Dozzz: Exploring the Feasibility of a Voice-Based Sleep Diary for Children

Shanshan Chen
Department of Industrial Design
Eindhoven University of Technology
s.chen1@tue.nl

Panos Markopoulos
Department of Industrial Design
Eindhoven University of Technology
p.markopoulos@tue.nl

Jun Hu
Department of Industrial Design
Eindhoven University of Technology
j.hu@tue.nl

Sleep problems can impact many children's wellbeing and development adversely. Treating clinicians often rely on parents to record their child's sleep behaviors and experiences in specialized diaries. Earlier studies have aimed at empowering children to self-report their sleep experiences independently, thereby providing direct communication with clinicians and reducing parental bias. Digital diaries based on text entry or rating scales, may not be suitable for young children. We developed Dozzz, a voice-based sleep diary that enables children to interact with chatbot avatar to report their sleep experiences. In a within-subject experiment involving 35 children aged eight to twelve, Dozzz was compared to a text-based sleep diary. The results indicate that Dozzz enhances user experience and engagement, demonstrating the potential of a voice-based diary for children. Future research should examine sustained and independent use in daily life.

Voice-user interfaces, conversational interfaces, children, sleep diary, self-report

1. INTRODUCTION

Sleep disorders affect many children and adolescents, with a prevalence rate up to 50% (Danielle Pacheco (2023)). They can be common and chronic, and they extend beyond nighttime disruptions, significantly influencing children's daytime functioning, development, and overall well-being (Sivertsen et al. (2009)). To capture the nature and progression of these problems, as well as environmental influences, healthcare specialists such as pediatricians, psychiatrists, and psychologists use sleep diaries. These diaries are designed to survey to record the child's sleep-related behaviors and patterns, including bedtimes, sleep routines related to going to sleep and waking up (Wiggs and Stores (1996)), and parasomnias related to watching television or media usage (Blader et al. (1997); Brockmann et al. (2016)). They also cover nocturnal behaviors involving night fears, night walking problems (Wiggs and Stores (1996); Chervin et al. (1997)), and daytime behaviors influenced by sleep patterns (Stein et al. (2001)). Next to reporting on behaviors connected to sleep, these diaries also involve a subjective evaluation of each night's sleep (Carney et al. (2012)). Thus, complementing objective methods like actigraphy and polysomnography (van Rijssen et al. (2023)), sleep diaries offer a cost-effective way to gain a comprehensive understanding of children's subjective sleep profiles. Commonly, parents are tasked with recording the sleep details of younger children in these diaries. However, previous research has revealed low correlations between parental reports and children's own accounts regarding their behaviors and emotions (Henry (2000)). This underscores the unique and irreplaceable value of insights derived from children's self-reports, emphasizing that they should not be replaced by proxy reports from parents (Riley (2004)).

Despite the benefits of sleep diaries to clinicians and children, along with evidence suggesting that young children can effectively communicate various somatic and psychiatric symptoms, the currently available sleep diaries do not meet all children's needs. Existing sleep diaries, Carney et al. (2012) and graphical diaries (Văcărețu et al. (2019)), are primarily designed for adults, and do not meet the needs of children as users. Children may be hindered by the phrasing of questions, their ability to report experiences using numeric scale and their limited motivation and ability to comply to a repetitive self-reporting routine. Consequently, the underdiagnosis of sleep disorders (Chervin et al. (2003); Lewandowski et al. (2011)) is exacerbated, as the gradual onset of some sleep disturbances often goes unnoticed for extended periods without access to a suitable sleep diary. Recent research in the field of child-computer interaction has examined ways to enable children to self-report independently,

© The Authors. Published by BCS Learning and Development Ltd.
Proceedings of . . .

so as to avoid biases of parental report, and allow children to communicate directly to clinicians about their sleep (Aarts et al. (2022); van Iterson et al. (2023)). Researchers have proposed digital diary applications relying on text-entry or rating scales which may not be suited for young with limited typing and writing skills and who may find it difficult to quantify evaluative statements or their agreement along a numeric scale, as is done in many questionnaires (Markopoulos et al. (2021)). Arguably developing self-report instruments suitable for children, can provide an important contribution to understanding and treating children's sleep problems.

We set out to develop methods that put a limited cognitive burden (Carney et al. (2012)), while improving the user experience (Pina et al. (2020)), to lower the threshold for children's active participation in self-reporting with sleep diaries. We argue that voice-based technologies provide potential to reduce the barrier of interaction compared with traditional text-based sleep diaries that involve writing or typing. Voice-based technologies may offer a faster and more efficient means of interaction, providing special benefits for children with limited typing and writing skills. Furthermore, since much of everyday interactions with adults is conducted orally (Brown and Kennedy (2011)), leveraging conversational and voice-based modes of interaction could lay the groundwork for a more accessible and child-friendly design. Such an approach could lead to more effective and sustained engagement from children in self-reporting their sleep patterns (Winstone et al. (2017)).

We arguably self-report on sleep experiences and sleep related behaviors can be made accessible to younger children in the form of a spoken conversation, where children are asked to comment on their sleep the night before. We expect that children will be more likely to engage in self-tracking activities in a casual and friendly conversation compared to responding to questions using a form-filling sleep diary as is done with diaries for adults (Tudor Car et al. (2020)). A voice-based user interface can alleviate the reading and editing requirements of current sleep diary solutions, thus lowering the threshold for self-report, potentially supporting a sustained use of the diary (Lewandowski et al. (2011)). With a voice-based conversational agent that runs on accessible hardware (smartphones), voice-based sleep diaries can reach a far greater number of children. This approach will facilitate access to evaluate children's sleep quality and engage them for a sustained period (Lewandowski et al. (2011)).

To deepen understanding the voice-user interfaces (VUIs) in sleep diaries for children engaging, we developed Dozzz, a voice-based chatbot designed as a sleep diary, incorporating Speech-to-Text and Text-to-Speech APIs to facilitate voice interactions. We conducted a within-subject experiment with 35 children aged eight to twelve. Each child interacted with both types of sleep diaries: the voice-based Dozzz and the text-based Snoozy (Aarts et al. (2022)), aiming for a comprehensive understanding of children's experience and preferences in relation to different sleep diaries. The outcomes of the evaluation demonstrate difference of children's performance and engagement in two sleep diaries. Notably, children exhibited a preference for the voice version, although certain advantages were observed with the text alternative. This work contributes to the field in the following ways:

- Innovative approach to sleep assessment for children: The implementation of a voice-based sleep diary represent an innovative method in capturing and understanding children's sleep patterns. It provides valuable insights into the potential for engaging children consistently in sleep assessment.
- Comparative analysis of voice and text input modalities: This work provides an in-depth understanding of children's experience regarding sleep diaries by comparing the voice and text versions. This comparative analysis adds a nuanced layer to our understanding of how different interaction modalities impact children's engagement and preferences.

2. RELATED WORK

In this section, our work provides the overview of sleep assessments for children and reviews prior studies on the use of voice-user interfaces (VUIs) to enhance child engagement.

2.1. Sleep Assessment and Sleep Diary for Children

Healthcare specialists primarily employ two methods to assess children's sleep: a) objective measurements, including actigraphy and polysomnography, and b) subjective measurements, such as sleep diaries and questionnaires. Polysomnography is considered the "gold standard" for objectively measuring sleep parameters (Mazza et al. 2020), which typically necessitates specialized medical equipment and a collaborative team effort. Compared with polysomnography, actigraphy presents a more cost-effective and less time-intensive alternative, and not requiring the extensive resources or infrastructure. It employs a non-invasive, wrist-worn device that uses movement detection to deduce sleep and wake cycles [1]. However, challenges include the lack of subjective parameters and discomfort associated with prolonged device wear. The latter impacts children's willingness to be engaged (Schaefer et al. 2014). Another category, the subjective report, relies on individuals' subjective recollection of sleep experiences through sleep diaries and questionnaires. These methods, unlike actigraphy that focuses on motor activity, are subject to cognitive biases, related to memory, time estimation, and motivation to recall sleep parameters (Short et al. 2012). Despite these concerns, numerous studies have demonstrated the relatively high accuracy of subjective sleep parameters (Mazza et al. 2020). Therefore, it should be emphasized that objective measurements for children's sleep do not replace subjective measures and are not equivalent to them, providing a valuable and accessible perspective on sleep experiences (Mazza et al. 2020; Mallinson et al. 2019; Short et al. 2012).

When considering subjective measurements for assessing children's sleep, sleep diaries are valuable tools compared to sleep questionnaires that are commonly used in sleep centers. Unlike questionnaires that rely on patients' recollections over one week or a month, sleep diaries provide a daily record of a patient's subjective experiences, which can help improve the accuracy of self-report.

Most existing digital sleep diaries (Văcărețu et al. (2019); Carney et al. (2012)) are primarily designed for adults. Consequently, clinicians typically rely on proxy reports from adults for continuous daily completion. Adults, such as parents, are reliable informants on concrete and unambiguous issues. However, they may not always be aware of their child's thoughts and emotions (Ivens and Rehm (1988); Barrett et al. (1991); Bird et al. (1992); Loeber et al. (1989)). In contrast, children's self-reports in sleep diaries are argued to be more valid due to their active engagement in their own experiences (Sturgess et al. (2002); Riley (2004)).

Digital diaries are increasingly favored for children's self-reporting, addressing issues such as the "parking lot syndrome" — the tendency to retrospectively complete several days' entries at once, a common problem with paper diaries (Ibáñez et al. 2018). To facilitate the easy self-reporting and motivate children's engagement, researchers have explored feasibility of sleep diary through storytelling (van Iterson et al. (2023)), as well as chatbot (Aarts et al. (2022)). However, participants in the evaluation of the chatbot (Aarts et al. (2022)), expressed a preference a voice-based version, suggesting it would be more engaging. Addressing this gap, Dozzz (Chen et al. (2023)) was designed with text-to-speech and speech-to-text APIs to facilitate a voice conversational interface, demonstrating the potential feasibility of voice-based sleep diaries in a small group of children. Despite these advancements, it remains unclear how voice input modality affects the feasibility of children's self-reporting in sleep diaries. Concerns have also been raised about the redundancy of certain interactive actions that do not involve voice interaction, which could hinder the accessibility of the diaries. Additionally, there have been suggestions for improvements in the content of conversations. To address these issues, we propose further enhancements in interaction design and a deeper exploration of user experience through technological and systemic refinements.

2.2. Voice-user Interfaces (VUIs) for Children

Voice-user interfaces (VUIs) offer distinct advantages over text-based methods when catering to the needs of primary school-aged children (Garg and Sengupta (2020)). Voice entry offers a time-saving alternative, particularly beneficial for those with limited typing or spelling skills, lowering potential barriers associated with text-based interfaces (Issenman and Jaffer (2004)). Previous research has illustrated that voice input can make simplify tasks for children and improve the accessibility. For instance, Xu and Warschauer demonstrated that conversations through voice input can lead to more cognitively and linguistically beneficial interactions for children (Xu and Warschauer (2020)). Similarly, TurtleTalk, which employs voice interaction, enabled children to become more immersed in the game and understand programming elements with greater ease and confidence (Jung et al. (2019)). Hence, we suppose that VUIs emerge as an attractive and user-friendly option for self-reporting children's sleep diaries.

Engaging in playful conversations not only aligns with children's interests and fosters an enjoyable experience (Issenman and Jaffer (2004)), but also has the potential to enhance engagement in dialogue and activities. This increased engagement can lead to children asking more questions and solving more problems (Tewari and Canny (2014)), maintaining emotional engagement with open-ended questions posed by agents (Xu and Warschauer (2020)), and participating in more learning activities such as language practice (Chervin et al. (1997); Mazza et al. (2020)) and reading (Xu and Warschauer (2020)). Additionally, voice input can aid children with diagnosed language impairments, leading to improved performance (Spitale et al.

[1] Actigraphy. `https://stanfordhealthcare.org/medical-tests/s/sleep-disorder-tests/procedures/actigraphy.html`. Accessed: 2023-12.

(2020)). Despite these benefits, there is insufficient evidence to suggest that children are more willing to engage in self-reporting through voice input compared with text input.

This study aims to enhance the design of the voice-based sleep diary for children, known as Dozzz, and to evaluate its feasibility. By comparing Dozzz with the traditional text-based digital sleep diary–Snoozy, we intend to uncover the potential benefits and challenges of implementing VUIs in children's sleep diaries. Our research is guided by the following questions:

1. Can Dozzz accurately recognize children's voice? (Transcription accuracy)
2. How do children interact with Dozzz during conversation? (Turn-taking analysis)
3. What is the user experience of children using voice-based sleep diaries?(Attitudes)

3. EXPERIMENT MATERIALS

3.1. The Design of Dozzz

The content of Dozzz. Based on Snoozy that special design for children (Aarts et al. (2022)) (see Table 1), Dozzz includes about ten questions on behavoirs that related to children's sleep patterns. These questions were based on the Consensus Sleep Diary (Carney et al. (2012)), a standardizing prospective sleep self-monitoring, and tailored for children. This format facilitates gathering sleep data through daily dialogues with children.

Table 1: *The question list in the sleep diary*

Questions
Did you sleep well last night?
Do you know why you sleep well?
How are you feeling today?
Did you wake up for a few times last night?
Did you watch a lot of tv or play video games yesterday?
Are you going to do something nice and exciting today?
What did you eat and drink before sleeping last night?
Where do you need to go today?
What have you planned for today?
What is something you like to do?

The Implementation of Voice-based Interface. To improve the smooth and naturalness of conversation for young children, Dozzz was improved by using the Google Cloud Speech-to-Text V1 API instead of the native voice recognizer on Android platform in our previous version (Chen et al. (2023)). This API facilitates a more natural conversation flow with the streaming recognition (gRPC) for real-time recognition purpose, enabling seamless switching between multi-threaded tasks.

Considering the limitations of the sleep diary's simplified and static questions, Dozzz improved with rule-based logic, enhanced by regular expressions (Regex), to discern keywords from children's input. This simpler alternative to complex machine learning algorithms involves a database of keywords, sorted into positive and negative sentiment groups. When a child speaks, the words are converted into text and matched with these groups to facilitate relevant and empathetic responses. For example, if a child answers positively to a question, Dozzz follows up with an appropriate question from the positive group, like "Good to hear! How are you feeling today?" This method, centered on semantic keyword classification, enables personalized and engaging dialogues that resonate with each child's unique sleep experience. The children's responses are seamlessly synchronized and stored as text in Firebase, as illustrated in Figure 1.

In the interface, an animated character, Dozzz, who is a duck. children's responses appear at the bottom of the interface, adding an interactive and visually engaging element to the experience.

Conversation Timestamps. To enhance our understanding of the reliability of voice-based conversations in Dozzz for young children, a timestamp mechanism is implemented to record key moments in the interaction. These timestamps include:

- Starting point of question: This timestamp marks when a new question begins.
- Ending point of question: This marks the end of a posed question.
- Starting point of answer: This indicates when a child begins the response.
- Ending point of answer: This marks the completion of the child's answer.

As shown in Figure 2, 'wait time' is the interval between the end of a question and the beginning of the child's response. The agent's questions are marked in yellow, and the child's responses in blue. The start and end of a question are indicated by black 1 and black 2, respectively, while the start and end of an answer are marked by purple 1 and purple 2. The wait time is calculated from black 2 to purple 1. This timing is tracked using a webhook fulfillment in Firebase, aiding in analyzing the conversational flow and children's engagement patterns with the voice-based sleep diary.

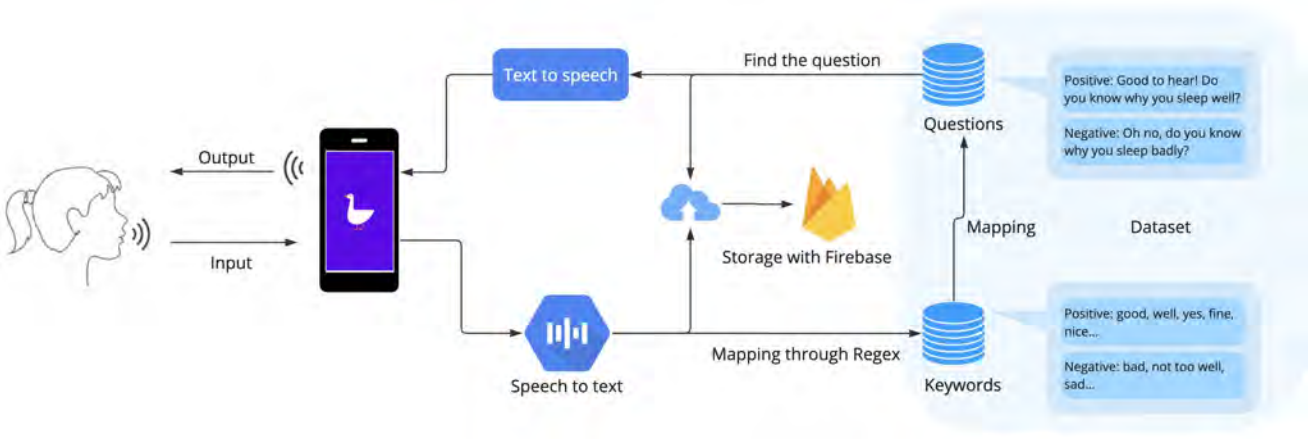

Figure 1: *Overview of the data processing procedure.*

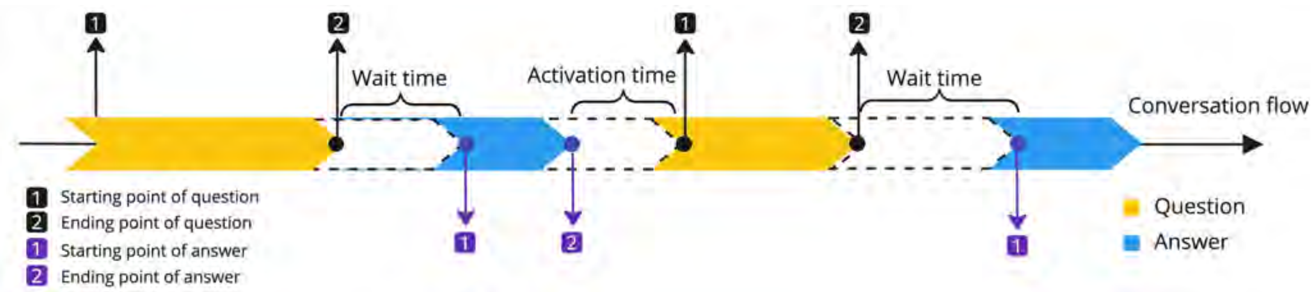

Figure 2: *Conversation timestamps.*

For added transparency and flexibility, the configuration files are publicly accessible through the project's Github examples repository (Chen (2024)).

3.2. The Text-based Interface – Snoozy

Snoozy, designed by Aarts et al. (2022), is a chatbot-based sleep diary for children built using the Landbot platform. Children response to questions by typing their answers. Figure 3 shows the interfaces between Dozzz and Snoozy, with Dozzz (the voice-based interface) on the left and Snoozy (the text-based interface) on the right. These two sleep diaries, representing voice-based and text-based interfaces, will be used as the materials in our experiment.

3.3. The Usability Scale for Children – SUStory

We redesigned the System Usability Scale (SUS) for kids into a game-board format called SUStory (Chen et al. (2024)) to enhance children's engagement with the questionnaire and to better reflect their actual experiences with systems. Specifically, we crafted two backstories about the two characters in Dozzz and Snoozy, a duck and snail (Figure 4). We then integrated the questions from the SUS for children (Putnam et al. (2020)) into these stories. Each question was presented in a cloud, symbolizing

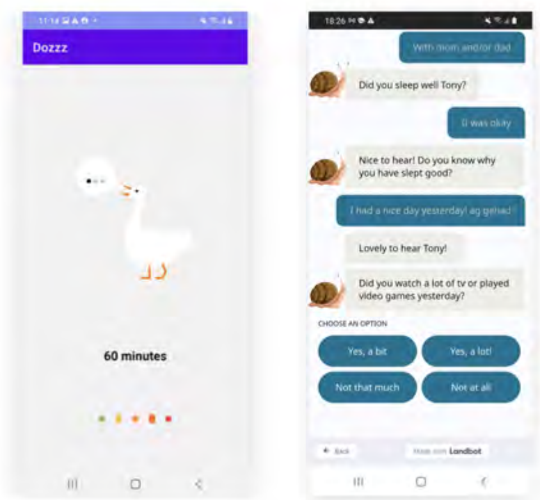

Figure 3: *The interfaces between Dozzz and Snoozy.*

a challenge for children to overcome. The questions include five options based on the Likert Scale.

To make the process more engaging, we removed colors from the five options and created colorful icon cut-out cards. Children could place these vibrant cards on their chosen selections, effectively indicating their answers (Figure 5). While the design is playful, the questions and scales align with the

Table 2: *Subject demographics*

Variables	Statistics
Gender	14 girls, 21 boys
Age	M = 9.39 (SD = 1.18)
Nationalities	Netherlands (15), Indian (7), Romanian (3), Italian (2), Japanese (2), Chinese (2), Vietnam (1), Kazakhstan (1)

standard SUS for kids (Putnam et al. (2020)). This approach aims to provide a more engaging and child-friendly assessment, leveraging the story context and the visual appeal of the chatbot characters.

Figure 4: *The usability scale for children*

4. EXPERIMENT

To evaluate the voice-based self-report interface, we conducted an experiment comparing it to the text-based Snoozy. This was with a 1x2 within-subject experiment with the interaction technique as the within-subjects independent variable having two levels: text-based and voice-based. Our assessment focused on two key aspects: the efficacy of the voice-based conversation and the overall user experience of the sleep diary. We set out to compare the user experience of each interactive sleep diary, and the children's preferences regarding the user interface.

4.1. Participants Demographics and Recruitment Process

We recruited 35 children aged eight to twelve, including 14 girls and 21 boys, for our study, mainly from a public library and local primary schools (Table 2). The recruitment process involved providing parents with information two weeks in advance, and all parents who agree with participation signed a consent form. The university's Ethical Review Board approved the study.

4.2. Procedure

The study was conducted in a separate room in our department. The experiment occurred in November 2023. Each experiment comprised three sessions, in total lasting thirty minutes. It was conducted on a one-on-one basis with a child and a researcher. A camera and an audio recorder were used to record children's behavior. The experiment sessions were structured as follows:

4.2.1. Session 1: Pre-Assessment.
In the first session, after completing consent forms with parents and participants, the participant was leaded to a separated room and introduced the study's purpose and structure. We then taught them how to use the chatbot, guiding them until they could independently complete the self-reporting task.

4.2.2. Session 2: User Test and Assessment.
Each child engaged with these two sleep diaries, Dozzz and Snoozy, completing SUStory after each interaction. This meant that every participant completed two sleep diaries and two sets of questionnaires. To minimize potential counterbalance effects, the order in which they tested the diaries was randomized.

4.2.3. Session 3: Post-Assessment.
A semi-structured interview assessed general preferences in terms of perceptions such as ease of use, attractiveness, efficiency, creativity, dependability, and stimulation, by comparing Dozzz with Snoozy.

5. MEASURES

5.1. Transcription accuracy

To evaluate the conversational efficiency of Dozzz, we used the transcription accuracy evaluation method as described by Park (Park et al. (2008)), which suggested that the word error rate (WER) is the primary metric for evaluating the precision of speech recognition. As shown in Equation 1.

$$WER = \sum_{u=1}^{m} \sum_{i=1}^{n} WER_{u_i} \frac{L_{u_i}}{L_T} \qquad (1)$$

In this equation, WER$_{u_i}$ represents the individual word error rate for each utterance, where $i \in \{1, \ldots, n\}$, and n is the total number of utterances. L$_{u_i}$ denotes the number of words in a given utterance, where $u \in \{1, \ldots, m\}$, and m is the total number of participants. L$_T$ is the total number of words (TNW) in the entire transcription. WER$_{u_i}$ for each utterance is calculated by summing the number of insertions, substitutions, and deletions, then dividing by the TNW, comparing the transcript with the actual recordings.

In practice, a child responds to questions from Dozzz, and these answers are converted to text and stored in the Firebase. Simultaneously, a voice recorder captures the child's actual voice responses as the baseline. WER is then calculated for each response using Equation 1.

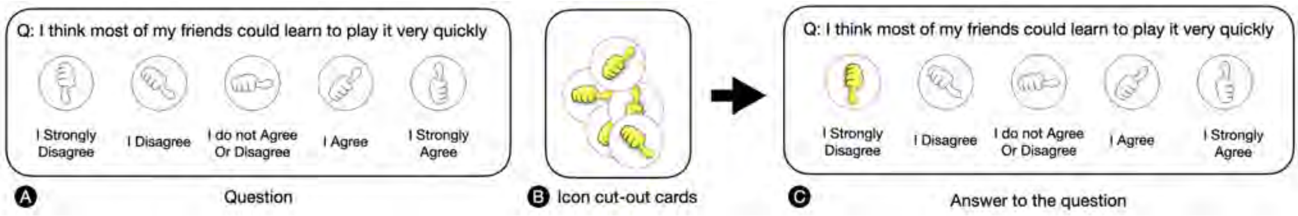

Figure 5: *The question in SUStory*

5.2. Turn-taking Analysis

Effective conversation relies on a smooth transition from one speaker to the next, without excessively long pauses between turns (O'Connell et al. 1990). In our analysis of the interactions between children and Dozzz, we observed that children tend to pause slightly longer before answering certain questions compared to others. These pauses are significant in understanding interactive behaviors and emotions, as hesitations or moments of silence in a conversation can indicate a lack of confidence or signal potential misunderstandings (Park et al. 2008).

To gain a deeper insight into the children's conversational performance, we concentrated on analyzing the "wait time" — the duration before a child began to respond to Dozzz's questions during the turn-taking procedure (Sari 2020). This focus allowed us to further explore how children engage in and manage the flow of conversation with the voice-based sleep diary.

5.3. Preference and Attitude

We analyze children's preference through calculating the result in SUStory for Dozzz and Snoozy, based on the System Usability Scale adapted for children (Putnam et al. (2020); Lewis (2018)). Additionally, we conducted semi-structured interviews with children, asking about the usefulness and enjoyment of using Dozzz and Snoozy. The interviews focus on six aspects: efficiency, attractiveness, creativity, ease of use, dependability, and stimulation. For analysis, we used thematic analysis (Clarke and Braun (2013)) with several stages (Miles and Huberman (1994)), including codebook development with 108 initial codes, discussion and refinement to 98 codes, independent coding with NVIVO for consistency, convergence meetings resulting in four themes and 91 agreed-upon codes, and an inter-rater reliability check that confirmed a 91.42% consistency in coding.

This process provides a comprehensive understanding of children's perspectives, allowing us to identify key themes and patterns in their interactions with both the voice-based and text-based chatbots in the context of the sleep diaries.

6. RESULTS

6.1. Data processing

All 35 children participated in the user test, but two children, one with dyslexia and another with autism, required the experimenter's assistance rather than completed tasks independently. To minimize potential bias, we chose to exclude their data, resulting in data from 33 participants being analyzed.

6.2. The Transcription Accuracy of Dozzz is Acceptable for Children

In the data collected from the sleep diary, for a more detailed conversational analysis, we classified each of the ten questions and the corresponding child's answer as one Turn Constructional Unit (TCU) (Goodwin and Heritage 1990). This implies that the conversation comprised a total of ten TCUs. The analysis involves comparing voice recordings with the corresponding transcript data from the backend platform by the speech-to-text transformation. It calculates the TNW and Error Number of Words (ENW) for each question. As shown in the left side of Figure6, open-ended questions (Q3, Q7, Q8, Q9, and Q10) prompted notably longer and more detailed responses from the children, with TNW values consistently exceeding 3. This indicated a willingness among children to engage in more extended conversations, particularly when responding to questions that allowed for expressive answers. The TNW in each TCU enables the calculation of the WER using Equation1 to evaluate transcription accuracy. Both the global WER and the question-specific WER for each TCU were found to be under 0.2 (Table 3), suggesting an acceptable level of efficiency in the conversation (eric urban (2023)) (The WER under 20% is considered acceptable).

Furthermore, the study focuses also on TCU – categorizing them into 'Wh-group' for open-ended questions and 'Yes-no group' for yes/no questions. The examination of WER distribution between these groups revealed intriguing insights. The findings indicated that the WER in the 'Wh- group', was lower $(M = 0.068, SD = 1.17)$ compared to the WER in the 'Yes-no group' $(M = 0.11, SD = 0.32)$ (The right side of Figure 6). This difference is significant when examined using a one-way, within-subject AN, revealing a significant effect of question types in two groups on the WER. This suggests a relatively higher accuracy $(M = 0.932, SD = 1.17)$ in transcribing responses to open-ended questions than to close-ended questions $(M = 0.89, SD = 0.32)$. Notably, the recognition error patterns differ between the two groups. In the 'Wh- group', where open-ended questions prompted more extensive responses, the primary issue lies in word substitution, for example, "go to school" is replaced to "go to scan". On the other hand, the 'Yes-no group', featuring simpler yes/no questions, exhibited word deletion as the primary error, for example, "yes, I have time" is deleted "yes".

6.3. Children Engaged in Conversation with Dozzz actively

To explore the difference in wait time between the two question types, we calculated the mean wait time for each question (Figure 7 (left side)). From the mean wait time for each question, we can discern the contrast between the two types more clearly. The analysis discerned distinct patterns between open-ended questions ('Wh-group') and closed questions ('Yes-no group'). Figure 7 shows the difference in wait time between these two groups of questions. Results indicate that the mean wait time of 33 participants in the 'Wh-group' $(M = 3.69s, SD = 5.21)$ was significantly longer than the wait time in the 'Yes-no group' $(M = 2.21s, SD = 4.12)$. A one-way ANOVA revealed a significant effect of the question types on test wait time, $F(2, 64) = 5.16, p < -0.001$. The effect size, eta square (η^2), was 0.29, indicating a larger effect.

This difference suggests that children exhibited less confidence and faced bigger challenges when dealing with open-ended questions (Kroeger and Thuesen 2013). For instance, when Dozzz asked an open question like "What do you want to do today? (Q9)", children would spend an extended time contemplating or looking at their parents for suggestions. The wait time was significantly longer when children answered "What did you eat and drink? (Q8)" compared to "Did you watch a lot of TV or play video games? (Q4)".

6.4. Usability Assessment for Sleep Diaries

The results of the SUS for kids highlight noticeable differences between the experiment condition (Dozzz) and the control condition (Snoozy). Figure 8 depicts the distribution of scores, revealing substantial differences.

In terms of the standardization of overall usability, Dozzz demonstrated superior performance $(M = 81.75, SD = 8.84)$ compared to Snoozy $(M = 73.17, SD = 12.08)$. A statistical analysis using one-way ANOVA, $F(2, 64) = 2.77, p < 0.05$, within an eta squared (η^2) = 0.11, highlighted a significant effect of the interactive method on usability. While the upper limits (maximum values) in both types of sleep diary did not differ significantly, the lower limit (minimum value) with Dozzz was higher than that with Snoozy. This suggests a more favorable and consistent perception of usability with Dozzz compared to Snoozy among the participants. Furthermore, we analyzed individual subject responses.

6.5. Children's Attitudinal Perception with Sleep Diaries

Our study's semi-structured interviews revealed children's attitudes towards Dozzz and Snoozy, focusing on six aspects: efficiency, attractiveness, creativity, ease of use, dependability, and stimulation. Out of 33 children, 72% (24) preferred Dozzz, 18% (6) favored Snoozy, and the remaining had no clear preference. Table 4 details the user experience metrics for both options across these factors and overall preference.

To further understand the underlying reason for their preference, we asked the children to explain why they thought that in the semi-structured interview. Subsequently, we conducted a thematic analysis of the interview content to uncover underlying patterns. Four children expressed a preference for Snoozy over Dozzz due to its dependability, as Snoozy allows them to edit or modify their input (C4, C10, C30, C33). This limitation in the voice-based interface seems to hinder the flexibility and control that children desire over their responses. One child explained, "If you typed something wrong, you could also remove it and then type again (C33)." It shows that children expected a sense of control over the activities (Hourcade (2015)). Four children appreciated the text version's capability to edit the content they input. It allowed them to ensure accuracy and provided a greater sense of autonomy. Seven children explained their preference for efficiency also based on this reason (C4, C6, C10, C17, C18, C22, C30). Additionally, the activation time of Dozzz is longer than Snoozy because the Speech-to-text APIs are more time-consuming threads tasks. Children may feel that Dozzz responds slower than Snoozy.

When we inquired about what made Dozzz more appealing than Snoozy, several children highlighted specific features that attracted them.

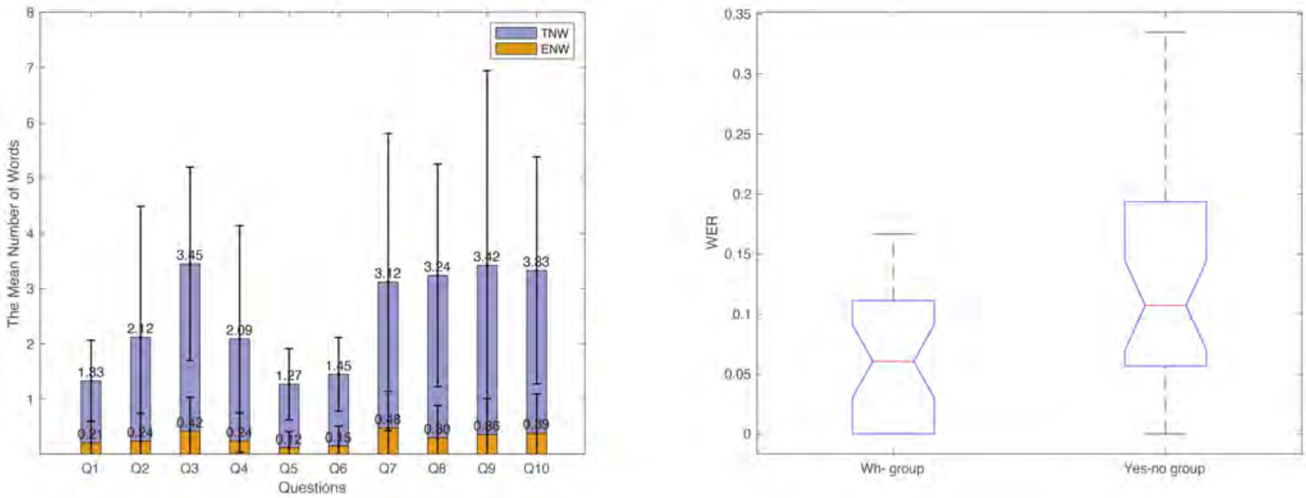

Figure 6: *The Mean Number of Words and the Word Error Rate (WER).*

Table 3: *The WER for Each Question and the Global WER. WER$_w$ represents the global word error rate. $u \in \{1, \ldots, m\}$ indicates the word error rate in each TCU.*

WER$_w$	**WER$_1$**	**WER$_2$**	**WER$_3$**	**WER$_4$**	**WER$_5$**	**WER$_6$**	**WER$_7$**	**WER$_8$**	**WER$_9$**	**WER$_{10}$**
0.12	0.13	0.11	0.12	0.12	0.06	0.10	0.16	0.09	0.11	0.12

Table 4: *User experience metrics with Dozzz and Snoozy.*

	Dozzz	**Snoozy**	**No preference**
Ease of use	84.85%	9.09%	6.06%
Attractiveness	63.64%	24.24%	12.12%
Efficiency	33.33%	54.54%	12.12%
Creativity	64.64%	33.33%	3.03%
Dependability	45.45%	48.48%	6.06%
Stimulation	72.73%	24.24%	3.03%
General preference	72.73%	18.19%	9.09%

Three participants (C4, C6, C11) were drawn to the engaging character imagery and animations within the interface. One child said, "I like the duck, and it can move (C4)". Additionally, the voice feature of Dozzz captivated five children (C3, C8, C13, C29, C31). For instance, one child shared, "It sounds like Google every time. I like to play with Google (C31)". Furthermore, two children (C7, C15) expressed that they simply found the act of speaking to be enjoyable. C7 explained, "I think speaking is fun because I don't need to use my fingers". These responses suggest that the visual and auditory elements of Dozzz, along with the ease of voice interaction, contributed significantly to its appeal among the children.

Regarding their willingness to use Dozzz every day, 24 of them expressed they would like to play with it continuously, but they also proposed some improvements. For example, C1, C9, and C13 hoped the content could be different every day, saying, "I hope the conversation can be different (C9)." This seems an important observation as it is known that children's interest and engagement may decline over time, even if they are initially excited about an activity (Ingram and Elliott (2014)). Dynamic conversations with personalization can present a potential solution to sustain their interest. Additionally, two children mentioned a desire for social interactions, hoping to engage in similar activities with their peers (C13 and C26), and expressed the need to play with peers together. "If my friends can do the same thing with me, I like to use it every day (C13)." Incorporating social elements is an interesting possibility as social interaction has been shown to enhance children's engagement (Fernaeus et al. (2010); Schneider (2010)).

7. DISCUSSION

The results of our evaluation indicated that Dozzz effectively engaged children in interaction with the voice conversational agent during self-reporting. Our qualitative analysis shed light on the children's preferences for voice user interfaces (VUIs) in sleep diaries, as well as the potential challenges associated with their use. In this section, we explore the design considerations that informed the development of Dozzz. We reflect on insights gained from our experiments, discuss the interface design tailored specifically for children, and address the limitations of our study along with potential directions for future research.

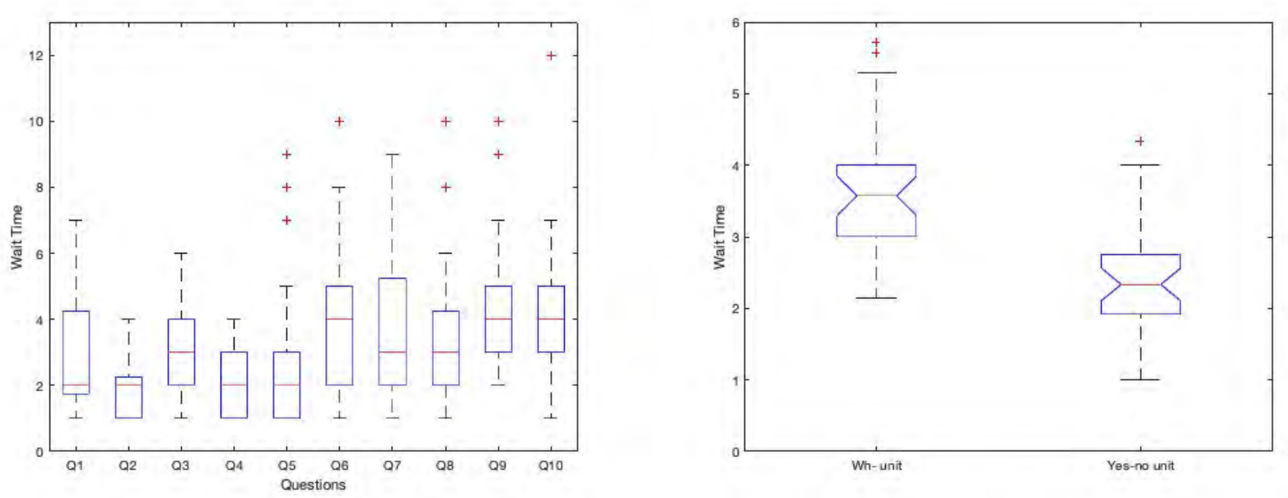

Figure 7: *Wait time analysis for each question and question groups.*

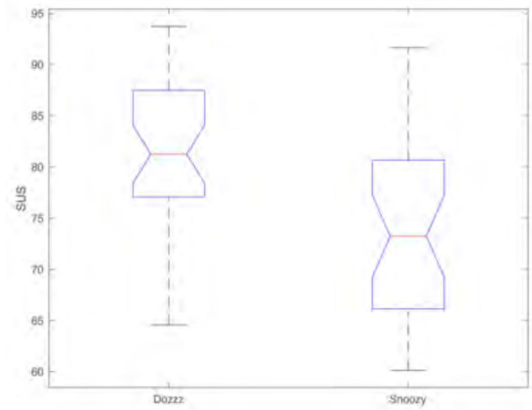

Figure 8: *Comparative analysis of SUS scores between Snoozy and Dozzz*

7.1. Design Considerations

Throughout the design and implementation of Dozzz, we integrated additional features, to explore voice interfaces for self-reporting and create a more personalized conversation experience for the children.

7.1.1. Designing Voice-Based Interaction
Voice interfaces offer a unique advantage in presenting complex or abstract content without necessitating literacy skills. A simple and uncluttered interface minimizes distractions, thereby helping children maintain focus on the task at hand (Chiong and DeLoache (2013)). In our approach, we opted to move away from visual symbolic representations to convey ideas. Instead, we demonstrated the feasibility and effectiveness of guiding children to through sleep reporting via speech, employing just-in-time dialogues. This method suggests that a combination of sound and images could be a more suitable medium for communication with children than written text (Oakes et al. 2009).

The interactive and playful nature of engaging in voice-based conversations by allowing direct speaking with chatbot aligns well with the interests and capacities of children at their age (Markopoulos et al. 2008; Xu et al. 2021). Additionally, the voice-based interface does not require accurate motor skills on a touchscreen, which may be difficult for children. The results showed that most children (73%, 24/33) expressed a preference for the voice-based interface, with a positive experience.

However, some children also expressed a preference for the text-based sleep diary in specific scenarios, such as classrooms or public places where quietness is required (Mirvis 1991). It's important to note that a child's reluctance to speak can be influenced by their mood or emotional state (Anthony 2019), underscoring the significance of situational factors in shaping children's preferences. This aspect was evident in our study, where we observed that children sometimes provided an invalid or unintended input during the interaction with the chatbot. In such instances, they expressed a desire to revisit and correct their previous response, which allowed them to edit their inputs. This finding suggests that children value having a sense of control over their interactions (Hourcade 2015), highlighting an area for potential enhancement in voice-based interfaces.

Considering these insights, a multi-modal interaction with the sleep diary seems preferable, with some redundancy allowing children to self-report through either typing or speaking could offer flexibility while mitigating the disadvantages associated with the voice version. Another option is for the voice-based conversation to roll back, enabling children to change their earlier responses, and addressing concerns related to accuracy and control. These considerations respond to the need to design child-friendly interfaces that consider individual differences and contextual factors.

7.1.2. Supporting Personalization

Responding to the answers to children empathetically was another important decision when designing Dozzz, recognizing the importance of creating a supportive and friendly conversation for children (Fu et al. 2022). Understanding that humans extend their emotions to software systems based on their responses (Casas et al. 2021), a conversational interface should depart from traditional sleep diaries, which often rely on rigid and static question-based structures (Văcărețu et al. 2019; Carney et al. 2012). Instead, Dozzz maintains a warm and empathetic tone in response to children's answers, fostering a kind and understanding interaction. Furthermore, we intentionally designed Dozzz to personalized engage in conversations by addressing the child and Dozzz itself by their own name, aligning with children's preferences (Yarosh et al. 2018). Unlike formulaic questioning processes found in traditional diaries, we encourage an introductory approach where Dozzz introduces itself to children, remembers their names, and addresses them by name when posing follow-up questions. This design choice aims to enhance the intimate relationship between children and the agent. Finally, to address potential language issues arising from children's speech characteristics, including limited vocabulary, atypical pitch, pronunciation, and syntax errors (Druga et al. 2017; Vance et al. 2005), we incorporated a feature that kindly rephrases previous questions in cases of misunderstanding. This serves to reduce misunderstanding during the conversation, contributing to a more inclusive user experience.

7.1.3. Dozzz as a Conversational Partner

Our study indicates a clear preference among children for interactions that resemble real human-like conversations with the sleep diary. In response to this preference, we have evolved the interface of Dozzz to support a more immersive conversational experience. This was achieved by transitioning from button-controlled speech input to a free-speaking, turn-taking style. Such an approach enables Dozzz to more closely emulate the dynamics of a conversation with a human-like partner.

Additionally, we enhanced the interface with visual feedback elements that correspond to children's speech. This includes animations of Dozzz characters and dynamic fluctuations of speaking dots, which visually represent the system's attractiveness to children's speech input. These visual cues not only add to the engagement factor but also reinforce the perception of Dozzz as an active, responsive participant in the conversation.

7.2. Towards sustained engagement with Conversational Sleep Diaries

Following the evaluation of Dozzz, our future focus is on exploring how it can effectively support children in continuously self-reporting their sleep experiences over extended periods. This study has highlighted several design considerations crucial for developing conversational dynamics that foster long-term engagement with the sleep diary (Chubb et al. 2022).

Our envisioned scenario involves children participating in engaging, daily conversations with Dozzz. In this setting, Dozzz would extend its role beyond simply asking questions to collect information. It would provide relevant feedback, respond to the children's queries, and engage them in a supportive and interactive dialogue. Dozzz aims to act as a companionable entity, encouraging children to openly discuss their sleep patterns. By tailoring conversations to individual preferences and promoting a natural, flowing dialogue, Dozzz could significantly enhance the experience of tracking and discussing sleep, thereby encouraging sustained use.

Clinicians typically require patients to keep a sleep diary for one or two weeks. This allows them to observe the daily routine which may be very dependent on the weekly schedule. Thus, future work should evaluate the usage of Dozzz in context, for such a longer duration and examine whether it helps to capture reliably children's sleep quality experience and sleep-related behaviours over time. Additionally, while our observations indicated children's certain level of acceptance towards voice-based sleep diaries, there was not sufficient evidence regarding the suitability of this technology in various contexts. Questions remain, such as whether it is an ideal choice for moments upon waking, during the night, or even when children are engaged in playing with friends. The other limitation is that most of the children in the current study had no sleep problems. The study is more for investigating the feasibility of a voice-based sleep diary, than for comprehensive insights into longitudinal engagement and adherence with children with and without sleep disorders. This should be addressed in the next longitudinal study. Therefore, further exploration with Dozzz, the voice-based sleep diary, is needed to investigate the conditions and situations that children encounter in their daily lives for an extended duration.

We observed that approximately half of the children were entirely new to smartphones during the experiment, often due to parental restrictions on smartphone usage. These restrictions stem from concerns related to smartphone addiction observed in children and adolescents, including physical and psychological problems associated with excessive use and difficulties in performing daily activities, as reported in various studies (Haug et al. 2015; Lopez-Fernandez et al. 2014; Virzi 1992; Hill et al. 2006). Recognizing these concerns, we plan to explore other smart devices that parents may find more acceptable for use by their children. This includes devices like smart speakers, watches, and smart toys designed for children, offering varied options for voice-based interaction that could potentially address concerns related to excessive smartphone use.

8. CONCLUSION

We introduced Dozzz, a voice-based sleep diary utilizing Speech-to-Text APIs to empower children in self-reporting sleep-related behaviors and experiences. Our analysis conducted through a within-subject experiment, comparing two different input modalities– voice and text– in sleep diaries. The study involved 35 participants aged eight to twelve. We presented (1) our design of voice-based sleep diary, and (2) the result of an evaluation demonstrating Dozzz's effectiveness and user experience in a controlled experiment by comparing it with a text-based sleep diary.

Dozzz's voice-based chatbot interface for sleep diaries effectively engaged children and was perceived as more user-friendly for ongoing sleep experience reporting than text-based options. Our findings highlight the substantial potential of voice-based chatbot interactions in maintaining children's interest in documenting their sleep patterns. This innovative approach demonstrated the effectiveness of voice-based interfaces in sleep sampling, suggesting significant benefits for both user engagement and data accuracy in children's sleep health assessment.

REFERENCES

Aarts, T., Markopoulos, P., Giling, L., Vacaretu, T., and Pillen, S. (2022). Snoozy: a chatbot-based sleep diary for children aged eight to twelve. In *Interaction Design and Children*, pages 297–307.

Anthony, L. (2019). Physical dimensions of children's touchscreen interactions: Lessons from five years of study on the mtagic project. *Int. J. Hum.-Comput. Stud.*, 128(C):1–16.

Barrett, M., Berney, T., Bhate, S., Famuyiwa, O., Fundudis, T., Kolvin, I., and Tyrer, S. (1991). Diagnosing childhood depression. who should be interviewed-parent or child?: The newcastle child depression project. *The British Journal of Psychiatry*, 159(S11):22–27.

Bird, H. R., Gould, M. S., and Staghezza, B. (1992). Aggregating data from multiple informants in child psychiatry epidemiological research. *Journal of the American Academy of Child & Adolescent Psychiatry*, 31(1):78–85.

Blader, J. C., Koplewicz, H. S., Abikoff, H., and Foley, C. (1997). Sleep problems of elementary school children: a community survey. *Archives of pediatrics & adolescent medicine*, 151(5):473–480.

Brockmann, P. E., Diaz, B., Damiani, F., Villarroel, L., Núñez, F., and Bruni, O. (2016). Impact of television on the quality of sleep in preschool children. *Sleep medicine*, 20:140–144.

Brown, K. and Kennedy, H. (2011). Learning through conversation: Exploring and extending teacher and children's involvement in classroom talk. *School Psychology International*, 32(4):377–396.

Carney, C. E., Buysse, D. J., Ancoli-Israel, S., Edinger, J. D., Krystal, A. D., Lichstein, K. L., and Morin, C. M. (2012). The consensus sleep diary: standardizing prospective sleep self-monitoring. *Sleep*, 35(2):287–302.

Casas, J., Spring, T., Daher, K., Mugellini, E., Khaled, O. A., and Cudré-Mauroux, P. (2021). Enhancing conversational agents with empathic abilities. In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*, IVA '21, page 41–47, New York, NY, USA. Association for Computing Machinery.

Chen, S. (2024). Dozzz source code and related documents. `https://github.com/Shanshan-css/Dozzz-v2.0.git`. Accessed: 2024-06-22.

Chen, S., Markopoulos, P., and Hu, J. (2023). Dozzz: Exploring voice-based sleep experience sampling for children. In *International Conference on Pervasive Computing Technologies for Healthcare*, pages 490–500. Springer.

Chen, S., Markopoulos, P., and Hu, J. (2024). Sustory: Usability evaluation of conversational interfaces for children with a narrative on a game board. In *Proceedings of the 6th International Conference on Conversational User Interfaces*.

Chervin, R. D., Dillon, J. E., Archbold, K. H., and Ruzicka, D. L. (2003). Conduct problems and symptoms of sleep disorders in children. *Journal of the American Academy of Child & Adolescent Psychiatry*, 42(2):201–208.

Chervin, R. D., Dillon, J. E., Bassetti, C., Ganoczy, D. A., and Pituch, K. J. (1997). Symptoms of sleep disorders, inattention, and hyperactivity in children. *Sleep*, 20(12):1185–1192.

Chiong, C. and DeLoache, J. S. (2013). Learning the abcs: What kinds of picture books facilitate young children's learning? *Journal of Early Childhood Literacy*, 13(2):225–241.

Chubb, J., Missaoui, S., Concannon, S., Maloney, L., and Walker, J. A. (2022). Interactive storytelling for children: A case-study of design and development considerations for ethical conversational ai. *Int. J. Child-Comp. Interact.*, 32(C).

Clarke, V. and Braun, V. (2013). Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. *The psychologist*, 26(2):120–123.

Danielle Pacheco, D. N. V. (2023). Children and sleep-an introduction to the importance of sleep in children and how to help them sleep better.

Druga, S., Williams, R., Breazeal, C., and Resnick, M. (2017). "hey google is it ok if i eat you?": Initial explorations in child-agent interaction. IDC '17, page 595–600, New York, NY, USA. Association for Computing Machinery.

eric urban, n. (2023). Test accuracy of a custom speech model.

Fernaeus, Y., Håkansson, M., Jacobsson, M., and Ljungblad, S. (2010). How do you play with a robotic toy animal? a long-term study of pleo. In *Proceedings of the 9th International Conference on Interaction Design and Children*, IDC '10, page 39–48, New York, NY, USA. Association for Computing Machinery.

Fu, Y., Michelson, R., Lin, Y., Nguyen, L. K., Tayebi, T. J., and Hiniker, A. (2022). Social emotional learning with conversational agents: Reviewing current designs and probing parents' ideas for future ones. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(2).

Garg, R. and Sengupta, S. (2020). He is just like me: a study of the long-term use of smart speakers by parents and children. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–24.

Goodwin, C. and Heritage, J. (1990). Conversation analysis. *Annual review of anthropology*, 19(1):283–307.

Haug, S., Castro, R. P., Kwon, M., Filler, A., Kowatsch, T., and Schaub, M. P. (2015). Smartphone use and smartphone addiction among young people in switzerland. *Journal of behavioral addictions*, 4(4):299–307.

Henry, A. D. (2000). *Pediatric interest profiles: Surveys of play for children and adolescents, kid play profile, preteen play profile, adolescent leisure interest profile*. Psychological Corporation.

Hill, A. L., Degnan, K. A., Calkins, S. D., and Keane, S. P. (2006). Profiles of externalizing behavior problems for boys and girls across preschool: the roles of emotion regulation and inattention. *Developmental psychology*, 42(5):913.

Hourcade, J. P. (2015). Child-computer interaction. *Self, Iowa City, Iowa*.

Ibáñez, V., Silva, J., and Cauli, O. (2018). A survey on sleep questionnaires and diaries. *Sleep medicine*, 42:90–96.

Ingram, J. and Elliott, V. (2014). Turn taking and 'wait time'in classroom interactions. *Journal of Pragmatics*, 62:1–12.

Issenman, R. M. and Jaffer, I. H. (2004). Use of voice recognition software in an outpatient pediatric specialty practice. *Pediatrics*, 114(3):e290–e293.

Ivens, C. and Rehm, L. P. (1988). Assessment of childhood depression: correspondence between reports by child, mother, and father. *Journal of the American Academy of Child & Adolescent Psychiatry*, 27(6):738–741.

Jung, H., Kim, H. J., So, S., Kim, J., and Oh, C. (2019). Turtletalk: An educational programming game for children with voice user interface. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–6.

Kroeger, O. and Thuesen, J. M. (2013). *Type talk: The 16 personality types that determine how we live, love, and work*. Dell.

Lewandowski, A. S., Ward, T. M., and Palermo, T. M. (2011). Sleep problems in children and adolescents with common medical conditions. *Pediatric Clinics*, 58(3):699–713.

Lewis, J. R. (2018). The system usability scale: Past, present, and future. *International Journal of Human–Computer Interaction*, 34:577 – 590.

Loeber, R., Green, S. M., Lahey, B. B., and Stouthamer-Loeber, M. (1989). Optimal informants on childhood disruptive behaviors. *Development and Psychopathology*, 1(4):317–337.

Lopez-Fernandez, O., Honrubia-Serrano, L., Freixa-Blanxart, M., and Gibson, W. (2014). Prevalence of problematic mobile phone use in british adolescents. *CyberPsychology, Behavior, and social networking*, 17(2):91–98.

Mallinson, D. C., Kamenetsky, M. E., Hagen, E. W., and Peppard, P. E. (2019). Subjective sleep measurement: comparing sleep diary to questionnaire. *Nature and Science of Sleep*, pages 197–206.

Markopoulos, P., Read, J. C., and Giannakos, M. (2021). Design of digital technologies for children. *Handbook of human factors and ergonomics*, pages 1287–1304.

Markopoulos, P., Read, J. C., MacFarlane, S., and Hoysniemi, J. (2008). *Evaluating children's interactive products: principles and practices for interaction designers*. Elsevier.

Mazza, S., Bastuji, H., and Rey, A. E. (2020). Objective and subjective assessments of sleep in children: comparison of actigraphy, sleep diary completed by children and parents' estimation. *Frontiers in Psychiatry*, 11:495.

Miles, M. B. and Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. sage.

Mirvis, P. H. (1991). Flow: The psychology of optimal experience.

Oakes, L. M., Newcombe, N. S., and Plumert, J. M. (2009). Are dynamic systems and connectionist approaches an alternative to "good old fashioned cognitive development". *Toward a new unified theory of development*, pages 279–294.

O'Connell, D. C., Kowal, S., and Kaltenbacher, E. (1990). Turn-taking: A critical analysis of the research tradition. *Journal of psycholinguistic research*, 19:345–373.

Park, Y., Patwardhan, S., Ramanan, K. V., and Gates, S. C. (2008). An empirical analysis of word error rate and keyword error rate. In *Interspeech*.

Pina, L., Sien, S.-W., Song, C., Ward, T. M., Fogarty, J., Munson, S. A., and Kientz, J. A. (2020). Dreamcatcher: exploring how parents and school-age children can track and review sleep information together. *Proceedings of the ACM on Human-computer Interaction*, 4(CSCW1):1–25.

Putnam, C., Puthenmadom, M., Cuerdo, M. A., Wang, W., and Paul, N. (2020). Adaptation of the system usability scale for user testing with children. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*, pages 1–7.

Riley, A. W. (2004). Evidence that school-age children can self-report on their health. *Ambulatory Pediatrics*, 4(4):371–376.

Sari, C. C. (2020). Conversation analysis: Turn-taking mechanism and power relation ij classroom setting. *Celtic: A Journal of Culture, English Language Teaching, Literature and Linguistics*, 7(2):118–136.

Schaefer, S. E., Van Loan, M., and German, J. B. (2014). A feasibility study of wearable activity monitors for pre-adolescent school-age children posted. 11(130262).

Schneider, W. (2010). Memory development in childhood. *The Wiley-Blackwell handbook of childhood cognitive development*, pages 347–376.

Short, M. A., Gradisar, M., Lack, L. C., Wright, H., and Carskadon, M. A. (2012). The discrepancy between actigraphic and sleep diary measures of sleep in adolescents. *Sleep medicine*, 13(4):378–384.

Sivertsen, B., Hysing, M., Elgen, I., Stormark, K. M., and Lundervold, A. J. (2009). Chronicity of sleep problems in children with chronic illness: a longitudinal population-based study. *Child and adolescent psychiatry and mental health*, 3:1–7.

Spitale, M., Silleresi, S., Cosentino, G., Panzeri, F., and Garzotto, F. (2020). " whom would you like to talk with?" exploring conversational agents for children's linguistic assessment. In *Proceedings of the interaction design and children conference*, pages 262–272.

Stein, M. A., Mendelsohn, J., Obermeyer, W. H., Amromin, J., and Benca, R. (2001). Sleep and behavior problems in school-aged children. *Pediatrics*, 107(4):e60–e60.

Sturgess, J., Rodger, S., and Ozanne, A. (2002). A review of the use of self-report assessment with young children. *British Journal of Occupational Therapy*, 65(3):108–116.

Tewari, A. and Canny, J. (2014). What did spot hide? a question-answering game for preschool children. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1807–1816.

Tudor Car, L., Dhinagaran, D. A., Kyaw, B. M., Kowatsch, T., Joty, S., Theng, Y.-L., and Atun, R. (2020). Conversational agents in health care: scoping review and conceptual analysis. *Journal of medical Internet research*, 22(8):e17158.

Văcăreţu, T., Batalas, N., Erten-Uyumaz, B., Van Gilst, M., Overeem, S., and Markopoulos, P. (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*, pages 113–122.

van Iterson, H. C., Markopoulos, P., and Gerritsen, L. E. (2023). Motivating children's engagement withnbsp;sleep diaries through storytelling. In *Human-Computer Interaction – INTERACT 2023: 19th IFIP TC13 International Conference, York, UK, August 28 – September 1, 2023, Proceedings, Part II*, page 214–223, Berlin, Heidelberg. Springer-Verlag.

van Rijssen, I. M., Hulst, R. Y., Gorter, J. W., Gerritsen, A., Visser-Meily, J. M. A., Dudink, J., Voorman, J. M., Pillen, S., and Verschuren, O. (2023). Device-based and subjective measurements of sleep in children with cerebral palsy: A comparison of sleep diary, actigraphy, and bed sensor data. *Journal of Clinical Sleep Medicine*, 19(1):35–43.

Vance, M., Stackhouse, J., and Wells, B. (2005). Speech-production skills in children aged 3–7 years. *International Journal of Language & Communication Disorders*, 40(1):29–48.

Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? *Human factors*, 34(4):457–468.

Wiggs, L. and Stores, G. (1996). Severe sleep disturbance and daytime challenging behaviour in children with severe learning disabilities. *Journal of Intellectual Disability Research*, 40(6):518–528.

Winstone, N. E., Nash, R. A., Parker, M., and Rowntree, J. (2017). Supporting learners' agentic engagement with feedback: A systematic review and a taxonomy of recipience processes. *Educational psychologist*, 52(1):17–37.

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

Xu, Y. and Warschauer, M. (2020). Exploring young children's engagement in joint reading with a conversational agent. In *Proceedings of the interaction design and children conference*, pages 216–228.

Yarosh, S., Thompson, S., Watson, K., Chase, A., Senthilkumar, A., Yuan, Y., and Brush, A. J. B. (2018). Children asking questions: Speech interface reformulations and personification preferences. IDC '18, page 300–312, New York, NY, USA. Association for Computing Machinery.