Preprint version

Dozzz: Exploring Voice-based Sleep Experience Sampling for Children

Shanshan Chen [1][0000-0002-6475-179X], Panos Markopoulos [1] [0000-0002-2001-7251], and Jun Hu [1] [0000-0003-2714-6264]

[1] Department of Industrial Design, Eindhoven University of Technology, 5612 Eindhoven, AZ, The Netherlands
{s.chen1, p.markopoulos, j.hu}@tue.nl

Abstract. Text-based digital diaries are an essential tool for sleep clinicians to assess how their patients experience sleep. However, text-entry can be challenging for children. Voice entry represents a plausible and yet unexplored alternative for supporting children's self-report in sleep diaries. We introduce Dozzz, a voice-based digital sleep diary that empowers children to record their sleep experiences using a smartphone. We present the result of usability evaluation involving ten children aged six to twelve. This evaluation confirmed that children were able to understand and interact with Dozzz effectively. Our study demonstrates the feasibility of voice-user interfaces (VUIs) to support sleep diaries for children. Future work needs to assess the use of diaries in real-life settings and evaluate the quality of responses children provide when using the system independently at home.

Keywords: voice-user interfaces, experience sampling method, children-computer interaction, sleep diary, self-tracking

1 Introduction

Sleep problems can have a significant impact on children's mental and physical health and development; they are very common and often chronic, with a prevalence estimated between 25-40% [4]. Sleep problems may refer to difficulties in falling asleep, interrupted sleep, drowsiness during the day, etc. To assess and subsequently treat sleep problems health care specialists like pediatricians, psychiatrists and psychologists, need to identify sleep-related patterns and behaviours such as bedtimes, routines related to going to sleep and waking up, co-sleeping, habits relating to watching television or using digital media, nocturnal behaviours such as snoring and nightmares, and daytime behaviours affected by sleep [1].

There are various ways for health care specialists to collect information about sleep, which include discussing with children and their parents and technical means such as actigraphy and polysomnography. Next to these, diaries can be useful for tracking daily sleep behaviour and experiences, allowing the assessment of night-to-night variability and other patterns over time. Diaries can be filled by the parents, but also there are sleep diaries specially developed for children to report their sleep experience independently [16][17]. Children's reports in a sleep diary have been argued to be more valid as children are actively engaged in their experiences throughout the entire time, compared with a sleep diary for parents [8][18]. Additionally, previous research has found low correlations between parent and child reports regarding a child's behaviors and emotions [19]. Hence, it is necessary to develop sleep diaries that enable children to self-report their sleep experiences accurately and reliably, which may be particularly difficult for children.

Traditional paper-based diaries are not always completed within the instructed timeframe, potentially leading to inaccurate responses [7]. For this reason, digital media, such as websites and mobile applications, are increasingly used for self-tracking purposes. These digital systems can help in recording and managing personal data, eliminating the inconvenience of editing, recording, and analyzing data with pen and paper [6][10], especially with the widespread use of digital devices such as mobile phones [22], laptops [23], and wearables [24]. On the other hand, most digital sleep diaries and sleep questionnaires, like the Consensus Sleep Diary [20] and Graphical diaries [12], have primarily targeted adults and have not been tailored for use by children. Snoozy [11] is a chatbot-based sleep diary that assists children in recording their sleep experiences. In a week-long evaluation with children (N=5) having sleep disorders, Snoozy was deemed feasible and enjoyable for children, offering clinical practitioners and sleep scientists a direct source of sleep quality data from children instead of relying on parents. Participants, however, indicated their preference for a voice-based version of the chatbot as it would be more fun to have a vocal conversation.

This paper introduces Dozzz, a voice-based sleep diary designed for children, using Google recognizer technology on the Android platform. It employs voice-user interfaces (VUIs) to facilitate children's self-reporting of sleep related behaviours and experiences. The evaluation involved ten children aged six to twelve, combining quantitative and qualitative methods, including a user test and a semi-structured interview to assess the system's usability and opinions of users regarding its potential for long-term usage.

This study revealed that the children were able to successfully complete the sleep survey entirely by independent voice input. Through data analysis, we found that children were willing to use voice entry for sleep diary self-reporting. Dozzz provides an alternative for researchers to explore voice-based sleep diaries for children in a conversational context, which paves the way for future work to facilitate more natural conversation on portable devices and in more complex contexts, while maintaining the sustained engagement of children.

2 Materials and Methods

2.1 Materials

Software Dozzz, our chatbot-based self-reporting method with voice-based response, was developed in Java, based on the children's sleep diary [11]. Dozzz implements voice input/output using the Google recognizer. It utilizes the voice recognizer APIs for backend infrastructure and capabilities such as voice recognition and transcription. SQLite is used to store user's answers on client devices. (see in Fig 1).

To enhance the chatbot's conversational naturalness with human-like sound, the system utilizes a matching regulation approach, which seeks to offer responses that are relatable to children. Rather than utilizing complex machine learning algorithms, we employ preset search rules based on keyword classification using Regular Expressions (Regex), as depicted in Fig 1.

Dozzz has been meticulously designed to accommodate the following scenarios:

(1) Clarification of Questions: Dozzz responds by kindly rephrasing the last question to ensure children's comprehension when they express confusion.

(2) Empathetic Responses: In response to children's answers, Dozzz maintains a warm and empathetic tone, fostering a kind and understanding interaction.

(3) Personalization: Dozzz enhances the conversational experience by remembering children's names and addressing them by name when asking follow-up questions, creating a more personal connection.

For added transparency and flexibility, the configuration files are publicly accessible through the project's GitHub examples repository [30].

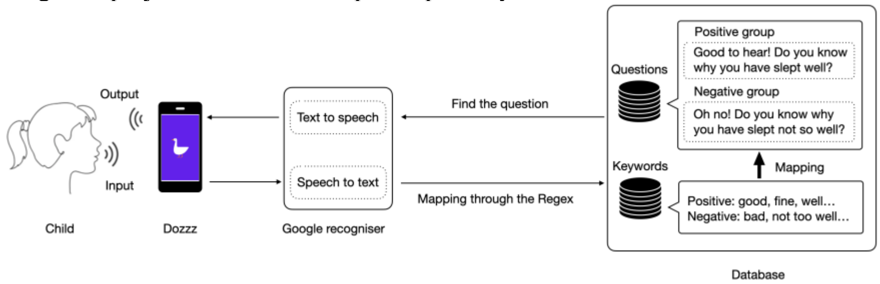

Fig 1 Overview of the data processing approach involves mapping emotional responses based on children's answers by utilizing keyword searches with predefined regex regulations.

2.2 Methods

User Study

A structured user test was carried out in a field setting, where we asked children to interact with Dozzz and respond to its questions about their sleep. Our aim was to examine whether children understood Dozzz and could carry their end of the conversation and to gauge their attitudes towards it.

Recruitment We recruited ten children by making announcements on social media platforms and sharing them widely in parent-social groups, and child centers in public libraries. The recruitment process aimed to reach a diverse group of children.

Experiment duration Each session with a participant lasted 20 to 30 minutes and was conducted one-on-one with a researcher. To ensure the child's comfort, sessions were held in locations like the public library or their home.

Number of participants Due to recruitment constraints and privacy concerns, we received consent from ten children's parents, concluding the evaluation with children aged six to twelve (M = 8.5 years, SD = 1.65), including four boys and six girls.

User test Each child completed the chatbot survey independently, with their behavior recorded by a camera during the 20 to 30 minutes.

Semi-structure interview We conducted a semi-structured interview with the children, covering their chatbot experience, technology usage at home (including voice technology), and daily routine, with parental consent for audio recording.

Data collection In the current system, participant responses lack timestamps in SQLite database. To address this, video recordings are employed to monitor response duration and wait time.

Ethical considerations We received ethics approval from the Ethical Review Board of our university. Data will be deleted after the study, with consent obtained from ten participants and their parents.

Data Analysis

Transcription accuracy We assessed transcription accuracy using methods like Google Cloud Speech and real-time transcription, comparing them to a gold-standard reference corpus. The word error rate (WER) was the primary metric for evaluating speech recognition precision [31]. We also introduced Carly's calculation method for accuracy evaluation [3]. See Equation 1.

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

WER_{u_i} is each utterances' individual word error rate, where $i \in 1, \ldots, n$, n is the number of utterances. L_{u_i} is the number of words in the utterance. L_T is the total number of words (TNW) in the transcription. WER_{u_i} is calculated by summing the number of insertions, substitutions, and deletions, and then divided by the TNW through comparing the transcript and real recordings.

Turn-taking analysis We introduce "pause" in turn-taking, a mechanism for analyzing discourse through phrasing, intonation, and pausing. These pauses, also known as "wait time," are essential for interpreting interactional behaviors and emotions [32], potentially signaling lower confidence. We aim to define wait time for different types of Turn Constructional Units (TCU) [26] (see Fig 2).

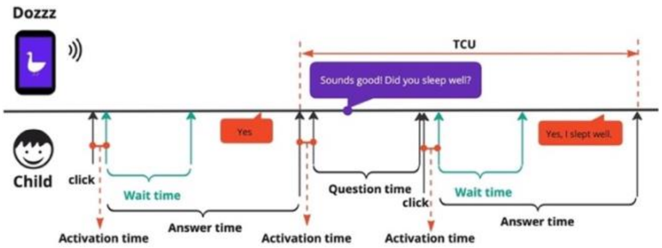

Fig 2 The conversation with Dozzz starts when the user activates the system by clicking a button.

Affinity diagramming and thematic analysis Affinity diagramming and thematic analysis [26][28], involving spatially clustering related insights or ideas to analyze qualitive data obtained from observational studies and user interviews [25][26], was used to group and cluster insights derived from the semi-structured interviews with children in this study and then generate themes from the clustered data. In this study, we use thematic analysis to identify patterns, themes, or categories within the data and using those themes as codes to organize and analyze the children's interview data. Fig 3 shows the resulting clustering of children's statements.

Fig 3 The visualization of affinity diagramming and thematic analysis for children's interviews with experience of voice interaction and daily activities with voice-based devices

3 Results

The analysis of the video recordings provides insights into the participants' activities with Dozzz, an overview of temporal aspects is shown in Table 1. The average answer duration was found to be 2.49s (SD =1.73). The average wait time was 2.02s (SD=1.53). Notably, when children answered yes-no questions, the average wait time was just 1.54s (SD=0.37). However, when participants were asked wh-questions like "Where do you need to go today?", they took more time to think, resulting in an average wait time of 4.04s (SD = 1.15).

Table 1 . Indices on ten children's answers

Answer duration	Wait time	Wait time for yes-no unit	Wait time for wh- unit	WER
2.49s (1.73)	2.02s (1.53)	1.54s (0.37)	4.04s (1.15)	14.9%(0.061)

3.1 Transcription accuracy

Based on Equation 1, we calculated a Word Error Rate (WER) of 0.149 (SD = 0.061) (see table 1), indicating acceptable transcription accuracy [2]. The left of Fig 4 displays WER distribution for the 'wh-unit' and 'yes-no-unit' groups. Although no significant difference was observed, it is notable that the primary cause of recognition errors differs between the groups. In the 'wh-unit' group, word substitution is the primary issue, while in the 'yes-no-unit' group, word deletion prevails. Importantly, many 'yes-no-unit' recognition errors occurred when the child pressed the button before the system fully activated (as shown in Fig 2), suggesting a readiness issue with voice recognition.

3.2 Turn-taking analysis

In the right of Fig 4, we observe a notable difference in wait times between the 'wh-unit' and the 'yes-no-unit' group. Upon reviewing video recordings, it became evident that when children were presented with questions pertaining to future plans or preferences that had not yet occurred, such as "What do you plan to do today?" or "Where do you want to go today?", they frequently responded with phrases like "I don't know" or "I have no idea" after contemplating for a few seconds. These responses suggest that children often felt uncertain or lacked specific intentions or preferences in those situations. This also suggests younger children may face challenges in comprehending and responding to certain questions due to factors like memory limitations, restricted vocabulary, and social factors. Designing voice-based interfaces for children should consider these challenges to enhance the user experience [21][29].

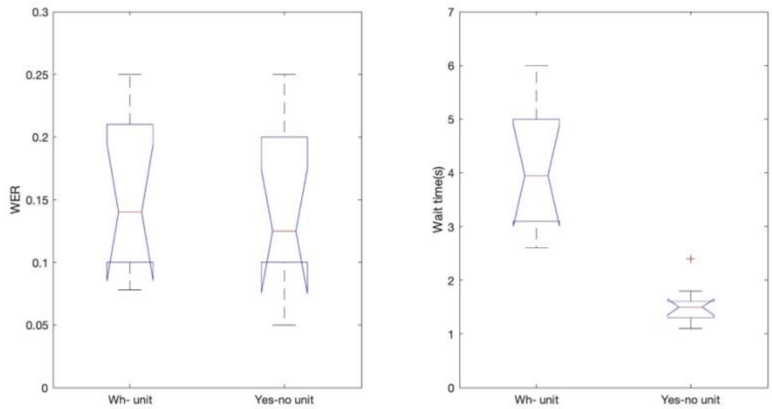

Fig 4 Box plots show WER and wait time for ten children in 'wh-unit' and 'yes-no-unit' groups. In the 'wh-unit' group, WER is slightly higher, while wait time are significantly higher compared to the 'yes-no-unit' group.

3.3 Qualitative Feedback

All the children had previous experience with VUIs on different terminal devices such as smart speakers, phones, and watches. Among the participants, five children (50%) had speakers at home, four children (40%) owned smartwatches, and three children (30%) had tablet computers. Additionally, all the children had the opportunity to interact with smartphones in their daily lives.

The youngest participant favored smart speakers over smartphones. Listening to music and stories was popular among all children, averaging 30 minutes daily. Asking questions through speakers was also popular, averaging ten minutes daily. They used smartwatches to stay in touch with friends and for location tracking. Of the three tablet owners, gaming and online courses were common activities. Preferences for the moments, devices, and activities of engagement varied among different age groups.

Out of the ten participants, 90% expressed willingness to continue self-reporting with their sleep diary. Among these, 66.6% preferred reporting at night, while 33.4% reported whenever they remembered. Additionally, 30% of children aged nine or older had prior experience with similar diary apps using templates instead of typing.

Summary of Themes

The analysis of the data revealed two main themes: enablers and barriers in conversations, and enablers and barriers in digital sleep-diaries. These themes provide insights into children's needs, preferences, and communication activities.

Theme 1: Enablers and Barriers in conversations

Subtheme 1.1: Voice input / output (Enablers). Children found Dozzz easy to use and enjoyable. They also found it funny, enhancing their positive experience. Children who owned smart speakers at home were particularly enthusiastic users of Dozzz, as they were familiar with voice-based interactions and preferred using such devices for communication and entertainment.

> *'I couldn't play with it today.' Her father explained she asked the speaker to play songs and tell stories all day long. However, it's too noisy, especially at night, and disrupted our sleep, so I turned it off. (C2's father)*

Children without smart speakers at home also enjoyed conversational interactions with Dozzz, indicating a universal positive inclination towards such interactions regardless of their exposure to smart speakers.

> *Dozzz likes my friend. I like to talk with her every night before sleeping. (C6)*
> *She looks beautiful. (C7)*

Subtheme 1.2: Redundant activities during speaking (Barriers). Three children above 9 years old preferred Dozzz's hands-free voice interaction over the button-based voice recording method. This highlights the importance of user-friendly interfaces for children of different ages.

> *I don't like to click each time. It feels a bit inconvenient. (C9)*

Theme 2: Enablers and barriers in sustained engagement with digital sleep diaries

Subtheme 1: Gamification as the continuous trigger for children (Enablers). About half of the children suggested adding gamification elements like point-based rankings and progress tracking to boost motivation and encourage regular app use. This indicates that gamification features and progress tracking can enhance engagement, promote commitment to the self-reporting process, and provide a sense of achievement for consistent sleep diary reporting.

> *I like 'Jumping Every Day'. Because I enjoy earning points every day. (C3)*
> *I use the learning app primarily to keep a record. I clock in everyday. (C5)*

One child highlighted the significance of social features in sustaining app usage. This indicates that incorporating features like progress sharing and friend interactions can boost children's motivation and engagement in self-reporting, fostering a sense of community and encouraging ongoing app use.

> *I love playing Pokémon Go on my phone almost every day because I can team up with friends. (C8)*

Subtheme 2: Parent's role in promoting sustained engagement (Enablers). Parent's habits, like exercise and eating, can shape children's long-term behavior, emphasizing the role of parental modeling and the family environment. Including family-related questions and promoting healthy behaviors in the self-reporting method can reinforce positive habits and offer a holistic view of children's well-being and sleep patterns.

> *My mother and I use 'Balance' for meditation every morning and night. It helps me sleep better, and I enjoy the music. (C10)*

Subtheme 3: Limitations in younger ages (Barriers). Children above 8 years old show more prolonged app usage, potentially due to greater digital device familiarity, while younger children may have shorter attention spans. Design considerations should accommodate these age-related differences.

> *I cannot. Because I do not use smartphones very often. (C6)*

4 Discussion

In this section, we discuss more potential applications of Dozzz as well as the current limitations and future work.

Feasibility of voice-based diaries and experience sampling. Dozzz, employing Google's voice recognition on Android, demonstrates the feasibility of a chatbot for children's voice-based sleep diaries, with potential applications in ambulatory self-reporting for health conditions and catering to user groups favoring voice input, such as individuals with motor impairments and older adults. While the potential is evident, it's important to note that the current feasibility evidence is limited to short-term sessions. First, while this study aimed to capture natural daily fluctuations, it was conducted in non-controlled settings like a public library and at home, potentially missing variations in the social and physical context of children's homes that could influence their independent system interactions [13]. Hence, long-term field studies are necessary to fully understand the impact of introducing voice-based response methods into a child's environment [15]. Second, this study demonstrates the feasibility of children using a voice-based sleep diary. However, the sample size of the study is very limited. Testing with a larger sample of children is essential to gain more confidence in their ability to use the system and uncover potential usability issues. Future studies will build upon this empirical evidence and address limitations. Third, to safeguard the data validity for children' sleep, we prefer they respond to the voice diary shortly after waking up. However, nine children preferred reporting their sleep at night before bedtime, possibly due to the convenience of interacting with media at that time. Future research should address this issue by reducing morning interaction requirements and allowing reports on daily sleep-related activities before bedtime.

Tangibles vs touchscreens. The voice-based self-reporting system developed in this study is currently implemented on the Android platform. However, requiring children to click the touchscreen for each response disrupts communication, particularly for younger kids. It is noteworthy that the percentage of children who own smart speakers at home (50%, 5/10) is lower than anticipated, while the penetration rate of smartphones (100%, 10/10) is higher among the participants in their daily lives. To address this issue, it is possible to further improve the system on smartphones through iterative design, enabling children to interact directly with voice-based diaries in a more natural communication context. This would involve refining the design to minimize distractions, enhance the usability of touchscreens, and create a more seamless and engaging interaction for young use.

Conditions that Influence continuous self-reporting. It is interesting to note that 90% of children in the experiment expressed a willingness to continue self-reporting with Dozzz, although the extent to which this positive attitude translates into regular reporting remains uncertain due to potential social desirability bias. This positive attitude can be attributed to several factors. First, children at this age are often driven by their developing interests, such as fantasy, individual play, and sensorimotor actions [14]. Dozzz's interactive nature, involving playful sounds and engaging conversations, aligns perfectly with these interests and offers an enjoyable experience for children [5]. Second, introducing gamification features has the potential to boost children's compliance and motivation. Participants in this study expressed enthusiasm for earning points, ranking, and tracking their progress, indicating a strong interest in gamified elements within the self-reporting experience. Based on these findings, future research could explore the integrating gamification elements with the voice-based response method to enhance children's proactive self-reporting. This integration can make the self-reporting process more engaging and motivating for children, encouraging their active participation in self-tracking activities.

5 Conclusion

This paper examined the feasibility of a voice-based sleep diary for children. To investigate this research objective, we developed a chatbot called Dozzz, which was built on the Android platform and utilized the Google recognizer engine to support voice input. The sleep diary was originally a text-based chatbot for children to self-report sleep information by typing answers to chatbot questions [9]. In Dozzz these questions and answers are supported by voice-based interactions.

A total of ten children aged 6 to 12 were recruited to participate in the experiment. The study involved two main components: 1) participants' experience with using Dozzz, 2) semi-structured interview with the children, focusing on three themes: their experience with Dozzz, their experience with electronic technologies in their daily life, and their daily routines. The observations and interviews provided initial evidence regarding the feasibility of voice-based sleep diaries, a positive attitude of children towards a voice-based diary. Future research should examine the independent use of the diary by children at home, and what interactive features are needed to sustain engagement and adherence over sustained reporting periods.

Reference

1. Danielle Pacheco, Nailing Vyas (2023). An introduction to the importance of sleep in children and how to help them sleep better. https://www.sleepfoundation.org/children-and-sleep
2. Eric Urban. (2023). Test accuracy of a custom speech model. https://learn.microsoft.com/en-us/azure/ai-services/speech-service/how-to-custom-speech-evaluate-data?pivots=speech-studio
3. Fox, C. B., Israelsen-Augenstein, M., Jones, S., & Gillam, S. L. (2021). An evaluation of expedited transcription methods for school-age children's narrative language: automatic speech recognition and real-time transcription. Journal of Speech, Language, and Hearing Research, 64(9), 3533-3548.
4. Meltzer, L. J., & Mindell, J. A. (2008). Behavioral sleep disorders in children and adolescents. Sleep Medicine Clinics, 3(2), 269-279.
5. Dele-Ajayi, O., Sanderson, J., Strachan, R., & Pickard, A. (2016). Learning mathematics through serious games: An engagement framework. Paper presented at 2016 IEEE Frontiers in Education Conference (FIE).
6. Maharjan, R., Rohani, D. A., Doherty, K., et al. (2022). What Is the Difference? Investigating the Self-Report of Wellbeing via Conversational Agent and Web App. IEEE Pervasive Computing, 21(2), 60-68.
7. CiteClegg-Kraynok, M., Barnovsky, L., & Zhou, E. S. (2023). Real, misreported, and backfilled adherence with paper sleep diaries. Sleep Medicine, 107, 31-35.
8. Riley, A. W. (2004). Evidence that school-age children can self-report on their health. Ambulatory Pediatrics, 4(4), 371-376.
9. Werner, H., Molinari, L., Guyer, C., et al. (2008). Agreement rates between actigraphy, diary, and questionnaire for children's sleep patterns. Archives of pediatrics & adolescent medicine, 162(4), 350-358.
10. Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. Annual Review of Clinical Psychology, 4,1–32.
11. Aarts, T., Markopoulos, P., Giling, L., et al. (2022, June). Snoozy: A Chatbot-Based Sleep Diary for Children Aged Eight to Twelve. In Interaction Design and Children (pp. 297-307).
12. Văcărețu, T., Batalas, N., Erten-Uyumaz, B., et al. (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 (pp. 113-122).

13. Beal, D. J. (2015). ESM 2.0: State of the art and future potential of experience sampling methods in organizational research. Annu. Rev. Organ. Psychol. Organ. Behav., 2(1), 383-407.
14. Markopoulos, P., Read, J. C., & Giannakos, M. (2021). Design of digital technologies for children. Handbook of human factors and ergonomics, 1287-1304.
15. Oakes, L. M., Newcombe, N. S., & 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, 279-294
16. Meltzer, L. J., Brimeyer, C., Russell, K., et al. (2014). The children's report of sleep patterns: validity and reliability of the sleep hygiene index and sleep disturbance scale in adolescents. Sleep Medicine, 15(12), 1500-1507.
17. Spruyt, K., & Gozal, D. (2011). Pediatric sleep questionnaires as diagnostic or epidemiological tools: a review of currently available instruments. Sleep medicine reviews, 15(1), 19-32.
18. Sturgess, J.,Rodger,S.,&Ozanne, A. (2002). A review of the use of self-report assessment with young children. British Journal of Occupational Therapy, 65(3), 108-116.
19. Henry AD (2000) Paediatric Interest Profiles: surveys of play for children and adolescents. USA: The Psychological Corporation, a Harcourt Assessment Company.
20. Carney, C. E., Buysse, D. J., Ancoli-Israel, S., et al. (2012). The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep, 35(2), 287-302.
21. Ceci, S. J., & Bruck, M. (1995). Jeopardy in the courtroom: A scientific analysis of children's testimony. American Psychological Association.
22. Vilaysack, B., Cordier, R., Doma, K., et al. (2016). Capturing everyday experiences of typically developing children aged five to seven years: A feasibility study of experience sampling methodology. Australian Occupational Therapy Journal, 63(6), 424-433.
23. Chen, Y. W., Cordier, R., & Brown, N. (2015). A preliminary study on the reliability and validity of using experience sampling method in children with autism spectrum disorders. Developmental Neurorehabilitation, 18(6), 383-389.
24. Csikszentmihalyi, M., Csikszentmihalyi, M., & Hunter, J. (2014). Happiness in everyday life: The uses of experience sampling. Flow and the foundations of positive psychology: The collected works of Mihaly Csikszentmihalyi, 89-101.
25. Lucero, A. (2015). Using affinity diagrams to evaluate interactive prototypes. In Human-Computer Interaction–INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14-18, 2015, Proceedings, Part II 15 (pp. 231-248). Springer International Publishing.
26. Ingram, J., & Elliott, V. (2014). Turn taking and 'wait time'in classroom interactions. Journal of Pragmatics, 62, 1-12.
27. Braun, V., & Clarke, V. (2012). Thematic analysis. American Psychological Association.
28. Clarke, V., & Braun, V. (2013). Teaching thematic analysis: Overcoming challenges and developing strategies for effective learning. The psychologist, 26(2), 120-123.
29. Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: a historical review and synthesis. Psychological bulletin, 113(3), 403.
30. Chen, S.: Dozzz source code and related documents (2023). https://github.com/Shanshan-css/Dozzz.git
31. Park, Y., Patwardhan, S., Visweswariah, et al. (2008). An empirical analysis of word error rate and keyword error rate. In 9th Annual Conference of the International Speech Communication Association. ISCA Archive.
32. Sari, C. C. (2020). Conversation Analysis: Turn-Taking Mechanism and Power Relation in Classroom Setting. Celtic: A Journal of Culture, English Language Teaching, Literature and Linguistics, 7(2), 118-136.