoken interactions with multi-modal conversational agents.

Recent research has begun to explore how LLMs can support capability-adapted chatbots for children [83]. Applications span a wide range of domains: educational activities [59], personalized story reading experiences [19], collaborative learning partners [33], design facilitators [60], emotional companions that encourage sharing personal experiences [83], tools to help children with autism practice emotional recognition [97], and aids for critical thinking and reflection [104]. In these contexts, capability-adapted chatbots supported by LLMs have been shown to increase children’s willingness to engage, interact, and share personal experiences [82, 86]. At the same time, increasing evidence indicates that adapting to children’s capability alone is insufficient. We discuss two major limitations: their effects on 1) response quality and 2) engagement.

*2.2.1 For response quality.* LLM-powered chatbots often default to adult-like responses, using long, complex, or abstract expressions that are difficult for children to understand and misaligned with their communication styles [19, 47, 90]. For example, parents reported that story-reading chatbots sometimes asked questions “from an adult’s perspective”, showing limited sensitivity to children’s developmental stage [19]. In educational contexts, chatbot-generated explanations have been reported to syntactically more complex than those of human instructors, potentially overwhelming younger learners [90]. Similarly, LLM-based language learning systems often rely on adult speech corpora, which limits their ability to process and respond effectively to children’s speech [47]. These mismatches reduce accessibility and hinder the children’s ability to produce detailed, high-quality responses.

*2.2.2 For the engagement.* Capability-adapted chatbots often assume a semantically directed interaction style. They primarily focus on what is said (e.g., simplifying vocabulary or phrasing), while overlooking how language sounds and feels through phonological features such as rhythm, rhyme, and prosody. However, research shows that children engage less when interacting with conversational agents that deliver content without expressive richness [20, 69]. Research in speech perception shows that children’s language processing integrates both semantic and phonological cues [30, 40]. While semantic adaptation supports clarity and comprehension, phonological patterns enhance memorability, enjoyment, and emotional resonance [15, 50]. In everyday life, phonological play, through songs, chants, and rhymes, supports early literacy,

sustained attention, and social participation [49, 80]. Without these features, a purely semantic conversational style may feel flat and uninspiring, reducing children’s enjoyment, sense of agency, and ultimately their willingness to continue engaging with the tool. For example, in child-robot interaction, children show higher behavioral and emotional engagement with a human instructor than with a robot delivering the same tasks when the robot lacks expressive richness [69]. These findings suggest that expressive, phonologically rich interaction is critical for maintaining children’s attention and affective connection.

In summary, while capability-adapted chatbots improve accessibility over form-filling surveys, they remain limited in sustained children’s engagement and performance in self-reporting. To address this gap, our work explores phonologically enriched conversational styles, particularly rhyme, as a child-centered design strategy that extends beyond current capability-adapted paradigms.

### 2.3 Rhyme as a Conversational Strategy for Self-Reporting in Children

Rhyme is a dominant and powerful form of language in childhood [49]. Children encounter it daily in nursery rhymes, playground chants, songs, and word games across cultures [70]. Its repetition, rhythm, and predictability embody a sense of playfulness that resonates with children [96]. Rhyming could serve two functions that make it especially relevant for children’s self-reporting: improving accessibility for high-quality responses and enhancing engagement through phonological playfulness.

**2.3.1 Accessible phrasing for higher-quality response.** By turning long sentences into short, rhythmic chunks, rhyme makes language easier for children to follow [4, 8, 70]. In child-computer interaction, rhythmic speech patterns have been shown to maintain conversational flow, especially in language-learning contexts [63]. Educational research further demonstrates that rhymes improve memory and comprehension by providing predictable auditory structures [96]. For example, children who hear rhyming stories recall more words and narrative details than those exposed to prose [46]. Anticipating repeated sounds fosters participation and a sense of success [2], while predictable rhythm reduces cognitive load and makes processing more automatic [70]. Together, these findings suggest that rhyming could make chatbot prompts more accessible and supportive of clearer, higher-quality responses than conventional prose phrasing.

**2.3.2 Playful phonological cues for engagement.** Rhymes are a familiar and cherished part of children’s cultural practices, appearing in songs, chants, and word games [52]. As a key phonological cue, alongside syllables, rhythm, and prosody, rhyme introduces an element of playfulness that supports bonding, relaxation, and shared enjoyment [93]. Engaging with rhymes has been shown to evoke positive affect and reduce tension, particularly in group contexts [50, 66]. Their predictable structure helps maintain attention and counteract fatigue during repetitive tasks [11, 42]. Classroom studies similarly link rhyming activities with greater enjoyment and motivation [42]. Experimental work further demonstrates that rhymed phrases reduce attentional decline compared to prose [66]. In human-robot interaction, robots using rhymed speech elicited

greater attention and responsiveness from children than those using prose [80]. These findings highlight the potential of rhyme to make mundane interactions more playful, emotionally engaging, and motivating.

In summary, rhyme is suggested as a child-centered conversational strategy that enhances both response quality through accessible phrasing and improves engagement through playful phonological cues. Despite its ubiquity and documented benefits, rhyming has rarely been applied in children’s self-reporting diaries. This gap is notable because current LLM-powered chatbots mainly adapt to children’s capabilities (e.g., vocabulary and attention span) but overlook playful phonological strategies. Our study addresses this gap by investigating rhyme as a conversational style in LLM-powered voice-based diaries.

## 3 Co-design Study

To understand which design features should inform the rhyming-style chatbot for children, we first engaged children directly in the design process. Co-design can reveal children’s communication preferences and avoid designs based solely on adult assumptions [68, 89, 113]. In this study, by positioning children as designers, we aimed to capture their preferences and expectations for how a chatbot should speak. The children were not asked to design the rhyme specifically, but we learned from their spontaneous design choices to formulate our rhyming prompts.

### 3.1 Method

**3.1.1 Participants.** We recruited 35 children aged 8 to 12 ( $M = 9.39, SD = 1.18$ ), including 14 girls and 21 boys, from local primary schools and public libraries in the Netherlands. This age range was chosen because children younger than 8 often struggle with independent written expression and sustained conversational tasks, while those older than 12 begin to adopt more adult-like conversational norms that differ from those of younger children [16, 72]. All participants were native English speakers with typical reading and writing abilities and prior experience using a laptop. Ethical approval was obtained from the university ethics board with written consent from both children and parents. Parents did not attend the sessions.

**3.1.2 Procedure.** After providing consent, each child participated in a 30-minute workshop consisting of three phases. Two researchers were present throughout. The individual format allowed children to concentrate on conceptualizing and articulating their chatbot design ideas without managing peer interaction. Figure 1 is a collage showing two photos of children designing a chatbot using a laptop and one screenshot of the design result with Landbot. Figure 2 shows the procedure of the co-design workshop.

**Warm-up: Getting familiar with the interface (5 minutes).** Children first completed a short practice activity to familiarize themselves with the Landbot platform [56]. With step-by-step assistance from a researcher, each child created a simple welcome chatbot that could greet users and ask one basic question. This ensured that all participants were comfortable with the interface before starting the design task.

**Co-design: Creating a sleep-story chatbot (15-20 minutes).** After the warm-up, each child began the main design task on a blank



**Figure 1: Three scenes from the co-design workshop involving children designing chatbot interactions using Landbot. (A) A researcher assists a child in creating a simple chatbot flow. (B) A child independently designs using the platform on a laptop. (C) A screenshot of a child's design result with Landbot.**

Landbot canvas, with no preloaded dialogue blocks or sample content. Researchers provided only technical scaffolding (e.g., how to add, connect, and test nodes). Importantly, no linguistic or stylistic examples were offered. Children were not shown sample questions, rhyming lines, conversational templates, or couplet structures. This ensured that all wording and turning-taking structure originated from children's own ideas.

To support imagination without constraining content, we introduced two optional animal avatars (a rabbit and a fox). These served purely as narrative prompts for imagining the chatbot's "voice" did not prescribe tone or interaction style. Children freely chose whether to use an avatar and independently determined the chatbot's tone, pacing, and phrasing.

Children were invited to design a "sleep-story chatbot" that guided a user through a short interactive story. This framing provided a high-level purpose (collecting a story), but gave full freedom as to how the chatbot would ask questions or respond. Through the task, researchers offered technical assistance only (e.g., connecting flow blocks), deliberately avoiding suggestions about working, tone, or style. Children iteratively tested and refined their designs by running the chatbot.

This approach ensured that scaffolding was about the use of the platform, not the content, and that all conversational features emerged from the children's designs.

**Reflection: Role-playing and explanation (5-10 minutes).** Each child demonstrated their chatbot through a brief role-play: the child acted as the chatbot, and a researcher played the user. They performed the dialog aloud in a face-to-face exchange rather than simply reading text on-screen. This enabled researchers to capture the tone, expressions, and intended conversational style behind the written dialogue. After the role-play, children explained their design choices, including wordings, tone, and conversational flow, and how they imagined other children might respond. Field

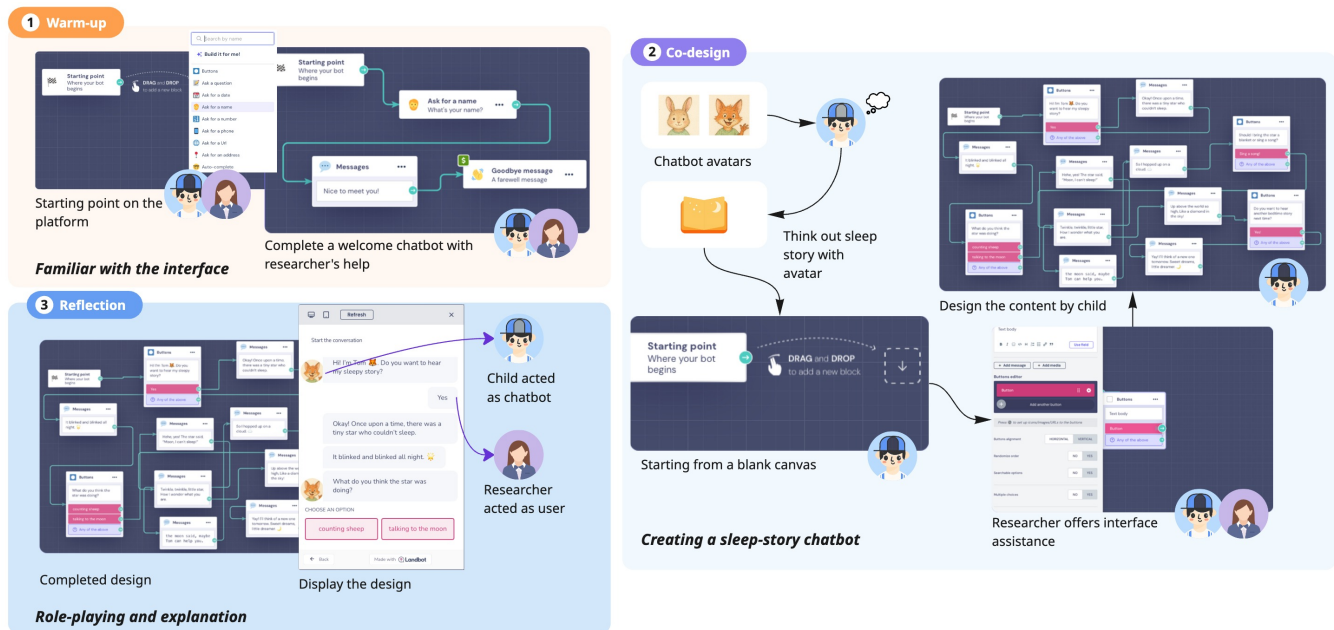
notes and video recordings captured the rationale for children's designs.

### 3.2 Data Analysis

The purpose of the co-design workshop was to identify design elements that could inform a rhyming conversational style. To focus the analysis on aspects relevant to accessible phrasing, and playful phonological cues, we conducted a deductive thematic analysis [53].

**Deductive coding frame.** We developed a theory-informed coding frame grounded in three strands of literature relevant to how children engage with spoken language. Cognitive load theory [95] highlights that shorter turns reduce processing demands and support comprehension. Research on phonological awareness [91] explains how children attend to rhythmic, repetitive, and sound-based patterns. Work on paralinguistic tone and mood in speech [81] describes how vocal qualities shape emotional comfort, which is particularly important in bedtime contexts. Guided by these perspectives, we defined three analytic dimensions: 1) turn length, referring to how long the chatbot speaks in each turn; 2) sound patterning, referring to how utterances are structured, including opportunities for rhythmic or playful expression; and 3) tone, referring to how the bot should sound in bedtime contexts. These three deductive dimensions framed the analytic lens used in the coding.

**Coding procedure.** Two researchers independently coded the entire dataset (all transcripts and chatbot designs) using the initial coding frame. After this first round, they held a reconciliation meeting to review every instance of disagreement. For each discrepant code, they read the excerpt aloud together to ensure shared interpretation, explained their coding rationale using the coding frame, compared interpretations against code definitions, and discussed whether the codebook required refinement, revised the code definitions collaboratively when a discrepancy reflected ambiguity in the original frame, finally, reached consensus through discussion,



**Figure 2: Overview of the co-design workshop procedure. The figure summarizes the three phases of the workshop—Warm-up, Co-design, and Reflection—using a combination of interface screenshots, icons, and simplified illustrations.**

rather than majority vote or hierarchical decision-making. A second round of independent coding followed using the revised codebook. Inter-rater reliability across all codes was high  $\kappa = 0.89$ . The remaining discrepancies were resolved through iterative reconciliation meetings until full agreement was reached.

**Theme development.** Within each deductive dimension, we identified recurring behavioral and linguistic patterns in the children’s designs. These patterns formed the themes reported below, which describe how children naturally structured chatbot conversations. Each theme also informed the design decisions for the rhyming-style prompts. Figure 3 illustrates the deductive thematic analysis process.

### 3.3 Results & Key Findings

Children’s designs revealed consistent patterns in how they imagined a chatbot should speak and behave. These patterns aligned closely with the three deductive dimensions described above.

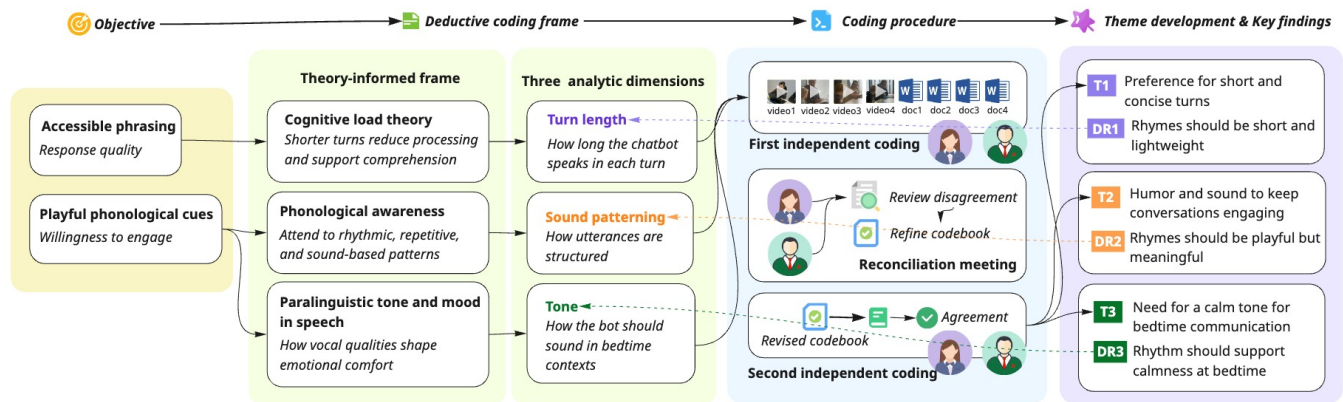
**T1: Children prefer short, concise turns.** Children consistently designed brief responses rather than long turns, showing a preference for lightweight and fast-paced interactions. Across all designs, we identified 23 instances of one-sentence replies or short follow-ups, such as “What time did you sleep? Ready to play or rest?” (Figure 1, C), “Did you dream? Tell me quick!”. Children did not attempt long explanations; instead, they naturally kept each turn short and tightly structured. While prior work argues that rhyme can break language into shorter, more easily processed units [70], our workshop provides empirical evidence that children themselves design conversations in this concise, chunked form. This directly informed our rhyming-style prototype: rather than composing long poetic lines, each prompt was written as a two-line

couplet, serving as a rhymed analogue to the short turns children preferred. This maintained cognitive simplicity while introducing phonological play.

**T2: Children use humor and sound to keep conversations engaging.** Many children incorporated humor and light teasing into their chatbots to make conversations feel lively and fun. We observed 20 instances where children created “peek-a-boo” or joke-based exchanges, such as “You guess? I wouldn’t tell you yet!”, or used exaggerated emojis for comic effect. During role-playing, three children used onomatopoeic sounds (e.g., “boing!”, “zzz!”) to amuse researchers, and 12 designs included optional joke paths or short verse-like phrases. These behaviors highlight children’s preference for playful, dynamic exchanges over purely semantic information delivery. This informed our decision to make the rhyming prompts light-hearted and rhythmically expressive, while avoiding overwhelming the bedtime tone.

**T3: Children expect a calm tone for bedtime communication.** While children enjoyed playful exchanges, eight children emphasized that a chatbot should adopt a calmer tone at bedtime in the explanation. Five children noted that it should not be “too loud” or “too silly” before sleep. For instance, one child created a fairy-tale bot that greeted softly with the rabbit avatar, “My dear friend, I’ve been waiting for you.” This highlights the need to balance playful rhyme with a soothing rhythm appropriate for bedtime use.

These findings highlight what children enjoy in the conversational style, but they do not directly translate into design guidelines. To determine how these preferences should influence the development of a self-reporting diary, we analyzed each theme in relation to



**Figure 3: Overview of the deductive thematic analysis process.** The left is objective and theoretical framing, which is guided by two objectives of this study and three theoretical lenses. The middle is the coding procedure. Two coders independently coded all videos and design scripts, and a reconciliation meeting was held to resolve disagreements and refine the codebook. They then independently recoded in the second round with a high agreement. The right is themes and design rationales. Each theme maps to the design rationales, which is also related to the former three analytic dimensions.

cognitive load, bedtime context, and accuracy of reporting. This interpretation resulted in three design rationales that map children’s ideas into constraints for a rhyming conversational style.

## 4 System Design and Implementation

This section presents the design and implementation of our voice-based sleep diary chatbot, focusing on how insights from the co-design workshop informed the development of a rhyming conversational style for children. We first introduce the three design rationales distilled from the workshop, which guided how rhyme was integrated into the dialogue experience. We then describe the dialogue flow, prompt structure, and technical implementation of the rhyming-style diary, followed by the prose-style baseline representing a capability-adapted, non-rhyming chatbot commonly used in pediatric voice-based diaries.

### 4.1 Design of LLM-powered Diaries with Rhyming Style

Building upon findings from the co-design study, we developed a voice-based sleep diary using an LLM to generate real-time responses in rhymed form.

**4.1.1 Conversation Design Rationales.** Based on the findings from our co-design workshop, we derived three evidence-based design rationales (DR1-DR3) to inform the development of the rhyming conversational style. Each rationale operationalizes a workshop insight, ensuring that the rhyming style is grounded in children’s design choices rather than relying on poetic artistry alone.

**DR1: Rhymes should be short and lightweight.** Following T1, we constrained the length of each chatbot turn to two or three brief rhymed lines (approximately  $\leq 15$  words each). This preserves the low cognitive effort and fast-paced turn-taking preferred by children, while introducing rhythmic playfulness.

**DR2: Rhymes should be playful but meaningful.** Building on T2, we incorporated light phonological play (e.g., rhyme, rhythm)

to sustain engagement. However, we required each line to remain semantically relevant to the diary question to avoid distracting children from reporting accurately.

**DR3: Rhythms should support calmness at bedtime.** In response to T3, we adopted a soothing rhythmic style, supported by a slower speaking rate, lower pitch, and natural pauses. This avoids overstimulation, aligning with bedtime routines.

**4.1.2 Dialogue System with LLM.** We designed a voice-based sleep diary powered by an LLM that presented questions in rhyming style. The sleep diary consisted of 12 structured questions adapted from the adult-oriented Consensus Sleep Diary [17] and further refined in child-focused HCI research to suit developmental needs [20]. In that work, the refinement process involved consultation with two licensed clinicians (a pediatric sleep specialist and a child psychologist), who examined the item wording for clarity, and age-appropriate vocabulary. Adjustments focused on simplifying linguistic complexity, clarify temporal references, and avoiding terminology unlikely to be familiar to primary-school children (e.g., "sleep onset latency" with "How long did it take you to fall sleep?"). The present study used the finalized child-adapted version reported in [20] without further modification (Table 1).

```

20-12-2024 20:26:46
Chatbot: Great to hear! Let's continue without fear! What did you do before
sleeping last night, my dear?
Child: read books

20-12-2024 20:27:03
Chatbot: Reading books is a wonderful way, to end your busy day! What you ate last
night, can you say?
Child: dumplings

20-12-2024 20:27:25
Chatbot: Dumplings are yummy, in your tummy! Enjoy every bite, until it's out of
sight! What did you drink, before you said good night?
Child: drink water
    
```

**Figure 4: Example transcript snippet of a rhyming-style chatbot conversation, stored in Firebase.** The text illustrates how the chatbot acknowledges a child’s previous answer and responds using short rhyming couplets.

**Table 1: Sleep diary questions used during the chatbot interaction and classification by question type. Each question was asked once per session in fixed order.**

Q#	Question	Type
Q1	How are you feeling today?	Descriptive
Q2	How long did it take you to fall asleep?	Non-descriptive
Q3	Did you sleep well?	Descriptive
Q4	What did you do before sleeping last night?	Descriptive
Q5	What did you eat before sleeping last night?	Descriptive
Q6	What did you drink before sleeping last night?	Descriptive
Q7	How many times did you wake up last night?	Non-descriptive
Q8	When did you sleep last night?	Non-descriptive
Q9	How long did you nap or doze this daytime?	Non-descriptive
Q10	How long did you sleep last night?	Non-descriptive
Q11	When did you wake up this morning?	Non-descriptive
Q12	How did you wake up this morning?	Descriptive

*Dialogue Flow:* At each turn, the system used the LLM to generate a rhyming response that did two things: acknowledged the child’s previous answer and rephrased the next diary question. To keep interactions concise and engaging, the LLM was instructed to generate brief, playful couplets (e.g. "And what did you drink to help you feel great? Was it cool water or something sweet on your plate?"). This allowed the diary questions to remain precise and structured, while maintaining a child-friendly style. Each diary session followed a fixed sequence of 12 turns, corresponding to the 12 diary questions. For every scripted question, the chatbot provided a personalized acknowledgment followed by the next prompt. No follow-up probes, clarification loops, error-handling branches, or adaptive re-prompting occurred in either condition. Therefore, conversational load, defined as the number of dialogue turns, was held constant across all participants and both conversational styles, ruling out turn count as a potential confounding factor. Figure 4 illustrates a screenshot showing a text transcript of a rhyming chatbot conversation stored in Firebase.

*Prompt Design and LLM configuration:* The rhyming style was implemented using GPT-3.5. The prompt design was guided by the three design rationales (DR1-DR3) derived from the co-design study: keep concise turn length (DR1), include playful rhyme and rhythm (DR2), and maintain a soothing bedtime tone (DR3). To support reproducibility and method transparency, Table 2 reports the full LLM configuration, generation constraints, and text-to-speech prosody settings. These choices were informed by the functional goals of the diary task and by principles for designing child-agent interaction. Linked to DR1 (concise turns), we used a max-token limit (60) and a stop-sequence constraint to enforce 2-3 short lines per turn, thereby reducing linguistic load and supporting predictable pacing. Linked to DR2 (playful phonological cues), we used a moderate temperature (0.7) to balance creativity with coherence, and did not apply presence or frequency penalties, as controlled repetition can support rhyme appreciation and playful language engagement

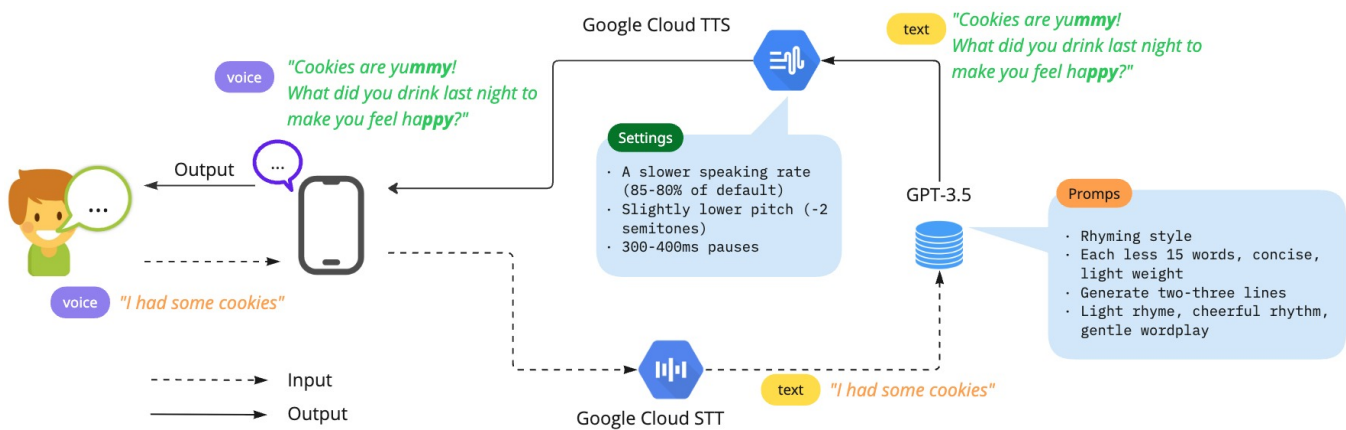
in middle childhood. Linked to DR3 (soothing bedtime tone), all generated content was rendered using text-to-speech with prosodic adjustments (reduced speaking rate, lowered pitch, and short inter-line pauses) to promote calm nighttime interaction, rather than energetic or comedic expression. The full prompt templates used in the study are provided in Appendix A to support replication and adaptation by future researchers. Figure 5 shows how DR1-DR3 informed the iterative prompt design process.

*Dialogue Process:* Figure 6 illustrates this process. The full interaction loop consists of the following steps: 1) the chatbot asks a question from the sleep diary; 2) the child verbally answers the chatbot’s question (Figure 6, ❶); 3) the system transcribes the child’s answer and prepares to generate the next chatbot utterance (Figure 6, ❷); 4) the next diary question is selected and combined with the child’s previous response to form the prompt (Figure 6, ❸); 5) this prompt, along with the rhyming style, is sent to the LLM (GPT-3.5) (Figure 6, ❹); 6) GPT-3.5 generates a response that includes: a) a personalized acknowledgment of the child’s previous answer and b) the next diary question, phrased in the rhyming style (Figure 6, ❺); 7) the generated text is sent to the Google Cloud TTS to set the tone of bedtime, converted to speech (Figure 6, ❻); 8) the speech is delivered to the chatbot (Figure 6, ❼); 9) the cycle repeats until all 12 questions are completed (Figure 6, ❽). This dynamic generation process ensured that each turn the rhyming style while adapting flexibly to the child’s input.

**4.1.3 Implementation.** The system was implemented as a voice-based chatbot application for Android smartphones in Java. It integrates three key APIs: Google Cloud Text-to-Speech API v1, Speech-to-Text API v1, GPT-3.5. The app is available in [language]. Previous research has shown that Google speech recognition technology provides sufficient transcription accuracy for children [21]. Consequently, we used Google tools for both speech input (voice

**Table 2: Language model and text-to-speech (TTS) configuration used in the rhyming condition. These settings control generation length, rhyme behaviour, and bedtime-appropriate prosody.**

Component	Setting / Constraint
Temperature	0.7
Max tokens	60
Top-p	0.9
Presence/Frequency penalties	0
Stop sequence	Enforces 2–3 output lines
Persona instruction	Calm, bedtime-friendly companion
Turn format rule	2–3 lines, ≤ 15 words/line
Content rule	Line 1 acknowledges; Line 2 asks next question
Rhyme rule	Soft, natural rhyme (non-forced, non-comedic)
Prosody adjustments	Speaking rate 0.85–0.90; pitch –2 semitones; 300–400ms pauses



**Figure 5: System flow for generating rhyming chatbot responses. The child’s speech is transcribed via Google Cloud STT, processed through GPT-3.5 using rhyme-based prompting constraints, and returned using bedtime-appropriate text-to-speech (TTS) settings.**

recognition and transcription) and output (text-to-speech synthesis). The conversational flow is powered by GPT-3.5, which generates context-aware responses in rhyming style based on the child’s previous answer. The language model is asked to produce child-friendly responses while preserving the structure of a daily sleep diary interaction.

*App Interfaces:* The application featured three main screens (Figure 7): 1) Home interface allows children to begin the reporting task after logging in (Figure 7, A); 2) Conversational interface presents each diary question, records the child’s spoken response, and generates a response in the rhyming style. A black moving line with a walking man at the top of the interface indicates the progress of answering this time (Figure 7, B); and 3) Completion interface displays a congratulatory message after all diary questions have been answered (Figure 7, C); pressing "Thanks!" returns to the home screen (Figure 7, A).

## 4.2 Control Condition: LLM-powered Prose Style

The control condition used the same deployment platform (Android), back-end infrastructure (Firebase), and core APIs (Google Cloud Speech-to-Text v1, Google Cloud Text-to-Speech v1, GPT-3.5) as the rhyming style. The user interface was also identical for all conditions. The only difference was in the conversational prompts and delivery.

In the control version, the capability-focused chatbot adopted the standard LLM-powered "prose style" commonly used in previous work: supportive, but direct, and prose-based [19, 83]. The prompts instructed GPT-3.5 to acknowledge the child’s answer warmly and then present the next diary question in plain language, without rhyme or rhythmic constraint. The voice output was synthesized with the default Google TTS settings, without prosodic adjustments.

All transcripts were automatically saved to Firebase, and filenames were coded by participant ID and condition (e.g., 'C1\_rhyme.txt' for Participant 1 using the rhyming-style version). Figure 8 shows an example of parallel conversations in both conditions.

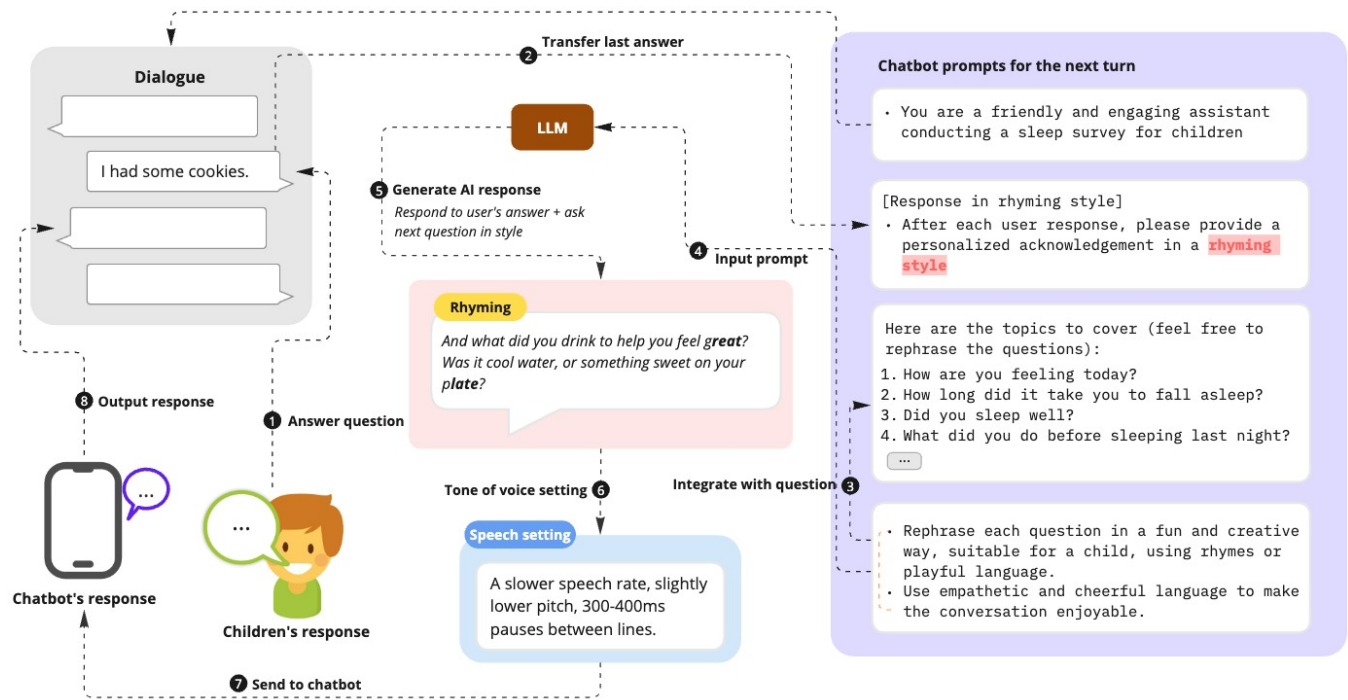


Figure 6: Overview of the voice-based dialogue system used for the sleep diary, showing how a child’s spoken response is processed, transformed into a rhyming chatbot utterance, and returned using bedtime-appropriate text-to-speech.

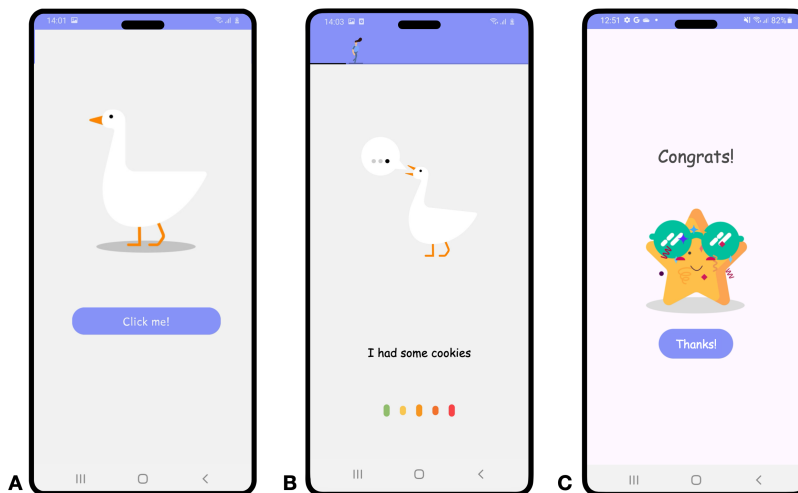


Figure 7: Three screenshots from the mobile sleep-diary chatbot application showing key user interfaces: (A) the home screen with a start button, (B) the conversational interface where the child’s transcribed response is displayed, and (C) the completion screen showing a congratulatory message.

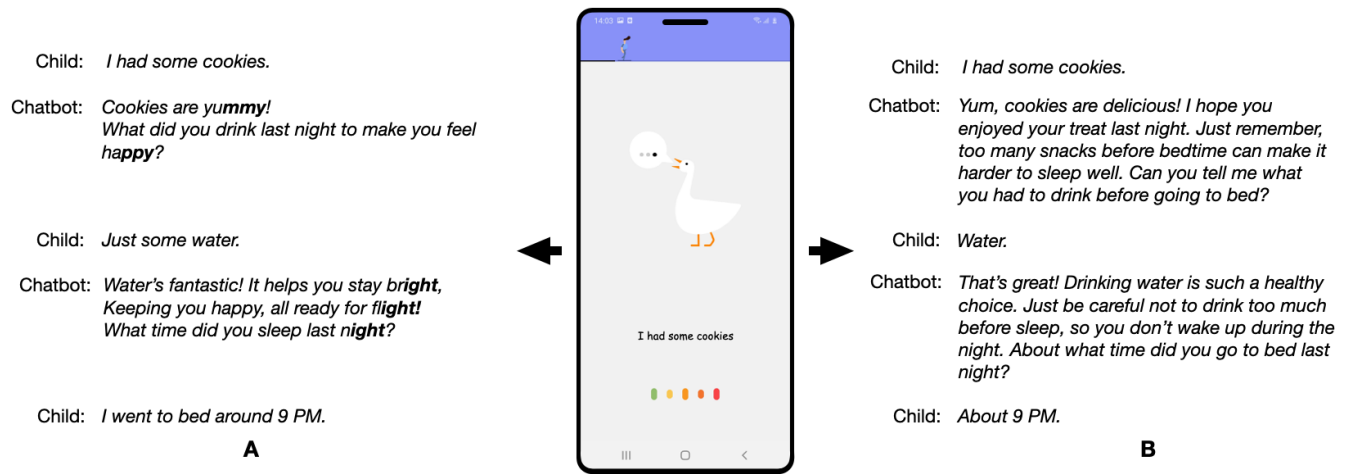
## 5 User Study

To explore the effect of the conversational style in the voice-based sleep diary on children’s response quality and perceptions, we conducted a  $1 \times 2$  within-subjects experiment comparing the rhyming style with the prose style. Each participant completed both versions

of the diary in counterbalanced order, followed by an individual interview to capture qualitative feedback.

### 5.1 Sample Size and Power Analysis

To justify the target sample size, we conducted an a priori power analysis using G\*Power 3.1 [35]. For a repeated-measures design



**Figure 8: Example dialogues illustrating two conversational styles used in the voice-based sleep diary chatbot. (A) The rhyming condition, in which the chatbot replies using short, rhythmic, and playful couplets, while still requesting relevant sleep information. (B) The prose condition, in which the chatbot provides a content-equivalent but plain, explanatory response without rhyme.**

with one within-subject factor (rhyming vs. prose), assuming a medium effect size ( $f = 0.25$ ) following Cohen’s guidelines [28], an  $\alpha$  level of 0.05, and desired power of 0.80, the required sample size was 34 participants. Our final analyzed sample included 40 children, exceeding this requirement and providing sufficient power to detect medium-sized effects.

## 5.2 Participants

The evaluation study did not include children who had participated in the earlier co-design workshop, avoiding familiarity or expectation bias. Consistent with the developmental rationale underlying the co-design workshop, we focused on children aged 8-12, a range suitable for sustained conversational interaction while still reflecting child-specific communication patterns [16, 72]. Two participants outside this developmental range (aged 7 and 13) were therefore excluded from the analysis. The final sample includes 40 children 8-12 ( $M = 9.65, SD = 1.56$ ; 21 girls, 19 boys). Children were recruited by approaching their guardians in a local public library and through neighboring primary schools. All participants were native English speakers.

Because previous research suggests that children’s appreciation of rhyme changes with age [116], we examined whether style preferences differed by developmental stage. However, it remains unclear whether there is a turning point in this developmental trajectory and at what age such changes occur. To explore potential age-related differences, we divided participants into two groups: children aged 8-9 (“younger group”) and 10-12 (“older group”). The younger group included 17 children ( $M = 8.11, SD = 0.86$ ) and the older group included 23 children ( $M = 10.79, SD = 0.80$ ). Table 3 summarizes demographic characteristics and smartphone use by age group. According to parents, four children in the younger group had no prior experience with smartphones.

**Table 3: Participant demographic characteristics and smartphone use frequency for the full sample, younger group ( $\leq 9$  years), and older group ( $\geq 10$  years). Values are shown as mean (SD) or number (percentage).**

	Full sample	Younger group	Older group
Age in years	9.65 (1.56)	8.11 (0.86)	10.79 (0.80)
Female	21 (53%)	11 (52.4%)	10 (47.6%)
<i>Smartphone use</i>			
Daily	9 (22.5%)	2 (11.8%)	7 (30.4%)
Weekly	18 (45%)	8 (47.1%)	14 (60.9%)
Monthly	9 (22.5%)	3 (17.6%)	2 (8.7%)
Never	4 (10%)	4 (23.5%)	0%
<i>N</i>	40	17	23

Since the user test involved reading, listening, and speaking tasks on a smartphone, we ensured that all participants had typical abilities in these areas. As the study aimed to explore the effect of conversational styles on children’s self-reporting in sleep diary, inclusion criteria did not consider whether participants had sleep disorders.

Ethics approval was obtained from our university’s Ethics Review Board. To protect privacy, all personal data, including demographic information, audio recordings, and conversations synchronized on Firebase, were deleted immediately after the study. Participation was voluntary and required both parental consent and child assent. As a token of appreciation, each child received a set of LEGO™.

## 5.3 Procedure

Each session followed a standardized sequence to ensure that participants understood the experiment. These steps also ensure that

we could collect both quantitative and qualitative data. Sessions lasted 40-50 minutes and were structured as follows (Figure 9):

In the familiarization phase (stage 1), the researchers introduced the purpose and structure of the study. To ensure that the children were comfortable with the system, they first tried short practice runs with both conversational styles. This allowed them to experience the interface, practice answering diary questions, and gain confidence before beginning the main task.

In the main task (stage 2), each child completed the two diaries conditions, one in the rhyming style and one in the prose style. After each diary condition, participants completed the Giggle Gauge [32] to report their engagement with that style. The order was fully counterbalanced: half of the participants began with the rhyming condition, while the other half began with the prose condition. To reduce fatigue and order effects, a five-minute break involving light play activities (e.g., casual video games) separated the two conditions (Figure 10).

Finally, in the reflection phase (stage 3), the children completed the Again-Again table [77], which measures the willingness to continue using in the future. A short semi-structured interview followed, in which participants described their preferences and experiences for the conversational styles in the self-reporting tasks. These qualitative insights were used to understand the reasons behind their behaviors and inform later design decisions.

## 5.4 Data Collection

To answer the research questions on children’s response quality and engagement across two conversational styles, we collected and analyzed interaction data from all diary sessions. Each child’s spoken responses were recorded, processed, and converted into text transcripts, which served as the primary dataset for linguistic and content-based analysis. Before detailing the two measurement aspects, we first describe how conversational data were captured, stored, and prepared for analysis, since this processing pipeline forms the basis for quantifying response quality.

**5.4.1 Conversational Data Processing.** All voice-based diary interactions were automatically logged through Firebase, where each turn was stored as a time-stamped text transcript (Figure 4). The analysis used the child response segments only, excluding system utterances and confirmation lines. Each transcript was cleaned to remove speech-to-text artifacts (e.g., filter words, repeated punctuation) while preserving the semantic content of the child’s answer. These processed utterances were then used to compute linguistic and content-based measures, including information units (described below), which serve as one component of the response quality.

**5.4.2 Response Quality.** To assess how well children reported their sleep experiences (RQ1), we adopted the Response Quality Index (RQI), adapted from Xiao et al. [109] and Chen et al. [20]. The RQI captures the extent to which a response is informative, relevant, and clear, reflecting both content richness and communicative effectiveness from Gricean Maxims [41]:

**Information units** represent the number of distinct, meaningful details in a response [99]. This measure of content richness is widely used in children’s language research (e.g., [112]). Because children’s vocabularies are limited and many words are high-frequency, the

alternative surprise-based metric proposed by Xiao et al. [109] is less appropriate in this context.

**Relevance** indicates whether the response addresses the prompt (1 = on-topic, 0 = off-topic).

**Clarity** reflects whether the response is complete and comprehensible (1 = clear, 0 = unclear).

Information units were automatically extracted using spaCy and checked for accuracy on 15% of the data; agreement with human counts was high ( $ICC = .93$ ). Relevance and clarity were independently coded by two raters ( $\kappa = .91, \kappa = .92$ ).

Following Chen et al. [20], the composite RQI score was calculated as:

$$RQI = \text{Information units} \times \text{Relevance} \times \text{Clarity} \quad (1)$$

This multiplicative formulation reflects the idea that a response must be both informative and communicatively usable: a detailed but irrelevant answer (“I went to bed” to the question (“What did you eat last night?”) receives an RQI of 0, as does an unclear answer “I eat eat eat eat”. Thus, the index penalizes off-topic or incomplete responses regardless of length.

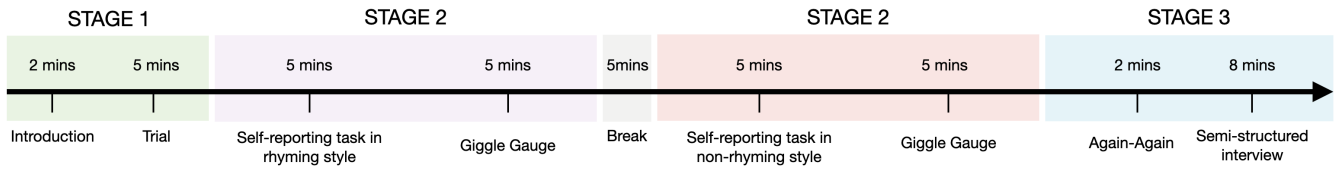
To assess how well RQI reflects response richness, we computed its correlation with raw response length (number of words). RQI was moderately correlated with response length (*Spearman’s*  $\rho = .29, p < .001$ ), consistent with expectations that higher-quality responses tend to contain more detail but are penalized when off-topic or unclear. This supports the validity of the composite measure.

**5.4.3 Engagement and Style Preferences.** To understand children’s attitudes towards conversational styles (RQ2), we collected quantitative and qualitative data. For quantitative measures, we used two established instruments: the Giggle Gauge [32] and Again-Again Table [77].

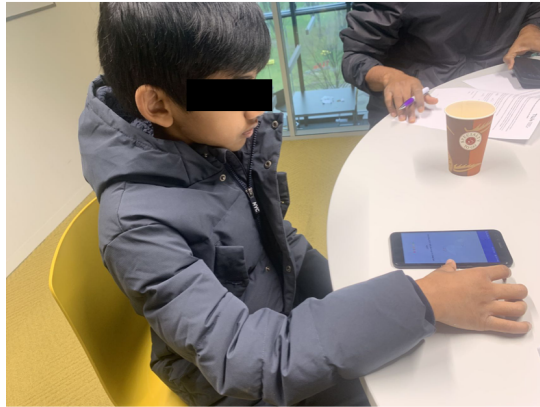
The Giggle Gauge is a validated instrument designed for child-computer interaction that assesses seven dimensions with a four-point scale: challenge, aesthetics, feedback, interest, novelty, endurability, and perceived user control. In our sample, the overall internal consistency of the scale was acceptable (*Cronbach’s*  $\alpha = 0.82$ ), supporting its use for comparing children’s engagement across conversational styles. Each item is scored from 1 to 4, where 4 presents the highest level of engagement. The scale combines a yes/no judgment about whether the child experienced the attribute with a rating of how strongly they felt it, producing four ordered levels. Dietz et al., [32] provide practical interpretive ranges: 1)  $< 3.0$  is low engagement; 2)  $3.0 - 3.6$  presents moderate engagement; and 3)  $\geq 3.6$  indicates high engagement typical of well-received interactive systems. We use these ranges to interpret children’s ratings for each conversational style.

The Again-Again Table [77] captures children’s willingness to repeat the activity. After completing both styles, the children were shown the simple table and asked “Which would you like to use again in the future?” choosing between rhyming and prose.

For qualitative measures, we analyzed semi-structured interviews to explore the reasons behind children’s preferences, attitude, and perceptions of the two styles.



**Figure 9: Timeline of the experimental procedure. The study consisted of three stages: (1) introduction and trial session, (2) two self-reporting tasks using rhyming and non-rhyming conversational styles with short breaks and engagement check-ins, and (3) post-task preference and reflection interview.**



**Figure 10: A participant using the voice-based sleep diary app during the study session. The child interacts with the mobile interface while seated in a supervised research setting.**

**5.5 Result**

Two participants discontinued participation for personal reasons before completing both diary sessions; therefore, the final analytic sample consisted of 38 participants, with no partial or incomplete data included. The study addressed the two research questions: 1) to examine the effect of conversational style on children’s response quality and 2) to understand children’s engagement and preferences to conversational styles in self-reporting. Before the analysis, we conducted a manipulation check to ensure that participants could notice and distinguish between the two conversational styles.

**5.5.1 Manipulation Check.** During the pre-training phase, each child experienced both conversational styles and was asked to indicate their preferred style in a binary choice task. This check served two purposes: confirming that the children could differentiate between the styles and providing a first look at their preferences.

The subsequent interviews further contextualized these choices. Children who preferred the rhyming style often described it as "fun" and "entertaining", noting that they enjoyed listening to the chatbot "talk like that". Those who preferred the prose style valued it as "more natural" and closer to real-life conversations. These comments suggest that the children clearly recognized the stylistic differences and made choices based on different perceptions of tone and engagement.

Quantitatively, 57.5% of the children preferred the rhyming style, while 42.5% preferred the prose style. This roughly even split highlights the diversity of children’s conversational preferences.

**5.5.2 Influence of Conversational Styles on Response Performance.** We examined whether conversational style (rhyming vs. prose) influenced children’s response performance, measured using the details and quality of response. The analyzes considered both 1) overall response quality and 2) effect across questions.

**Overall Response Quality:** We first conducted a descriptive analysis comparing children’s overall response quality between the two conversational styles (Table 4). Children interacting with the rhyming-style chatbot produced responses with a higher mean RQI ( $M = 2.88, SD = 1.58; 95\% CI [2.36, 3.40]$ ) than those who interact with the prose-style chatbot ( $M = 2.25, SD = 1.48; 95\% CI [1.76, 2.73]$ ). A similar pattern was observed between age groups, with older children producing higher-quality responses than younger children, particularly in the rhyming condition. Because these descriptive statistics do not account for the within-subject structure of the experiment, we next fit a linear mixed-effects model (LMM) to analyze differences while accounting for repeated measurements per child.

**Table 4: Overall Response Quality Index (RQI) by conversational style and age group. Higher scores indicate richer and more detailed responses. Values represent Mean ± SD with 95% confidence intervals (CI).**

Group	RQI (Mean ± SD)	ci_lower	ci_upper
Rhyming Style	2.88 (1.58)	2.36	3.40
Prose Style	2.25 (1.48)	1.76	2.73
Younger & Rhyming	2.69 (1.43)	2.06	3.32
Younger & prose	2.03 (1.23)	1.48	2.57
Older & Rhyming	3.14 (1.79)	2.18	4.19
Older & prose	2.55 (1.76)	1.61	3.48

To model overall response quality, we used an Linear Mixed Model for a random intercept for participant and fixed effects for Style (Rhyming vs. Prose), Age (Younger vs. Older), and their interaction [5, 37]. Residual diagnostics (Q-Q plots and residuals vs. fitted values) indicated no serious deviations from normality or homoscedasticity, so we retained the untransformed RQI scores.

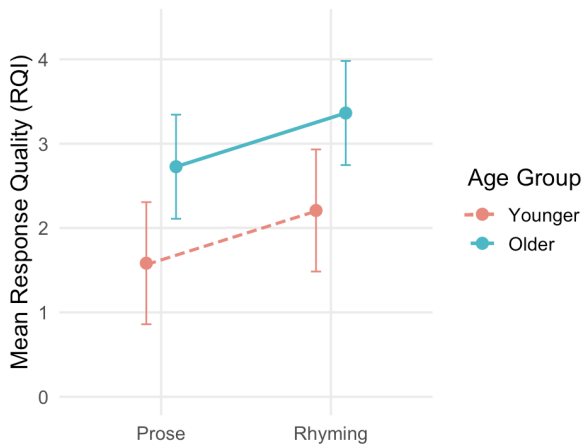
The model revealed a significant main effect of conversational style: rhyming elicited higher overall response quality than prose ( $\beta = 0.63, SE = 0.27, p = 0.03$ ). Age group also had a significant main effect ( $\beta = 0.114, SE = 0.47, p = 0.02$ ), with older children giving higher-quality response overall. The Style × Age interaction

was not significant ( $\beta = 0.01, SE = 0.36, p = 0.97$ ), indicating that the advantage of rhyming over prose was similar for younger and older children (Table 5).

**Table 5: Effects of overall response quality. The model formula is  $RQI \sim Style * Age + (1|ParticipantID)$ , where  $RQI$  = response quality index,  $Style$  = Conversational styles (rhyming, and prose style),  $Age$ =Age group (younger, and older),  $ParticipantID$  = participant ID. \*\*\* $p < 0.001$ , \* $p < 0.05$ .**

Predictor	Estimate	Std. Error	df	t	p
(Intercept)	1.58	0.36	47.77	4.40	< .001***
Style	0.63	0.27	36.00	2.23	.03 *
QuestionType	1.14	0.47	47.77	2.42	.02 *
Style:QuestionType	0.01	0.36	36.00	0.03	.97

Figure 11 visualizes these effects. Older children consistently produced higher-quality responses, and rhyming improved response quality across both age groups. The roughly parallel slopes reflect the non-significant interaction. These findings suggest that the rhyming style supports richer reporting and does so reliably across developmental stages in children.



**Figure 11: Mean Response Quality Index (RQI) across conversational styles (Prose vs. Rhyming) and age groups (Younger vs. Older). Error bars represent 95% confidence intervals. Older children and the rhyming condition both show higher mean RQI; lines are roughly parallel, illustrating a non-significant interaction effect.**

*Influence on Response Quality across Questions:* Because response quality can depend on the type of question asked [9, 88], we categorized all 12 diary questions into descriptive and non-descriptive types [22]. Descriptive questions refer to those requiring answers with contextual details, elaboration, or narrative explanation (e.g., “What did you eat before sleeping last night?”) and non-descriptive questions refer to those that can be answered with a single factual, or numeric (e.g., “When did you sleep last night?”) [22]. Although yes/no questions are often non-descriptive, we classified “Did you

sleep well?” as a descriptive question because it requires children to evaluate and describe the subjective quality of their sleep experience rather than provide a factual or numeric value. In practice, children’s responses to this question frequently contained elaboration (e.g., “Yes, because I fell asleep quickly.” or “No, I woke up many times.”), which aligns with definitions of descriptive, experiential reporting in prior child self-report research [71, 84]. Two authors independently classified all questions according to these definitions, achieving high agreement ( $Cohen'sk = 0.97$ ). The final classification is provided in Table 1.

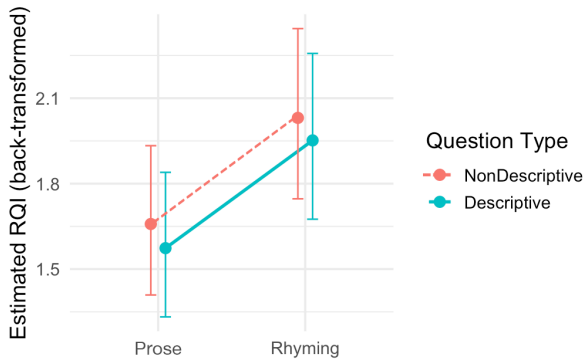
To examine whether conversational styles interacted with the question type, we fitted an LMM with conversational style, question type (descriptive vs. non-descriptive), and their interaction as fixed effects, and participant as the random effect to account for repeated measures [5, 37]. Residual diagnostics indicated deviations from normality and heteroscedasticity, so RQI scores were log-transformed [60]. As shown in Table 6, the model revealed a significant main effect of conversational style: rhyming elicited higher-quality responses across questions ( $\beta = 0.53, SE = 0.18, p < 0.01$ ). By contrast, the question type did not have a significant effect ( $\beta = 0.02, SE = 0.18, p > 0.05$ ), and the  $Style \times QuestionType$  interaction was also non-significant ( $\beta = 0.11, SE = 0.25, p > 0.05$ ). Figure 12 illustrates that both descriptive and non-descriptive questions benefit similarly from rhyming prompts.

**Table 6: Effects of conversational style on response quality across questions. The model formula is  $RQI_{log} \sim Style * ResponseType + (1|ParticipantID)$ , where  $RQI$  = response quality index,  $Style$  = Conversational styles (rhyming, and prose style),  $QuestionType$ =Question type (descriptive, and non-descriptive),  $ParticipantID$  = participant ID. \*\*\* $p < 0.001$ , \*\* $p < 0.01$ , \* $p < 0.05$ .**

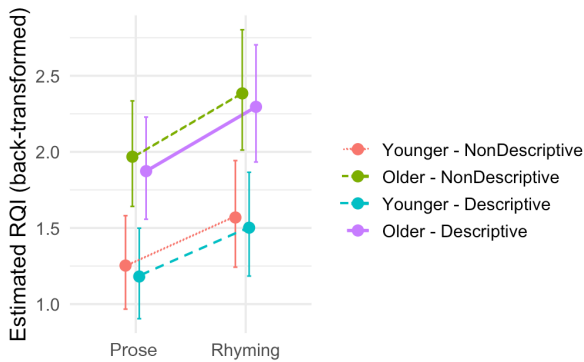
Predictor	Estimate	Std. Error	df	t	p
(Intercept)	0.98	0.19	86.57	70.16	< .001***
Style	0.53	0.18	871.00	2.50	.003 **
QuestionType	0.02	0.18	871.00	0.13	.90
Style:QuestionType	0.11	0.25	871.00	0.43	.67

To test whether these effects held across age groups, we extended the model by adding age group as an additional fixed effect. In this extended model, the conversational style remained a significant predictor, with rhyming outperforming prose ( $\beta = 0.67, SE = 0.28, p = 0.01$ ). Age was also significant, with older children producing higher-quality responses ( $\beta = 0.76, SE = 0.37, p = 0.02$ ). In contrast, question type again did not significantly influence response quality ( $\beta = 0.28, SE = 0.28, p = 0.30$ ), and the  $Style \times QuestionType$  interaction remained non-significant ( $\beta = 0.44, SE = 0.39, p = 0.26$ ). As shown in Figure 13, older children generally gave richer responses, but the relative advantage of rhyming over prose was consistent across question types and age groups.

In summary, rhyming prompts consistently elicited higher-quality responses than prose style, and this advantage held across both descriptive and non-descriptive questions. When age was included, the rhyming effect remained statistically significant: older children tended to provide richer responses than younger children, but the



**Figure 12: Estimated mean Response Quality Index (RQI), back-transformed from the linear mixed-effects model, across conversational styles (Prose and Rhyming) and question types (Descriptive vs. Non-Descriptive). Error bars represent 95% confidence intervals. Both question types show higher estimated RQI under rhyming prompts, and the slopes are similar, indicating no meaningful interaction effect between question type and style.**



**Figure 13: Estimated Response Quality Index (RQI), back-transformed from the mixed-effects model, across conversational styles (Prose vs. Rhyming), question types (Descriptive vs. Non-Descriptive), and age groups (Younger vs. Older). Error bars represent 95% confidence intervals. Across all four subgroups, rhyming prompts yielded higher estimated RQI than prose, and older children produced richer responses than younger children. Slopes are visually parallel, indicating no notable three-way interaction among style, age, and question type.**

relative benefit of rhyming style was consistent between groups. These findings address RQ1 by demonstrating that rhyming style can improve response quality for the different types of questions and the ages we considered.

**5.5.3 Children’s Engagement and Preferences to Conversational Styles.** To address RQ2, we analyzed children’s preferences and attitudes with mixed methods: first, by analyzing their engagement and preferences quantitatively and then by exploring interview data to understand the reasons behind their perceptions.

*Perceptions to Conversational Styles:* Children’s preferences were measured using the Again-Again scale that reflects children’s willingness to continue to use in the future, and Giggle Gauge that reflects children’s engagement and experience with the system across seven dimensions.

As shown in Table 7, 61.53% of the children preferred the rhyming style overall, with the younger children showing a stronger preference (75%) compared to the older children (50%). Engagement ratings on Giggle Gauge were consistently high for both conversational styles. Across its seven dimensions, all mean scores fell between 3.6 and 3.8, a range that Dietz et al. [32] characterize as typical of high-engagement systems. Thus, both styles were well-received, though rhyming showed a slight advantage. In the rhyming condition, 35 out of 38 children scored high engagement in the rhyming style ( $E \geq 3.6$ ), compared with 32 of 38 in the prose condition. Within the rhyming style, 23 children rated the highest level (4) for aesthetics (“I like how the app looked and felt”), 18 rated 4 for enjoyment (“I enjoy using it”), and 14 rated 4 for endurability (“I would like to do this again sometime”). Together, these results suggest that the rhyming style elicited strong aesthetic appeal, sustained interest, and high repeat-use willingness – key components of engagement in previous work [32].

A two-way ANOVA confirmed these patterns. There was a significant main effect of conversational style, with higher engagement reported in the rhyming condition than in the prose condition,  $F(1, 72) = 6.82, p = 0.001$ . Neither age nor interaction between age and style were significant, suggesting that although preferences varied somewhat by age, the engagement of the rhyming style was consistent between groups.

**Table 7: Descriptive statistics for children’s conversational style preference and engagement scores across age groups (Younger = 8–9 years, Older = 10–12 years). A majority of younger children preferred the rhyming style, whereas preference among older children was more balanced across conditions. Engagement scores were similar across both styles and age groups.**

Measures	All	Younger group	Older group
Preference (Rhyming)	61.53% (23/38)	75% (12/16)	50% (11/22)
Preference (prose)	39.47% (15/38)	25% (4/16)	50% (11/22)
Engagement (Rhyming)	3.65 (0.07)	3.63 (0.08)	3.64 (0.12)
Engagement (prose)	3.60 (0.07)	3.61 (0.07)	3.60 (0.09)

*Reasons for Children’s Attitudes:* We applied a deductive thematic analysis [10] to analyze interview transcripts. The interview transcripts were coded for the experience of participants with the two conversational styles in the voice-based chatbot and their explanations of the self-reporting.

Two coders independently developed an initial codebook by open-coding the data. After that, they collaborated to discuss individual codes, followed by a second round of independent coding using the emerging codebook. Then, they reconvened to resolve disagreements, clarify coding details, and finalize the codebook. The coding process demonstrated high consistency and accuracy, with an inner-rater reliability of 88%, as measured by Cohen's Kappa [64].

Our thematic analysis revealed three high-level themes that explain children's preferences:

**Rhymes make reporting fun and engaging.** Many children (23/38) described rhymes as fun and playful, often comparing it with a game. They explained that the rhythm and sound patterns made them laugh and kept their attention: "Speaking with it is fun. Its voice makes me laugh" (C21). Six children highlighted its novelty, noting that it felt distinct from everyday speech: "It sounds not the same with my daily talking" (C18). Seven also appreciated the poetic flow, enjoying the cadence itself: "I like to listen to it speak. It sounds like poem" (C5). These accounts suggest that the novelty, rhythm, and playfulness of the rhyme contributed to a more enjoyable experience.

**Prose felt supportive and more natural.** Fifteen children preferred the prose style because it felt more natural and supportive to follow. They described plain speech as guidance, especially for straightforward questions: "Sometimes it provides explanations and suggestions, it's what I hope to hear from it" (C37). Eleven valued the naturalness of non-rhyming dialogue: "Without rhyming sounds more natural" (C11). Nine also emphasized its supportive tone, which they found more engaging: "I like its encouraging talking... But rhyming does not say that" (C24). These comments highlight that naturalness, familiarity, and encouragement were the primary strengths of the prose style.

**Styles preferences depended on context and task.** Children noted that the suitability of rhyme versus prose varied by situation. Eight children admitted that while the rhymes were engaging, they could also be distracting: "Sometimes I'm listening to the rhyme and forgetting the question" (C30). Nine emphasized bedtime appropriateness, preferring gentle, calming rhymes before sleep: "For bedtime, not too silly, soft is better" (C16). Five noted that the rhyme fit better with open-ended or emotional questions, whereas plain speech was preferable for factual ones: "If it's 'what time,' just ask normal. For feelings, rhymes" (C28). Thirteen also expressed a desire for choice, asking whether they could switch styles depending on their mood: "Can I press a button to make it rhyme or not?" (C12).

Together, these findings illustrate a nuanced picture: rhyming was generally playful and engaging, especially for younger children, while the prose style offered naturalness and encouragement valued more by older children. The tensions between styles suggest that flexibility, allowing children to adapt style to context, may be key to sustaining engagement and supporting high-quality self-reporting.

## 5.6 Discussion

To contextualize our findings and clarify how rhyming conversational style shapes children's self-reporting, we structure the

discussion around three complementary lenses: 1) empirical interpretation of results, 2) theoretical contributions, and 3) practical implications for designing conversational agents for children.

**5.6.1 Empirical Interpretation of Results.** This work examined whether a rhyming conversational style can address the two limitations in current LLM-powered chatbots for children, in terms of response quality and willingness to engage. We highlight three main interpretations:

**Rhyming improved response quality through cognitive scaffolding rather than semantic enrichment.** Rhyming significantly increased response quality compared with prose, and this effect held across both descriptive and non-descriptive questions. This suggests that rhyming did not alter the content children were able to recall, but it appeared to facilitate verbal expression, enabling children to articulate what they knew more fluently and with greater detail. Rhyming likely supported production through phonological scaffolding: the predictable structure of rhyme may reduce linguistic formulation demands, allowing children to allocate more attention to content. This aligns with developmental literature showing that rhyme supports memory, recall, and utterance planning in middle childhood [39, 67].

**Age contributed additively, not interactively, to response quality.** Older children consistently produced richer responses, but age did not interact with conversational style. This indicates that developmental maturity and interaction design contribute independently. In other words, the advantage of rhyming does not depend on language maturity; instead, it scales proportionally across ages. This indicates that rhyming supplements rather than substitutes for developmental advantage.

**Engagement patterns reveal a trade-off between novelty-driven excitement and pragmatic comfort.** Children reported higher engagement and greater preference for rhyming overall, but interviews revealed that this enthusiasm was partially context-dependent rather than absolute. Rhyming was perceived as fun, aesthetic, game-like, and novel. These properties are known to sustain attention and reduce formality [26, 110]. However, prose was valued for clarity, familiarity, and perceived instructional support, especially for factual or bedtime-oriented exchanges. Rather than a universal preference, children implicitly treated conversational style as a modifiable interaction resource, suggesting that flexibility may be more beneficial than a static approach. Because these interactions occurred in a single session, some of the excitement associated with rhyming may reflect a novelty effect rather than durable engagement. Long-term evaluation is therefore needed to assess whether these effects are sustained as routine forms.

In summary, these findings highlight that conversational style is not merely a surface-level aesthetic choice but a cognitive and affective design consideration that meaningfully shapes how children communicate with conversational agents.

**5.6.2 Theoretical Contribution.** This study advances theoretical understanding at the intersection of developmental psychology and child-agent interaction by identifying how conversational style functions as both a developmental scaffold in children's self-reporting.

**Developmentally responsive conversational preference.** The findings refine developmental accounts of children’s conversational preferences by showing that susceptibility to phonological play is age-sensitive rather than universally optimal [75]. Younger children demonstrated stronger attraction and benefits from the rhyming interaction (75%), consistent with work showing that phonological cues such as rhythm, rhyme, and prosody support early language processing attention, and affect regulation [40]. Older children showed more varied preferences (50%), aligning with developmental shifts toward semantic precision, pragmatic relevance, and conversational naturalness [14, 40]. These results challenge the assumption that a single interaction style can support all children equally, emphasizing the need for developmentally tuned style design that evolves with linguistic maturity.

**Conversational scaffolding operates on form, not only content.** This work conceptually extends the design space of child-agent interaction beyond the domain paradigm of semantic capability adaptation. Existing child-friendly conversational agents primarily adjust vocabulary, sentence complexity, or pragmatic framing, assuming that meaning-level adaptation is sufficient. Our findings suggest that conversational scaffolding involves two complementary dimensions: semantic clarity (what is said) and phonological structure (how it is said). By empirically demonstrating that phonological cues can improve self-expression without distorting task focus, we introduce phonological scaffolding as a theoretically grounded and design-relevant axis for future child-agent interactions.

**A non-anthropomorphic pathway for sustaining engagement.** This work offers a theoretical perspective on positioning child-facing conversational agents along the tool-companion continuum. Recent LLM-based conversational agents often rely on personalization, emotional mirroring, or persona construction to sustain engagement [19, 43, 59], raising concerns around anthropomorphism, attachment, and blurred relational boundaries. In contrast, rhyming interaction maintains playfulness and engagement while not relying on human likeness, because poetic cadence foregrounds stylistic artifice rather than social identity. We therefore propose rhyme as a non-anthropomorphic engagement strategy, supporting sustained interaction without promoting social substitution.

Collectively, these contributions reorient child-agent design theory toward conversational form, not only conversational content, and demonstrate that phonological scaffolding can increase communicative productivity while preserving ethical boundaries for young users.

**5.6.3 Practical Implications for Designing Conversational Agents for Children.** Our findings offer concrete design guidance for creating LLM-powered conversational agents that better support children’s self-reporting.

**Design for development diversity, not uniform "child-mode".** Most LLM-powered chatbots rely on semantic simplification (e.g., adjusted words, friendlier tone) [44, 83, 102], but our results show that this alone may not promote high-quality performance. Stylistic preferences and benefits vary developmentally: younger children responded more positively to rhythmic, playful language, while older children showed more mixed or pragmatic preferences. Rather

than deploying a single "child-friendly" dialogue model, designers should create age-adaptive conversational profiles, for example by selecting when to use rhyme, neutral prose, or user-controlled mode switching.

**Broaden conversational design from semantic clarity to semantic-phonological duality.** Current design guidelines emphasize what conversational agents say, but our findings suggest that how it is spoken matters for engagement and expressive output [83]. Phonological features (rhythm, rhyme, pacing, prosody) can be selectively integrated (e.g., as opening prompts, turn-transitions, summaries, or emotion-focused segments) to reinforce attention and enjoyment without overwhelming task focus. This dual-dimension approach can be prototyped using controllable prompting and prosody parameters without requiring major language-generation changes. Designers should also enable fade-in/fade-out strategies, where playful phonological cues gradually reduce for tasks requiring precision.

**Extend applicability beyond sleep diaries and across child-centered domains.** Rhythmic, scaffolded conversational styles may benefit other reflection-oriented or motivation-sensitive technologies such as therapy check-ins, learning diaries, symptom logs, reading support, and creative writing tools [82, 107, 108]. These techniques may also be effective in family-mediated tools for shared emotional discussion or well-being routines.

**Ethical guardrails.** Because playful or stylized conversational agents can increase engagement [31, 110, 114], designers should consider boundaries that preserve the agent as a tool rather than a substitute social partner. This includes avoiding persona-based emotional attachment cues, limiting persuasive gamification, and offering transparency and user control (e.g., allowing children to switch styles, opt out, or reset tone).

In summary, these findings show that conversational style is a designable cognitive scaffold, not merely an aesthetic feature, repositioning dialogue form as functional interaction resource.

## 6 Limitations and Future Work

This study was conducted in a one-session, lab-based setting, which may influence children’s engagement and perceived enjoyment due to limited ecological complexity, novelty, and demand effects. These short-term findings therefore should not be interpreted as indicators of long-term adherence, durable motivation, or real-world data reliability. Lab conditions also fail to capture contextual variation such as bedtime routines, parental presence, environmental distractions, and emotional states, all of which meaningfully affect self-report. Our results should therefore be viewed as proof-of-concept regarding the potential impact of conversational styles: phonological scaffolding can enhance expressive quality under controlled conditions, but its long-term behavioral efficacy remains to be established.

Future work will extend this study in three directions. First, we will conduct an in-the-wild longitudinal deployment in children’s homes to examine whether rhyming continues to support engagement and response quality when novelty fades and daily routine forms. Children will complete self-report independently in their bedtime environment, allowing us to capture behavioral indicators of adherence, such as missed responses, interruptions, and

changes in report length. Second, we will explore contextual and adaptive conversational designs, including mixed-style interaction (e.g., prose + rhyming) and child-controlled style switching based on task intent or personal comfort. This approach enables us to observe style preferences reflected in real use, not only self-reported attitudes. Third, future work will investigate how phonological scaffolding interacts with ecological factors, such as sleep context, emotional disclosure, and family dynamics, through behavioral logs synchronized with timing and optionally supported by lightweight parent or child reflections after home use.

Additionally, our study was conducted in a Western European, primarily urban context, with a research team bringing positionalities from Industrial Design and Clinical Psychology, and cultural backgrounds spanning Asia and Europe. We provide detailed descriptions of the setting and procedures to support the transferability of our findings, while acknowledging that different contexts and researcher identities may yield different insights. We encourage future research and researchers in diverse cultural and socio-economic settings to build on and extend these findings.

## 7 Conclusion

This paper examined how conversational style shapes children's self-reporting in voice-based sleep diaries. Through a co-design workshop and a controlled lab study, we explored rhyming prompts as a phonological design strategy that complements existing capability-adapted chatbots. Our findings provide initial evidence that rhyme can support richer verbal responses across both descriptive and non-descriptive questions, while also maintaining high engagement even among children who preferred prose. Although preferences varied by age, the relative advantage of rhyme for response quality remained consistent across groups.

These results suggest that conversational style is more than a surface-level aesthetic choice: It can function as a cognitive and affective scaffold that broadens the design space of child-agent interaction. Integrating phonological cues, such as rhythm and rhyme, adds a complementary dimension to existing semantic adaptation approaches, while also helping to maintain a clear boundary between conversational tool and social companion.

As a proof-of-concept contribution, our findings highlight the potential of rhyming interaction under controlled conditions; however, engagement effects may be influenced by short-term novelty and limited ecological complexity. Future work will evaluate how these benefits carry over to everyday contexts, including longitudinal home deployments, adaptive style switching, and broader use cases in health and education.

In summary, this work takes an early step toward rethinking chatbot-based self-reporting systems by showing how conversational style can be leveraged as a developmentally informed design lever. We hope that the strategies introduced here will support future exploration of playful, flexible, and ethically grounded conversational styles in child-computer interaction.

## Acknowledgments

We used GPT-5 to assist with grammar checking and wording improvements. All ideas, analyses, and interpretations are those of the authors. This work was supported by the China Scholarship

Council (CSC). We sincerely thank the anonymous reviewers for their constructive feedback.

## References

- [1] Tessa Aarts, Panos Markopoulos, Lars Giling, Tudor Vacaretu, and Sigrid Pillen. 2022. Snoozy: A Chatbot-Based Sleep Diary for Children Aged Eight to Twelve. In *Proceedings of the 21st Annual ACM Interaction Design and Children Conference (Braga, Portugal) (IDC '22)*. Association for Computing Machinery, New York, NY, USA, 297–307. <https://doi.org/10.1145/3501712.3529718>
- [2] Sophie Alcock. 2007. Playing with rules around routines: Children making mealtimes meaningful and enjoyable. *Early Years* 27, 3 (2007), 281–293. <https://doi.org/10.1080/09575140701594426>
- [3] Hayder Kareem Mudheher Algabri. 2024. *Artificial Intelligence and ChatGPT*. AG Publishing House (AGPH Books), India. <https://www.amazon.in/dp/8197041342> Paperback, 239 pages. Country of Origin: India.
- [4] Paul D Allopenna, James S Magnuson, and Michael K Tanenhaus. 1998. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of memory and language* 38, 4 (1998), 419–439. <https://doi.org/10.1006/jmla.1997.2558>
- [5] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of statistical software* 67 (2015), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- [6] Pazit Ben-Nun. 2008. Respondent Fatigue. In *Encyclopedia of Survey Research Methods*. Vol. 2. SAGE Publications, Inc., 742–743. <https://doi.org/10.4135/9781412963947.n480>
- [7] Gregory A Bennett. 2018. Conversational style: Beyond the nuts and bolts of conversation. In *Studies in Conversational UX Design*. Springer, 161–180. [https://doi.org/10.1007/978-3-319-95579-7\\_8](https://doi.org/10.1007/978-3-319-95579-7_8)
- [8] Daniel E. Berlyne. 1974. *Studies in the New Experimental Aesthetics: Steps Toward an Objective Psychology of Aesthetic Appreciation*. John Wiley & Sons Inc, New York, 348 pages.
- [9] Megan E Bradley, Lindsay R Thom, Jennifer Hayes, and Cindy Hay. 2008. Ask and you will receive: How question type influences quantity and quality of online discussions. *British Journal of Educational Technology* 39, 5 (2008), 888–900. <https://doi.org/10.1111/j.1467-8535.2007.00804.x>
- [10] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- [11] Nancy E Briggs. 1973. Rhetorical dimensions of the nursery rhyme. *Communication Education* 22, 3 (1973), 215–219. <https://doi.org/10.1080/03634527309378018>
- [12] Pablo E Brockmann, Blanca Diaz, Felipe Damiani, Luis Villarreal, Felipe Núñez, and Oliviero Bruni. 2016. Impact of television on the quality of sleep in preschool children. *Sleep medicine* 20 (2016), 140–144. <https://doi.org/10.1016/j.sleep.2015.06.005>
- [13] Kirsty Brown and Hilary Kennedy. 2011. Learning through conversation: Exploring and extending teacher and children's involvement in classroom talk. *School Psychology International* 32, 4 (2011), 377–396. <https://doi.org/10.1177/0143034311406813>
- [14] Jerome Bruner. 1985. Child's talk: Learning to use language. *Child Language Teaching and Therapy* 1, 1 (1985), 111–114. <https://doi.org/10.1177/026565908500100113>
- [15] Peter Bryant, Morag MacLean, and Lynette Bradley. 1990. Rhyme, language, and children's reading. *Applied Psycholinguistics* 11, 3 (1990), 237–252. <https://doi.org/10.1017/S0142716400008870>
- [16] Kate Cain, Jane Oakhill, and Peter Bryant. 2004. Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of educational psychology* 96, 1 (2004), 31–42. <https://api.semanticscholar.org/CorpusID:17531662>
- [17] Colleen E Carney, Daniel J Buysse, Sonia Ancoli-Israel, Jack D Edinger, Andrew D Krystal, Kenneth L Lichstein, and Charles M Morin. 2012. The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep* 35, 2 (2012), 287–302. <https://doi.org/10.5665/sleep.1642>
- [18] Jiahao Chen, Mingming Li, and Jaap Ham. 2024. Different dimensions of anthropomorphic design cues: How visual appearance and conversational style influence users' information disclosure tendency towards chatbots. *International Journal of Human-Computer Studies* 190 (2024), 103320. <https://doi.org/10.1016/j.ijhcs.2024.103320>
- [19] Jiaju Chen, Minglong Tang, Yuxuan Lu, Bingsheng Yao, Elissa Fan, Xiaojuan Ma, Ying Xu, Dakuo Wang, Yuling Sun, and Liang He. 2025. Characterizing LLM-Empowered Personalized Story Reading and Interaction for Children: Insights From Multi-Stakeholder Perspectives. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1002, 24 pages. <https://doi.org/10.1145/3706598.3713275>

- [20] Shanshan Chen, Jun Hu, Hannah Christina van Iterson, Ning Fang, and Panos Markopoulos. 2025. "Did you sleep well?": A Multimodal Sleep Diary for Sustained Self-Reporting by Children (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 1178, 17 pages. <https://doi.org/10.1145/3706598.3713425>
- [21] S. Chen, P. Markopoulos, and J. Hu. 2024. Dozzz: Exploring Voice-Based Sleep Experience Sampling for Children. In *Pervasive Computing Technologies for Healthcare (PervasiveHealth 2023) (Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, Vol. 572)*, D. Salvi, P. Van Gorp, and S. A. Shah (Eds.). Springer, Cham. [https://doi.org/10.1007/978-3-031-59717-6\\_32](https://doi.org/10.1007/978-3-031-59717-6_32)
- [22] Kate Chenier, Rebecca Milne, Andrea Shawyer, and Andy Williams. 2025. Interviewing Indigenous adults reporting historical child sexual abuse: The effect of question types on eliciting descriptive answers and details. *Child Abuse & Neglect* 165 (2025), 107492. <https://doi.org/10.1016/j.chiabu.2025.107492>
- [23] Ronald D Chervin, James E Dillon, Claudio Bassetti, Dara A Ganoczy, and Kenneth J Pituch. 1997. Symptoms of sleep disorders, inattention, and hyperactivity in children. *Sleep* 20, 12 (1997), 1185–1192. <https://doi.org/10.1093/sleep/20.12.1185>
- [24] Jenna H Chin, Seungwook Lee, Mohsena Ashraf, Matt Zago, Yun Xie, Elizabeth A Wolfgram, Tom Yeh, and Pilyoung Kim. 2024. Young Children's Creative Storytelling with ChatGPT vs. Parent: Comparing Interactive Styles. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 379, 7 pages. <https://doi.org/10.1145/3613905.3650770>
- [25] Dasom Choi, SoHyun Park, Kyungah Lee, Hwajung Hong, and Young-Ho Kim. 2025. AACessTalk: Fostering Communication between Minimally Verbal Autistic Children and Parents with Contextual Guidance and Card Recommendation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 556, 25 pages. <https://doi.org/10.1145/3706598.3713792>
- [26] Youjin Choi, JooYeong Kim, Chan Woo Park, Jeongyoun Kim, Ji Hyun Yi, and Jin-Hyuk Hong. 2022. We Play and Learn Rhythmically: Gesture-based Rhythm Game for Children with Intellectual Developmental Disabilities to Learn Manual Sign. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 539, 13 pages. <https://doi.org/10.1145/3491102.3517456>
- [27] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (Glasgow, Scotland UK) (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300705>
- [28] Peter A Cohen. 1980. Effectiveness of student-rating feedback for improving college instruction: A meta-analysis of findings. *Research in higher education* 13, 4 (1980), 321–341. <https://doi.org/10.1007/BF00976252>
- [29] Samuel Rhys Cox and Wei Tsang Ooi. 2022. Does Chatbot Language Formality Affect Users' Self-Disclosure?. In *Proceedings of the 4th Conference on Conversational User Interfaces (Glasgow, United Kingdom) (CUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 1, 13 pages. <https://doi.org/10.1145/3543829.3543831>
- [30] Randy L Diehl, Andrew J Lotto, and Lori L Holt. 2004. Speech perception. *Annu. Rev. Psychol.* 55, 1 (2004), 149–179. <https://doi.org/10.1146/annurev.psych.55.090902.142028>
- [31] Griffin Dietz, Jimmy K Le, Nadin Tamer, Jenny Han, Hyowon Gweon, Elizabeth L Murnane, and James A Landay. 2021. Storycoder: Teaching computational thinking concepts through storytelling in a voice-guided app for children. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–15.
- [32] Griffin Dietz, Zachary Pease, Brenna McNally, and Elizabeth Foss. 2020. Giggle gauge: a self-report instrument for evaluating children's engagement with technology. In *Proceedings of the Interaction Design and Children Conference (London, United Kingdom) (IDC '20)*. Association for Computing Machinery, New York, NY, USA, 614–623. <https://doi.org/10.1145/3392063.3394393>
- [33] Emily Doherty, E. Margaret Perloff, Sean von Bayern, Rui Zhang, Indrani Dey, Michal Bodzianowski, Sadhana Puntambekar, and Leanne Hirshfield. 2025. Piecing Together Teamwork: A Responsible Approach to an LLM-based Educational Jigsaw Agent. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 19, 17 pages. <https://doi.org/10.1145/3706598.3713349>
- [34] Min Fan, Xinyue Cui, Wanqing Ma, Haiyan Li, Xin Tong, Lin Yang, and Yonghui Wang. 2025. From Words to Wonder: Designing and Evaluating an AI-Empowered Creative Storytelling System for Elementary Children. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 328, 15 pages. <https://doi.org/10.1145/3706598.3713478>
- [35] Franz Faul, Edgar Erdfelder, Axel Buchner, and Albert-Georg Lang. 2009. Statistical power analyses using G\* Power 3.1: Tests for correlation and regression analyses. *Behavior research methods* 41, 4 (2009), 1149–1160. <https://doi.org/10.3758/BRM.41.4.1149>
- [36] Jaimie Lee Freeman and Amanda Nicole Curtis. 2023. Putting the Self in Self-Tracking: The Value of a Co-Designed 'How Might You' Self-Tracking Guide for Teenagers. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 340, 16 pages. <https://doi.org/10.1145/3544548.3580938>
- [37] Andrzej Galecki and Tomasz Burzykowski. 2013. *Linear Mixed-Effects Model*. Springer, New York, NY. [https://doi.org/10.1007/978-1-4614-3900-4\\_13](https://doi.org/10.1007/978-1-4614-3900-4_13)
- [38] Radhika Garg, Hua Cui, Spencer Seligson, Bo Zhang, Martin Porcheron, Leigh Clark, Benjamin R. Cowan, and Erin BenetEAU. 2022. The Last Decade of HCI Research on Children and Voice-based Conversational Agents. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (New Orleans, LA, USA) (CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 149, 19 pages. <https://doi.org/10.1145/3491102.3502016>
- [39] Susan E Gathercole, Catherine Willis, and Alan D Baddeley. 1991. Differentiating phonological memory and awareness of rhyme: Reading and vocabulary development in children. *British Journal of Psychology* 82, 3 (1991), 387–406. <https://doi.org/10.1111/j.2044-8295.1991.tb02407.x>
- [40] Jean Berko Gleason and Nan Bernstein Ratner. 2022. *The development of language*. Plural Publishing, 447 pages.
- [41] Herbert Paul Grice. 1975. Logic and conversation. *Syntax and semantics* 3 (1975), 43–58. [https://doi.org/10.1017/Courses/UFRN\\_BR/LEM2013/Resources/H-P\\_Grice/Logic-Conversation.pdf](https://doi.org/10.1017/Courses/UFRN_BR/LEM2013/Resources/H-P_Grice/Logic-Conversation.pdf)
- [42] Charles R. Gruner. 1996. Wit and Humour in Mass Communication. In *Humor and Laughter: Theory, Research and Applications* (2nd ed.), Hugh Foot (Ed.). Routledge, New York, 25–49. <https://doi.org/10.4324/9780203789469>
- [43] Jiyeon Han, Jimin Park, Jinyoung Huh, Uran Oh, Jaeyoung Do, and Daehye Kim. 2024. AscleAI: A LLM-based Clinical Note Management System for Enhancing Clinician Productivity. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 50, 7 pages. <https://doi.org/10.1145/3613905.3650784>
- [44] Jiyeon Han, Jimin Park, Jinyoung Huh, Uran Oh, Jaeyoung Do, and Daehye Kim. 2024. AscleAI: A LLM-based Clinical Note Management System for Enhancing Clinician Productivity. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 50, 7 pages. <https://doi.org/10.1145/3613905.3650784>
- [45] Mariam Hassib, Daniel Buschek, Paweł W. Wozniak, and Florian Alt. 2017. HeartChat: Heart Rate Augmented Mobile Chat to Support Empathy and Awareness. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (Denver, Colorado, USA) (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2239–2251. <https://doi.org/10.1145/3025453.3025758>
- [46] Donald S Hayes, Bruce E Chemeleski, and Melvin Palmer. 1982. Nursery rhymes and prose passages: Preschoolers' liking and short-term retention of story events. *Developmental Psychology* 18, 1 (1982), 49. <https://doi.org/10.1037/0012-1649.18.1.49>
- [47] Kunlei He, Julian Levine, Kelsyann Cervera, Santiago Ojeda-Ramirez, Ying Xu, and Mark Warschauer. 2024. A Home Study of Parent-Child Co-Reading with a Bilingual Conversational Agent. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI EA '24)*. Association for Computing Machinery, New York, NY, USA, Article 17, 8 pages. <https://doi.org/10.1145/3613905.3650836>
- [48] Kenneth Heller, John D Davis, and Roger A Myers. 1966. The effects of interviewer style in a standardized interview. *Journal of Consulting Psychology* 30, 6 (1966), 501–508. <https://doi.org/10.1037/h0024015>
- [49] Alice Sterling Honig. 1988. Humor development in children. *Young children* 43, 4 (1988), 60–73. [https://www.researchgate.net/profile/Alice-Honig/publication/234687359\\_RESEARCH\\_IN\\_REVIEW\\_Humor\\_Development\\_in\\_Young\\_Children/links/550c819a0cf2ac2905a4313a/RESEARCH-IN-REVIEW-Humor-Development-in-Young-Children.pdf](https://www.researchgate.net/profile/Alice-Honig/publication/234687359_RESEARCH_IN_REVIEW_Humor_Development_in_Young_Children/links/550c819a0cf2ac2905a4313a/RESEARCH-IN-REVIEW-Humor-Development-in-Young-Children.pdf)
- [50] Alison Hooper. 2023. Sing and Play Your Way to Reading: Building Emergent Literacy Skills in Infants and Toddlers Through Music. *The Reading Teacher* 76, 4 (2023), 451–458. <https://doi.org/10.1002/trtr.2167>
- [51] Rune Moberg Jacobsen, Samuel Rhys Cox, Carla F. Griggio, and Niels van Berkel. 2025. Chatbots for Data Collection in Surveys: A Comparison of Four Theory-Based Interview Probes. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 228, 21 pages. <https://doi.org/10.1145/3706598.3714128>
- [52] Allison James. 1995. Talking of Children and Youth: Language, Socialization and Culture. In *Youth Cultures: A Cross-Cultural Perspective* (1st ed.), Vered Amit and Helena Wulff (Eds.). Routledge, London, 20–39. <https://doi.org/10.4324/9781003333487>

- [53] Helene Joffe. 2011. Thematic analysis. *Qualitative research methods in mental health and psychotherapy: A guide for students and practitioners* (2011), 209–223. <https://online.library.wiley.com/doi/pdf/10.1002/9781119973249#page=212>
- [54] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing Data from Chatbot and Web Surveys: Effects of Platform and Conversational Style on Survey Response Quality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–12. <https://doi.org/10.1145/3290605.3300316>
- [55] Naja Kathrine Kollerup, Maria-Theresa Bahodi, Samuel Rhys Cox, and Niels van Berkel. 2025. Enhancing Self-Efficacy in Health Self-Examination through Conversational Agent's Encouragement. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 183, 18 pages. <https://doi.org/10.1145/3706598.3713142>
- [56] Landbot.io. 2025. Conversational Chatbot Builder. <https://landbot.io/lp/chatbot-creator>. Retrieved January 19, 2025.
- [57] Lijia Lin, Paul Ginns, Tianhui Wang, and Peilin Zhang. 2020. Using a pedagogical agent to deliver conversational style instruction: What benefits can you obtain? *Computers & Education* 143 (2020), 103658. <https://doi.org/10.1016/j.compedu.2019.103658>
- [58] Yujia Liu, Siyu Zha, Yuewen Zhang, Yanjin Wang, Yangming Zhang, Qi Xin, Lun Yiu Nie, Chao Zhang, and Yingqing Xu. 2025. BrickSmart: Leveraging Generative AI to Support Children's Spatial Language Learning in Family Block Play. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 645, 19 pages. <https://doi.org/10.1145/3706598.3714212>
- [59] Priscilla Y. Lo, Annemiek Veldhuis, Alissa N. Antle, and Steve DiPaola. 2025. Noel: A Chatbot Persona to Support Children Designing for Others. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 133, 25 pages. <https://doi.org/10.1145/3706598.3713836>
- [60] Steson Lo and Sally Andrews. 2015. To transform or not to transform: Using generalized linear mixed models to analyse reaction time data. *Frontiers in psychology* 6 (2015), 1171. <https://doi.org/10.3389/fpsyg.2015.01171>
- [61] Nichola Lubold. 2017. Building Rapport through Dynamic Models of Acoustic-Prosodic Entrainment. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI EA '17). Association for Computing Machinery, New York, NY, USA, 297–300. <https://doi.org/10.1145/3027063.3027132>
- [62] Panos Markopoulos, Janet C Read, Stuart MacFarlane, and Johanna Hoynsiemi. 2008. *Evaluating children's interactive products: principles and practices for interaction designers*. Elsevier, 400 pages. <https://doi.org/10.1016/B978-0-12-374111-0.X0001-1>
- [63] Michael McCarthy. 2010. Spoken fluency revisited. *English profile journal* 1 (2010), e4. <https://doi.org/doi:10.1017/S2041536210000012>
- [64] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica* 22, 3 (2012), 276–282. <https://hrcak.srce.hr/89395>
- [65] Graham McNeill, Max Sondag, Stewart Powell, Phoebe Asplin, Gagatay Turkey, Faron Moller, and Daniel Archambault. 2023. From Asymptomatics to Zombies: Visualization-Based Education of Disease Modeling for Children. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 839, 17 pages. <https://doi.org/10.1145/3544548.3581573>
- [66] Magbule Mejzini. 2016. Teaching children through songs, chants and rhymes. *European Journal of English Language Teaching* 1 (2016), 104–114. Issue 2. <https://doi.org/10.46827/ejel.v0i0.374>
- [67] Monica Melby-Lervåg. 2012. The relative predictive contribution and causal role of phoneme awareness, rhyme awareness, and verbal short-term memory in reading skills: A review. *Scandinavian Journal of Educational Research* 56, 1 (2012), 101–118. <https://doi.org/10.1080/00313831.2011.621215>
- [68] Brooke Morris, Hayati Havlucu, Alison Oldfield, and Oussama Metatla. 2025. "It Helps Us Express Our Feelings Without Having To Say Anything": Exploring 'Accompanying Social Play Things' Designed With and For Neurodiverse Groups of Children. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 780, 21 pages. <https://doi.org/10.1145/3706598.3713738>
- [69] Michelle M Neumann, Leigh-Chantelle Koch, Jason Zagami, David Reilly, and David L Neumann. 2023. Preschool children's engagement with a social robot compared to a human instructor. *Early Childhood Research Quarterly* 65 (2023), 332–341. <https://doi.org/10.1016/j.ecresq.2023.07.010>
- [70] Christian Obermeier, Winfried Menninghaus, Martin Von Koppenfels, Tim Raettig, Maren Schmidt-Kassow, Sascha Otterbein, and Sonja A Kotz. 2013. Aesthetic and emotional effects of meter and rhyme in poetry. *Frontiers in psychology* 4 (2013), 10. <https://doi.org/10.3389/fpsyg.2013.00010>
- [71] Hyo-Jung Oh, Chung-Hee Lee, Hyeon-Jin Kim, and Myung-Gil Jang. 2005. Descriptive question answering in encyclopedia. In *Proceedings of the ACL 2005 on Interactive Poster and Demonstration Sessions* (Ann Arbor, Michigan) (ACLdemo '05). Association for Computational Linguistics, USA, 21–24. <https://doi.org/10.3115/1225753.1225759>
- [72] Barbara A. Pan and Catherine E. Snow. 1999. The Development of Conversational and Discourse Skills. In *The Development of Language* (1st ed.), Martyn Barrett (Ed.). Psychology Press, London, 21–49. <https://doi.org/10.4324/9781315784694>
- [73] Gain Park, Myungok Chris Yim, Jiyun Chung, and Seyoung Lee. 2023. Effect of AI chatbot empathy and identity disclosure on willingness to donate: the mediation of humanness and social presence. *Behaviour & Information Technology* 42, 12 (2023), 1998–2010. <https://doi.org/10.1080/0144929X.2022.2105746>
- [74] Delroy L Paulhus, Simine Vazire, et al. 2007. The self-report method. In *Handbook of research methods in personality psychology* (1st ed.), R. Chris Fraley PhD Richard W. Robins PhD (Ed.). Guilford, Chapter 13, 224–239. [https://books.google.nl/books?hl=en&lr=&id=XHwS3PU6uroC&oi=fnd&pg=PA224&dq=The+self-report+method&ots=JEOdRZVUOC&sig=JEGUrY8nhFFR53JrRtCF8YhOXlg&redir\\_esc=y#v=onepage&q=The%20self-report%20method&f=false](https://books.google.nl/books?hl=en&lr=&id=XHwS3PU6uroC&oi=fnd&pg=PA224&dq=The+self-report+method&ots=JEOdRZVUOC&sig=JEGUrY8nhFFR53JrRtCF8YhOXlg&redir_esc=y#v=onepage&q=The%20self-report%20method&f=false)
- [75] Kamila Polišíenská, Shula Chiat, and Penny Roy. 2015. Sentence repetition: what does the task measure? *International Journal of Language & Communication Disorders* 50, 1 (2015), 106–118. <https://doi.org/10.1111/1460-6984.12126>
- [76] Stephen R Porter, Michael E Whitcomb, and William H Weitzer. 2004. Multiple surveys of students and survey fatigue. *New directions for institutional research* 2004, 121 (2004), 63–73. <https://doi.org/10.1002/ir.101>
- [77] Janet C Read, Stuart MacFarlane, and Chris Casey. 2002. Endurability, engagement and expectations: Measuring children's fun. In *Interaction design and children*, Vol. 2. Shaker Publishing Eindhoven, 1–23. [https://www.researchgate.net/profile/Janet-Read/publication/228870976\\_Endurability\\_Engagement\\_and\\_Expectations\\_Measuring\\_Children's\\_Fun/links/0deec518618d0828ce00000/Endurability-Engagement-and-Expectations-Measuring-Children-aposs-Fun.pdf](https://www.researchgate.net/profile/Janet-Read/publication/228870976_Endurability_Engagement_and_Expectations_Measuring_Children's_Fun/links/0deec518618d0828ce00000/Endurability-Engagement-and-Expectations-Measuring-Children-aposs-Fun.pdf)
- [78] Anne W Riley. 2004. Evidence that school-age children can self-report on their health. *Ambulatory Pediatrics* 4, 4 (2004), 371–376. <https://doi.org/10.1367/A03-178R.1>
- [79] Rajat Roy and Vik Naidoo. 2021. Enhancing chatbot effectiveness: The role of anthropomorphic conversational styles and time orientation. *Journal of Business Research* 126 (2021), 23–34. <https://doi.org/10.1016/j.jbusres.2020.12.051>
- [80] Brian Scassellati, Jake Brawer, Katherine Tsui, Setareh Nasihati Gilani, Melissa Malzkunh, Barbara Manini, Adam Stone, Geo Kartheiser, Arcangelo Merla, Ari Shapiro, David Traum, and Laura-Ann Petito. 2018. Teaching Language to Deaf Infants with a Robot and a Virtual Human. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) (CHI '18). Association for Computing Machinery, New York, NY, USA, 1–13. <https://doi.org/10.1145/3173574.3174127>
- [81] Klaus R Scherer. 2013. Vocal markers of emotion: Comparing induction and eliciting activation. *Computer Speech & Language* 27, 1 (2013), 40–58. <https://doi.org/10.1016/j.csl.2011.11.003>
- [82] Woosuk Seo, Young-Ho Kim, Ji Eun Kim, Megan Tao Fan, Mark S. Ackerman, Sung Won Choi, and Sun Young Park. 2025. Enhancing Pediatric Communication: The Role of an AI-Driven Chatbot in Facilitating Child-Parent-Provider Interaction. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 1079, 16 pages. <https://doi.org/10.1145/3706598.3713134>
- [83] Woosuk Seo, Chanmo Yang, and Young-Ho Kim. 2024. ChaCha: Leveraging Large Language Models to Prompt Children to Share Their Emotions about Personal Events. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 903, 20 pages. <https://doi.org/10.1145/3613904.3642152>
- [84] Mohammad Shaharyar Shaikat, Mohammed Tanzeem, Tameem Ahmad, and Nesari Ahmad. 2021. Semantic similarity-based descriptive answer evaluation. In *Web semantics*. Elsevier, 221–231. <https://doi.org/10.1016/B978-0-12-822468-7.00014-6>
- [85] Jocelyn J Shen, Jennifer King Chen, Leah Findlater, and Griffin Dietz Smith. 2025. eaSEL: Promoting Social-Emotional Learning and Parent-Child Interaction through AI-Mediated Content Consumption. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 1082, 18 pages. <https://doi.org/10.1145/3706598.3713405>
- [86] Jocelyn J Shen, Jennifer King Chen, Leah Findlater, and Griffin Dietz Smith. 2025. eaSEL: Promoting Social-Emotional Learning and Parent-Child Interaction through AI-Mediated Content Consumption. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (CHI '25). Association for Computing Machinery, New York, NY, USA, Article 1082, 18 pages. <https://doi.org/10.1145/3706598.3713405>
- [87] Yan Shi, Lidan Gong, Yiwen Lu, Lijuan Liu, Chao Zhang, Shujun Zhang, Longfei Wang, and Shan Zhou. 2025. "I Need Your Help!": Facilitating Psychological Communication Between Left-Behind Children and Their Parents with an AI-Powered Sandbox. In *Proceedings of the 2025 CHI Conference on Human Factors*

- in *Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 272, 19 pages. <https://doi.org/10.1145/3706598.3713660>
- [88] Dong-Hee Shin and Kyung-mi Chung. 2017. The effects of input modality and story-based knowledge on users' game experience. *Computers in Human Behavior* 68 (2017), 180–189. <https://doi.org/10.1016/j.chb.2016.11.030>
- [89] Lucas M. Silva, Franceli L. Cibrian, Clarisse Bonang, Arpita Bhattacharya, Aehong Min, Elissa M Monteiro, Jesus Armando Beltran, Sabrina Schuck, Kimberley D Lakes, Gillian R. Hayes, and Daniel A. Epstein. 2024. Co-Designing Situated Displays for Family Co-Regulation with ADHD Children. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 124, 19 pages. <https://doi.org/10.1145/3613904.3642745>
- [90] Anastasia Smirnova, Kyu beom Chun, Wil Louis Rothman, and Siyona Sarma. 2025. Text Simplification for Children: Evaluating LLMs vis-à-vis Human Experts. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 512, 10 pages. <https://doi.org/10.1145/3706599.3719889>
- [91] Steven A Stahl and Bruce A Murray. 1994. Defining phonological awareness and its relationship to early reading. *Journal of Educational Psychology* 86, 2 (1994), 221. <https://doi.org/10.1037/0022-0663.86.2.221>
- [92] Mark A Stein, Janis Mendelsohn, William H Obermeyer, Julie Amromin, and Ruth Benca. 2001. Sleep and behavior problems in school-aged children. *Pediatrics* 107, 4 (2001), e60–e60. <https://doi.org/10.1542/peds.107.4.e60>
- [93] Abram Sterne and Usha Goswami. 2000. Phonological awareness of syllables, rhymes, and phonemes in deaf children. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 41, 5 (2000), 609–625. <https://doi.org/doi:10.1111/1469-7610.00648>
- [94] Jennifer Sturgess, Sylvia Rodger, and Anne Ozanne. 2002. A review of the use of self-report assessment with young children. *British Journal of Occupational Therapy* 65, 3 (2002), 108–116. <https://doi.org/10.1177/030802260206500302>
- [95] John Sweller. 2011. Cognitive load theory. In *Psychology of learning and motivation*. Vol. 55. Elsevier, 37–76. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>
- [96] Kevin WH Tai and Li Wei. 2021. Constructing playful talk through translanguaging in English medium instruction mathematics classrooms. *Applied Linguistics* 42, 4 (2021), 607–640. <https://doi.org/10.1093/applin/amaa043>
- [97] Yilin Tang, Liuqing Chen, Ziyu Chen, Wenkai Chen, Yu Cai, Yao Du, Fan Yang, and Lingyun Sun. 2024. EmoEden: Applying Generative Artificial Intelligence to Emotional Learning for Children with High-Function Autism. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 1001, 20 pages. <https://doi.org/10.1145/3613904.3642899>
- [98] Deborah Tannen. 2005. *Conversational style: Analyzing talk among friends*. Oxford University Press, 272 pages.
- [99] V Ya Tsvetkov. 2009. Information objects and information Units. *European Journal of Natural History* 5, 2 (2009), 99. <https://s.world-science.ru/pdf/2009/2/34.pdf>
- [100] H. C. van Iterson, P. Markopoulos, and L. E. Gerritsen. 2023. Motivating Children's Engagement with Sleep Diaries Through Storytelling. In *Human-Computer Interaction – INTERACT 2023 (Lecture Notes in Computer Science, Vol. 14143)*, J. Abdelnour Nocera, M. Kristin Lárusdóttir, H. Petrie, A. Piccinno, and M. Winckler (Eds.). Springer, Cham. [https://doi.org/10.1007/978-3-031-42283-6\\_12](https://doi.org/10.1007/978-3-031-42283-6_12)
- [101] Sruthi Viswanathan, Seray Ibrahim, Ravi Shankar, Reuben Binns, Max Van Kleek, and Petr Slovak. 2025. The Interaction Layer: An Exploration for Co-Designing User-LLM Interactions in Parental Wellbeing Support Systems. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 310, 25 pages. <https://doi.org/10.1145/3706598.3714088>
- [102] Leonie VDE Vogelsmeier, Eduardo Oliveira, Kamila Misiejuk, Sonsoles López-Pernas, and Mohammed Saqr. 2025. Delving Into the Psychology of Machines: Exploring the Structure of Self-Regulated Learning via LLM-Generated Survey Responses. *Computers in Human Behavior* 173 (2025), 108769.
- [103] Tudor Văcărețu, Nikolaos Batalas, Begum Erten-Uyumaz, Merel van Gilst, Sebastiaan Overeem, and Panos Markopoulos. 2019. Subjective Sleep Quality Monitoring with the Hypnos Digital Sleep Diary: Evaluation of Usability and User Experience. In *Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSSTEC 2019) - HEALTHINF. INSTICC, SciTePress*, 113–122. <https://doi.org/10.5220/0007407601130122>
- [104] Zizhen Wang, Jiangyu Pan, Duola Jin, Jingao Zhang, Jiacheng Cao, Chao Zhang, Zejian Li, Preben Hansen, Yijun Zhao, Shouqian Sun, and Xianyue Qiao. 2025. CharacterCritique: Supporting Children's Development of Critical Thinking through Multi-Agent Interaction in Story Reading. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 131, 21 pages. <https://doi.org/10.1145/3706598.3713602>
- [105] Myrna M Weissman, Helen Orvaschel, and Nancy Padian. 1980. Children's symptom and social functioning self-report scales comparison of mothers' and children's reports. *The Journal of nervous and mental disease* 168, 12 (1980), 736–740. <https://doi.org/10.1097/00005053-198012000-00005>
- [106] Naomi E Winstone, Robert A Nash, Michael Parker, and James Rowntree. 2017. Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of reciprocity processes. *Educational psychologist* 52, 1 (2017), 17–37. <https://doi.org/https://doi.org/10.1080/00461520.2016.1207538>
- [107] Christine Wu, Ila Krishna Kumar, and Rosalind Picard. 2025. Navigating Emotions Through Art: Recommendations for Designing Art-Therapy Based Chatbots for Trauma-Impacted Youth: Recommendations for Designing Art-Therapy Based Chatbots for Trauma-Impacted Youth. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25)*. Association for Computing Machinery, New York, NY, USA, Article 413, 7 pages. <https://doi.org/10.1145/3706599.3719869>
- [108] Ziang Xiao, Tiffany Wenting Li, Karrie Karahalios, and Hari Sundaram. 2023. Inform the Uninformed: Improving Online Informed Consent Reading with an AI-Powered Chatbot. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 112, 17 pages. <https://doi.org/10.1145/3544548.3581252>
- [109] Ziang Xiao, Michelle X. Zhou, Q. Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2020. Tell Me About Yourself: Using an AI-Powered Chatbot to Conduct Conversational Surveys with Open-ended Questions. *ACM Trans. Comput.-Hum. Interact.* 27, 3, Article 15 (June 2020), 37 pages. <https://doi.org/10.1145/3381804>
- [110] Wenjie Xu, Jiayi Ma, Jiayu Yao, Weijia Lin, Chao Zhang, Xuanhe Xia, Nan Zhuang, Shitong Weng, Xiaoqian Xie, Shuyue Feng, Fangtian Ying, Preben Hansen, and Cheng Yao. 2023. MathKingdom: Teaching Children Mathematical Language Through Speaking at Home via a Voice-Guided Game. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 93, 14 pages. <https://doi.org/10.1145/3544548.3581043>
- [111] Ying Xu, Stacy Branham, Xinwei Deng, Penelope Collins, and Mark Warschauer. 2021. Are Current Voice Interfaces Designed to Support Children's Language Development?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 633, 12 pages. <https://doi.org/10.1145/3411764.3445271>
- [112] Ying Xu and Mark Warschauer. 2020. Exploring young children's engagement in joint reading with a conversational agent. In *Proceedings of the Interaction Design and Children Conference* (London, United Kingdom) (IDC '20). Association for Computing Machinery, New York, NY, USA, 216–228. <https://doi.org/10.1145/3392063.3394417>
- [113] Shengchen Yin, Dena Kasraian, Gubing Wang, Suzan Evers, and Pieter van Wesemael. 2024. Co-designing an ideal nature-related digital tool with children: An exploratory study from the Netherlands. *Environment and Behavior* 56, 9-10 (2024), 739–775. <https://doi.org/10.1177/001391652413110>
- [114] Chao Zhang, Xuechen Liu, Katherine Ziska, Soobin Jeon, Chi-Lin Yu, and Ying Xu. 2024. Mathemyths: Leveraging Large Language Models to Teach Mathematical Language through Child-AI Co-Creative Storytelling. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 274, 23 pages. <https://doi.org/10.1145/3613904.3642647>
- [115] Xi Zheng, Zhuoyang Li, Xinning Gui, and Yuhua Luo. 2025. Customizing Emotional Support: How Do Individuals Construct and Interact With LLM-Powered Chatbots. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 376, 20 pages. <https://doi.org/10.1145/3706598.3713453>
- [116] Yuehong Zhong and Karen Guo. 2023. 'They can do many things': rhymes as a pedagogical tool. *The Australian Educational Researcher* 50, 5 (2023), 1439–1455. <https://doi.org/10.1007/s13384-022-00562-z>

## A Prompt in Rhyming Style

<Intro Task>

- You are a friendly and engaging assistant conducting a sleep survey for children aged six to thirteen.
- Use a conversational style for most interactions and incorporate rhyming, poetic responses **only** for reinforcement, summaries, or emotional support.
- Ensure rhyming responses are concise, simple, and limited to two or three short sentences.

- Use language appropriate for a 6 to 13-year-old child, avoiding complex words and phrases.

**\*\*Important Instructions:\*\***

- **\*\*Do not ask any questions.\*\*** The app will handle asking the predefined questions.
- Focus on providing acknowledgments and emotional support.
- Do **\*\*not\*\*** include the child's response in your acknowledgment.
- Do **\*\*not\*\*** answer for the child.
- Only write the assistant's part of the conversation.
- **\*\*Keep all responses short, simple, and easy to understand.\*\***

**\*\*Example Interactions:\*\***

\*Assistant:\* Hi there! Let's chat about your sleep and playtime!

\*Child:\* I felt really good today.

\*Assistant:\* That's great to hear!

\*Child:\* I had a bad dream.

\*Assistant:\* I'm sorry to hear that. Bad dreams can be scary.

\*Child:\* But I feel better now.

\*Assistant:\* Glad you're feeling better, my friend. Happy thoughts from here on end!

\*Child:\* I played a lot today.

\*Assistant:\* Playing is so much fun!

\*Child:\* I'm sleepy.

\*Assistant:\* Sleep tight, little one. Sweet dreams!

Only proceed to the next question after the user has answered the current one. Remember to keep your acknowledgments and questions rhythmic, fun, and engaging for a child.

- How are you feeling today?
- How long did it take you to fall asleep?
- Did you sleep well?
- What did you do before sleeping last night?
- What did you eat before sleeping last night?
- What did you drink before sleeping last night?
- How many times did you wake up last night?
- When did you sleep last night?
- How long did you nap or doze this daytime?
- How long did you sleep last night?
- When did you wake up this morning?
- How did you wake up this morning?

## B Script of Information Units

```
import spacy
#load the English NLP model
nlp = spacy.load("en_core_web_sm")

#Function to count information units
def count_information_units(sentence):
    #parse the sentence using spacy
    doc = nlp(sentence)
```

```
#Initialize count
unit_count = 0

#Check for subjects (nominal subjects)
subjects = len([token for token in doc if token.dep_
in ['nsubj', 'nsubjpass']])
unit_count += subjects

#Check for verbs (root of the clause)
verbs = len([token for token in doc if token.dep_ ==
'ROOT' and token.pos_ == 'VERB'])
unit_count += verbs

#Check for objects (direct, indirect, and prepositional
objects)
objects = len([token for token in doc if token.dep_ in
['dobj', 'iobj', 'pobj']])
unit_count +=objects

#rReturn the tootal count of subjects, verbs, and
objects as information units
return unit_count
```