Designing Robot-Mediated Phonological Awareness Activities: Child-Centered Approach and Kindergarten Integration

Jasmin Azarfar[1], Utku Norman[2,*], Irina Rudenko[1] and Barbara Bruno[1]

***Abstract*— Long-term child-robot interaction (CRI) in kindergartens is key for integrating robots into children's daily lives and supporting their development. This unsupervised 4-weeks study explores how the robot Pepper can support phonological awareness (PA) in kindergarten children through the 'Sound Game', leveraging robot logs, questionnaires and open feedback by teachers and experts' analysis of audio recordings. Our findings reveal a generally high engagement, with the game being played 12 times, for a total of 233 minutes. Teachers played a crucial role in mediating the game, especially for younger children, and noted its potential for long-term use. The game proved beneficial across age groups, highlighting its potential as a first step for sustained, effective PA development and the transformative impact of social robots in early childhood education.**

Fig. 1. Sound Game observed during our kindergarten deployment.

I. Introduction

Long-term, in-the-wild child-robot interactions (CRI) in natural settings, such as kindergartens, homes, or schools, offer valuable insights and reveal new design opportunities for developing social robots that assist children in daily life [1]. These interactions, often unsupervised by developers, provide a deeper understanding of the actual value a robot can bring, the role it may play in a child's school or kindergarten routine, and how it is genuinely perceived. They also help clarify how real activities unfold, the extent to which the robot aligns with its intended role—how children and educators wish to use it versus what roboticists envision—and how its introduction reshapes the physical and social environment [1].

Although robot-assisted learning and CRI are widely studied research areas, much remains unknown about social robots' real-world capabilities [2], [3]. CRI research tends to rely on controlled laboratory environments rather than naturalistic user settings, and long-term in-the-wild studies remain relatively rare compared to the large number of short-term studies [3]–[5]. As a result, it remains uncertain whether—and how—findings from CRI research translate from lab conditions to real-world applications and whether lab-based studies truly capture the complexity of interactions as they develop over time [3].

Social robots have shown significant potential in early childhood development [6], specifically in supporting early language and literacy learning [7], [8]. Phonological awareness (PA), the ability to detect and analyze the spoken components of language [9], is a critical skill for preschool children and a strong predictor of future reading success [9], with researchers recommending introducing PA training as early as age four [10].

This study investigates how the robot Pepper can be used to support the development of PA in kindergarten children. We share our experience in designing and implementing robot-mediated activities in kindergarten for practicing PA, from prototype development to a four-week deployment at a kindergarten where the activities were used by teachers without developer supervision. Our goal is to create activities that are long-lasting, engaging for children, usable by teachers, and effective for PA development.

To reach this goal, we adopted a child-centered perspective, setting age- and developmentally appropriate interaction goals and formats, which are two key factors that impact child-robot interactions [5]. We leveraged the robot's communicative capabilities to best support children's PA activity. We expected this approach to shape the interaction in ways that are engaging and educational. Thus, we developed a game, the *Sound Game*, where children, together with the robot Pepper, guess the source of sounds, which we evaluated during a four-week deployment at a kindergarten—see Fig. 1.

The contribution of this paper is twofold:

1) The iterative design of a robot-mediated PA training activity for deployment in kindergarten settings.
2) The analysis of a four-week deployment of the activity in a kindergarten, leveraging teacher questionnaires, open feedback, and interaction patterns from game event logs and audio coding, to evaluate its impacts.

*This work was funded by the Baden-Württemberg Ministry of Science, Research and Art (MWK), using funds from the state digitalization strategy digital@bw. Corresponding author's email: utku.norman@kit.edu

[1]Socially Assistive Robotics with Artificial Intelligence (SARAI) Lab, Karlsruhe Institute of Technology (KIT), Germany

[2]Institute for Technology Assessment and Systems Analysis (ITAS), Karlsruhe Institute of Technology (KIT), Germany

This work addresses the following research questions:

- *RQ1: How do children engage with the robot-mediated phonological awareness activities?*
- *RQ2: Do children perform well and show improvement in phonological awareness through these activities?*
- *RQ3: Does the activity have the potential to be sustained over a long-term period?*
- *RQ4: Can the activity be naturally integrated into the kindergarten routine?*

Sec. II reviews long-term in-the-wild CRI and activities to develop PA skills. Sec. III details our activity design. Sec. IV describes our metrics for evaluating the RQs and the user study conducted at a kindergarten. The results are presented and discussed in Sec. V, and we conclude in Sec. VI.

II. Related Work

A. Long-Term "in the Wild" CRI

CRI is particularly relevant in education [7], [8], therapy [11], [12], and developmental support [13]. Despite established research [1], [14], much remains unknown about the real-world capabilities of social robots in these domains [2], [3]. Most CRI studies are still conducted in controlled environments, and long-term, naturalistic studies are relatively rare [3]–[5], although the robotics community emphasizes the need for research "in the wild"—conducting extended studies in everyday environments—to understand real-world functionality [1], [15], [16].

On their way "in the wild", social robots actively explore spaces like shopping malls [17], pedestrian flows [18], workplaces [19], museums, restaurants, and more [20]. Such studies provide valuable insights not only into the robot's capabilities but also into human behavior in response to the robot. However, high costs, reliability/safety challenges, and the time and expenses required for field studies [15] mean they are often short-term and primarily involve only adults.

In-the-wild and long-term studies with young children are preferably conducted in familiar environments like home, school, or kindergarten [1]. Home robots for kindergarten children focus on reading books and simple interactions such as storytelling, playing music, and casual conversations [21]. In schools and kindergartens, long-term CRI studies are typically centered around learning tasks, particularly in language learning [8], dance promotion [16], [22], or free play [23]. Despite positive results reported in some studies [13], [21], sustaining long-term engagement with social robots remains a significant challenge [3], [24], [25], due to a lack of engaging functions and disappointment in the robot's capabilities (e.g., limited speech recognition and response flexibility), and certain behaviors being perceived as repetitive or even annoying [1].

Recent studies explore new solutions to overcome these limitations by designing more interactive and user-driven interactions. [26] conducted a two-week home deployment study focusing on language processing in CRI. Results show that providing mixed-initiative dialogue capabilities was beneficial for the user's experience, as they learned how to use the system and changed their interaction from mostly system-driven in the first days to mostly user-driven by the end of the experiment. [14] presents an application enabling in-the-wild CRI with a large group of children interacting at the same time with the robot, capable to propose different activities (i.e., story reading, dance, game with rules). The design was based on reactive robot behavior, where instead of merely presenting content, the robot reacted and integrated itself into the children's activity. [3] conducted a long-term study in school settings investigating the potential of a self-regulated learning system. The system provided children with compact learning tasks for use in short episodes of learning, while being adaptable enough for repetition over a longer period. Four months of unsupervised practice with a balance scale task showed that children could effectively self-navigate the available difficulty levels, consistently progressing from easier to harder assignments. These findings highlight the potential for CRI in long-term, in-the-wild settings. Our work goes further by demonstrating practical and developmentally appropriate applications of unsupervised CRI in real-world kindergarten environments, addressing the critical need for sustainable and engaging activities.

B. Activities to Develop Phonological Awareness Skills

Phonological awareness (PA)—the ability to detect and analyze the spoken components of language [27]—plays a critical role in literacy development. The better children are at identifying, segmenting, and blending sounds in words, the more easily they can connect these sounds to the corresponding letters [27]. Early PA is a strong predictor of later reading success [28], and its lack can affect both academic achievement and self-esteem [29]. Teachers recommend starting PA training as early as age four, as reading and spelling difficulties often become evident only in first grade [29].

PA represents a continuum of cognitive operations that vary in complexity and require different cognitive resources [30]. Developing levels of PA include syllable awareness, onset-rime awareness, and phoneme awareness. Broader phonological awareness involves larger language units like rhymes and syllables, typically emerging spontaneously in kindergarten. In contrast, narrower phonological awareness requires conscious manipulation of phonemes and usually develops with school-based learning [31].

PA training programs for young children usually consist of activities that progress from simple rhyming tasks to more complex exercises, such as working with syllables and individual phonemes [29], [32], preparing children for reading and writing. Special programs are developed for specific groups of children, such as those at risk for reading disabilities, older children with dyslexia, and children with spoken language impairments or complex communication needs [30]. Literature [33] reports on Information and Communication Technologies (ICT) promoting PA in preschoolers, suggesting a wide spectrum of reading or spelling tasks and games, as well as books for practicing reading skills.

The only reported study using robot-assisted PA training was conducted in [34] with fourteen Korean children (ages

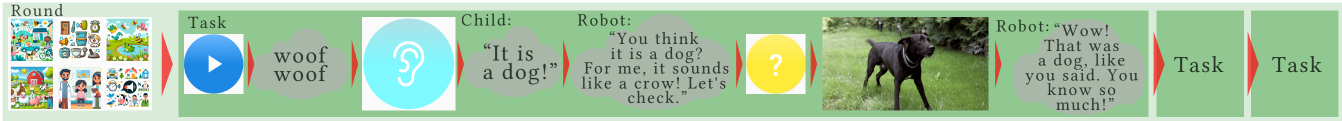

Fig. 2. Game flow: theme selection, press button to listen to a sound, robot's hearing touchpoint for answer, robot acknowledges child's answer and gives its own answer, hint button plays a video (e.g., dog barking), and the robot reflects on the answers and provides encouraging feedback. Theme images were AI-generated using OpenAI DALL·E. Themes from top-left: on the street, at home, in nature, on a farm, by the doctor, and mixed.

6–8) with reading difficulties. They were divided into three groups: traditional PA training, robot-assisted PA training, and a control group. Over 23 sessions, held twice a week for 30 minutes each, PA training included activities such as invented spelling, syllable discrimination, and blending of syllables and phonemes. All activities were conducted using the "Letter popcorn program"—on robot iRobiG's monitor, letters were hidden or covered by a popcorn shaped icon. The robot spoke a target word or syllable or phoneme. The child listened to it carefully and touched the letter popcorns in the correct sequence. Results showed that both PA training groups (traditional and robot-assisted) significantly improved their word spelling skills compared to the control group, which received no training. However, no significant differences were found between the two training methods.

Our study aims to apply robot-mediated PA activities to kindergarten children who have not yet started formal literacy instruction. The program outlined in [29] provided foundational insights for designing our activities. It consists of games and exercises across six areas that build on each other, helping children understand the structure of spoken language. It begins with eavesdropping games that aim to train the children's hearing and naming for sounds in their environment. Rhymes are then introduced, followed by concepts of sentences and words, syllable awareness, identifying initial sounds in words, and finally working with phonemes. Our current paper presents an activity related to the first step of the program. Ongoing work is tackling the second step, which targets rhyming and encourages children to recognize language through sound rather than meaning. We hope that the results from both steps will provide valuable insights for refining and advancing the design of PA practice with robots.

III. Activity Design

A. Sound Game Scenario

The game's goal is to introduce children to a variety of sounds as an initial step in practicing PA skills.

1) Child-Centered Design: To develop effective child–robot interaction (CRI), we adopted a child-centered perspective, setting age- and developmentally appropriate interaction goals and formats, which are key factors that impact CRI [5]. Leveraging the robot's communicative capabilities, we aimed for interactions that best support children's learning by being intuitive, engaging, sustainable, and educational. Our focus on PA in kindergarten-aged children was motivated by its crucial role in literacy acquisition and by the general recommendation to begin PA training around the age of four (see Sec. II-B). Additionally, teachers in our partner kindergarten expressed a need for PA-related activities, particularly rhyming.

Given that *play* is the leading activity at this stage of development, we designed the CRI accordingly—as a game in which children, together with the robot Pepper, encounter a mysterious sound, hypothesize about its source, and then watch a video to verify their guesses. The game, played on the robot's tablet, is structured to suit the attention span and engagement patterns of the target age group.

Thus, Pepper acts as a peer, maintaining interaction by agreeing or disagreeing with the children's ideas, making right or wrong guesses, and expressing emotional reactions based on the accuracy of their answers (e.g., praising, expressing regret). To keep children's attention, the game is divided into short episodes that children can choose from, featuring different locations where they can hear sounds. Kindergarten teachers confirmed that the game's pacing, number of tasks, and Pepper's speech rate were age-appropriate. Periodic prompts to check whether to continue allow children to stop the game any time. Physical actions, such as pressing a button to listen to a sound, making guesses, watching videos, and choosing the next sound location, keep the game dynamic and encourage children to directly steer the game. This game format is also well-suited for group play (preferred by both kindergarten children and teachers), allowing children to take turns.

2) Structure: The game has three hierarchical levels:

- **Task**: The basic unit where children identify sounds.
- **Round**: A round consists of *three* tasks within a selected theme.
- **Session**: The overall activity with one or more rounds.

3) Interaction Sequence: The robot orchestrates the game interaction through the tablet, as illustrated in Fig. 2:

1) **Welcome and Demo**: The robot greets the children and provides a demo explaining the game during the first interaction. This demo is skipped by default but can be (re)played at the teachers' or children's discretion.
2) **Theme Selection**: Children pick the theme for the first round. They can go back to the theme selection at any action point (e.g., pressing to listen) during the game.
3) **Rounds**: Children engage in one or more rounds, deciding whether to continue or end the game after each. The sound categories within each theme are chosen randomly from the audio and video dataset described in Sec. III-B, while ensuring no repetition in the immediate next round.

4) Task Flow: Each task follows these steps (see Fig. 2)[1]:

1) **Listen**: Children press the blue button to hear a sound.
2) **Answer**: Children press a touchpoint shaped like an ear representing Pepper's 'special ear' to indicate readiness to answer, which blinks while the robot listens. The robot detects all words but filters and recognizes the last non-stop word as the answer, comparing it to a list of synonyms for the sound category.
3) **Robot's Answer**: The robot internally makes a guess (50% correct, 40% incorrect, and 10% unsure). It acknowledges the answer it recognized, and communicates whether it agrees, disagrees, or is uncertain.
4) **Hint**: Children press the yellow button to play a video revealing the sound's source.
5) **Robot's Reflection**: The robot provides feedback, indicating correct answers and encouraging the children.

B. Audio and Video Dataset

To support the game's objectives, we curated a dataset of environmental sounds and video excerpts, designed to be extendable with new sounds, sound categories, and themes:

- **Sounds**: Builds upon the ESC-50 Dataset [35] with 2000 clips across 50 classes of common sound events. We selected 30 classes as sound categories with 10 sounds each, resulting in *300 sound clips*, each 5s long. Sounds are post-processed by trimming silence and repeating cuts to *7 s* for better audibility.
- **Videos**: Includes *150+ video clips* visualizing the sound categories, sourced from free-licensed platforms.[2].
- **Themes**: Groups sound categories into themes (e.g., "on the street", "in nature"). We formulated *5 themes and a randomized theme* ("sounds mixed together").
- **Synonyms**: Lists synonyms for each sound to allow alternative naming for the same sound category.

C. Setup Design

1) Hardware Setup: The setup includes a Pepper robot (v. 1.8) integrated with a Samsung Galaxy S8 tablet, mounted on top of the Pepper's own tablet using custom 3D-printed brackets to create a unified system. Both devices are connected to the local network via a dedicated portable router.

2) Software Setup: Pepper operates in autonomous life mode with enhanced tablet reachability. The game application, developed in Java for Android, uses the Pepper SDK (QiSDK, NAOqi 2.9), for robot control. Communication between the tablet and robot is established via SSH on the same network, using a custom version of the Pepper SDK.

Earlier testing showed Pepper SDK's recognition was inadequate for open-ended responses. We chose the Vosk Offline Speech Recognition[3] for its low-latency, open-ended recognition capabilities, ensuring variability and reliability, and competitive performance on a low compute budget [36].

IV. Study at the Kindergarten

A. Participants and Procedure

The study took place over four weeks in a kindergarten in Germany.[4] Initially, the robot's deployment was planned for two weeks, during which 38 children aged 3–6 participated: 6 were 3 years old, 15 were 4–5 years old, and 17 were 6 years old.[5]

Children played in age-mixed groups of up to six participants, with group members determined ad hoc based on children's preferences before each session. Two exceptions in group formation occurred when one teacher conducted sessions separately with children aged 3–4 and 5–6.

Three teachers participated during the first two weeks. A single one-hour introductory session was conducted for them. The teachers completed questionnaires at the end of each week for the age groups: 3, 4–5, and 6 years old. After the initial two weeks, Pepper remained in the kindergarten for an additional two weeks without extra requests to the teachers. Robot logs indicated continued use and new teacher participation, though no questionnaires were filled out during this extended period. Data from their game-play was included in the analysis. No experimenters were ever present except as invited observers in two sessions at the beginning of deployment.

B. Evaluation Metrics

Children's interactions with the robot and the game were evaluated using teachers' expertise (questionnaire and feedback), audio data coding, and robot logs. The coding scheme was developed by researchers with expertise in child psychology and human-robot interaction studies, the coding was conducted by a native speaker of the children's language and verified by an experienced researcher. The evaluation metrics are grouped according to the RQs for a structured analysis.

1) RQ1 on children's engagement with the activity: Teachers rated the following questionnaire items:

- **Motivation (Q1)**: What motivated children to play with the robot (Teacher, Group, Own)
- **Comfort (Q2)**: Children's comfort level during interactions with the robot (Shy, Reserved, Relaxed)
- **Difficulty (Q3)**: Game difficulty (Easy, Medium, Hard)
- **Interest (Q4)**: Level of interest during the game (High, Moderate, Low)
- **Response to Mistakes (Q5)**: Children's reactions to their own mistakes (Positive, Neutral, Negative)
- **Response to Success (Q6)**: Children's reactions to their own success (Positive, Neutral, Negative)
- **Response to Correction (Q7)**: Reactions to the robot's corrections in its reflection (Positive, Neutral, Negative)

Children's engagement was further examined by coding their verbal actions during the game, focusing on reactions to game events and the generation of ideas about the sounds.

[1]See the game in a lab demo at `https://youtu.be/p4MBqnw0rlo`.

[2]like Pexels via the Pexels API using Pexels' ranking algorithm, then filtered for appropriateness and complemented by manual downloads.

[3]`https://alphacephei.com/vosk/`, German small model v. 0.15

[4]Ethical approval was granted by KIT Ethics Committee, No. A2024-035.

[5]The three groups reflect the structure of the German kindergarten system, that identifies activities for newcomers (aged 3), children aged 4–5 and preschoolers (children nearing age 6).

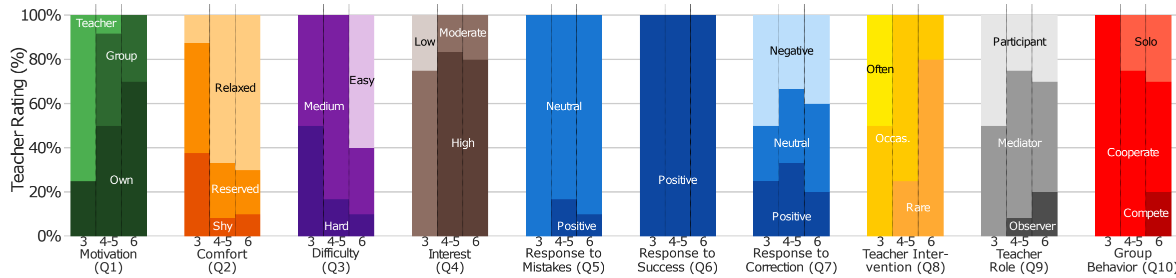

Fig. 3. Questionnaire results from three teachers showing the distribution of their responses to the items for age groups 3, 4–5, and 6 (preschool).

We assume that both metrics indicate active participation of children in the game:

- **Reactions**: Vocal expressions conveying attitude, opinion, or feedback toward their own or Pepper's actions.
- **Ideas Generation**: Guesses about the sound expressed immediately after hearing it, prior to providing the final answer to the robot. Each task is uniquely coded as:
  - **Multiple Voices**: Several children expressed ideas.
  - **One Voice**: Only one child expressed their idea.
  - **No Voice**: The teacher was the first to speak.

We computed the number of codes per task.

2) RQ2 on children's performance:

- **Right Answer**: (Coded) Correct answers by the children
- **Recognized Answer**: Correct answers of the children that are also correctly recognized by the robot (logs)

We compared these metrics, as well as other metrics from logs and audio, for game sessions with tutorial (first-time interactions), and sessions without tutorial.

3) RQ3 on long-term potential: The long-term potential of the game is assessed using longevity metrics derived from event logs, focusing on game content (categories & sounds):

- **Novelty**: Proportion of new content introduced in a session compared to previous sessions.
- **Variety**: Proportion of unique content within a session.
- **Explored**: Cumulative rate of content explored across all previous sessions.

4) RQ4 on integration into kindergarten: We determined the game's smooth conduct, the extent of teacher involvement needed to keep children engaged, its suitability for kindergarten-aged children, and its potential for future use based on teachers' experiences. Teachers rated the following:

- **Teacher Role (Q8)**: Observer, Mediator, Participant
- **Teacher Intervention (Q9)**: Frequency of teacher involvement (Rare, Occasional, Often)
- **Group Behavior (Q10)**: Children's social behavior, evaluating the game's potential to encourage both group and individual participation (Solo, Cooperate, Compete)

Teachers' actions during the game were further coded for:

- **Support**: Feedback on children's performance and emotional support on misunderstandings by the robot.
- **Explanation**: Explain game rules, encouraging teamwork, alternating answering and button pressing.
- **Hints**: Help children come up with answers by giving hints or guiding them to a decision, encouraging them to answer based on their own or group discussions.

V. Results and Discussion

There were 12 sessions over four weeks, totaling 233 min ($M = 19.4$, $SD = 25.2$, range: 5.7 to 37.7 min). The sessions included 154 tasks across 55 started rounds and 51 completed ($M = 4.6$ rounds, 1–10 rounds per session). Task durations averaged 98 s ($SD = 25.2$, from 37.1 to 132.4 s).

A. RQ1 on Children's Engagement with the Activity

1) Motivation to play: Older children were mostly self-motivated, while 4–5-year-olds were equally influenced by their peers and their own interest (Fig. 3, Q1). Younger children, however, were more likely to follow teachers' invitations, as robots were completely unfamiliar to them.

2) Comfort level: Children aged 4-6 were mostly relaxed, whereas younger children were more shy and reserved around Pepper (Fig. 3, Q2). Teachers observed that younger children hesitated at first, struggling to categorize Pepper as either familiar or unfamiliar. Pepper's height also seemed imposing for smaller children. However, as they interacted with Pepper more frequently, their confidence grew—for example, they began approaching Pepper to press buttons and answer into its ear.

3) Game difficulty: Teachers rated the difficulty level as mostly easy for older children, mostly moderate for 4–5-year-olds, and equally hard and moderate for younger children (Fig. 3, Q3). The primary challenge for children was articulation, which often made it difficult for Pepper to correctly recognize their responses. Many children were shy (especially at the beginning), spoke briefly and quietly, or answered while moving around, sometimes directing their responses away from Pepper. Beyond the inherent limitations of automatic speech recognition, children's speech patterns further complicated the recognition process. This issue was particularly pronounced among younger children and non-native speakers. However, this difficulty also created a learning opportunity. Teachers encouraged children to focus on their pronunciation and speak more clearly to ensure Pepper could understand them. This additional challenge proved beneficial—in some sessions children's speech clarity improved, and Pepper's speech recognition accuracy increased up to 33%—twice higher than average, see Fig. 4.

4) Interest during the game: Children aged 4–6 displayed high levels of interest in the game, while younger children showed moderate interest (Fig. 3, Q4). Despite their interest, teachers noted that children were often frustrated when

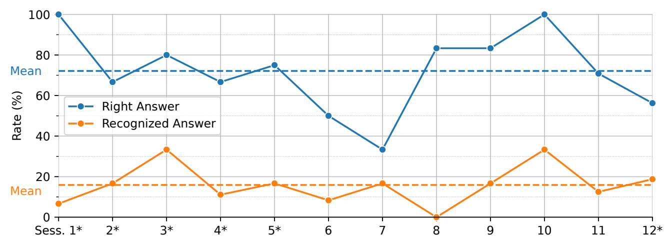

Fig. 4. The evolution of right and recognized answer rates through sessions. An asterisk (*) indicates the game was run with the tutorial.

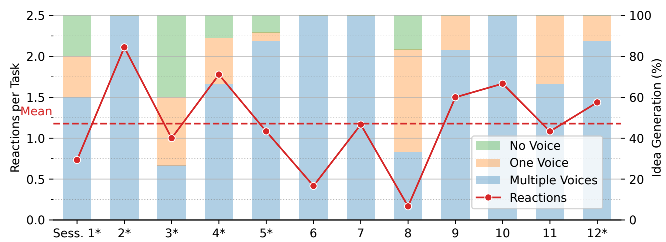

Fig. 5. Evolution of reaction rates and idea generation distributions.

Pepper failed to understand them. However, they remained eager to continue playing, likely motivated by the belief that improving their pronunciation would lead to better recognition. This challenge may have contributed to a unique form of sustained interest in the game.

5) Responses to performance and Pepper's utterances: Children of all ages were not particularly concerned about making wrong guesses but reacted positively to correct answers and Pepper's encouraging feedback (Fig. 3, Q5–7). However, when Pepper disagreed and suggested a different answer, children's reactions varied depending on whether Pepper had correctly recognized their response. If recognition was incorrect, children responded with negative comments or laughter.

6) Reactions to the game events: Children's engagement with the game was evident through their reactions, averaging 1.18 per task ($SD = 0.59$, range: $0.17 - 2.11$), as shown in Fig. 5. This supports the overall engagement levels observed.

7) Idea generation: The ideas were in majority multiple voices ($M = 76.0\%$, $SD = 25.8\%$), followed by one voice ($M = 16.0\%$, $SD = 16.5\%$), and only little no voice ($M = 8.0\%$, $SD = 12.4\%$), where the children did not know the answer or were too shy to answer, see Fig. 5.

The characteristics of children's interaction with the robot, as noted by the teachers and reflected in the audio coding, demonstrate high engagement, evident in both their willingness to solve tasks and their active reactions to game events. For 3-year-old children, the game was quite challenging—both in terms of completing tasks and following the game procedure—as also revealed by their heightened sense of tension around the robot, whose realistic perception they are not yet fully prepared for at this age [37].

B. RQ2 on Children's Performance

Since the goal of the game was to introduce children to the world of sounds (as a precursor to future work with rhymes), we can evaluate their performance based on whether they improved in identifying and naming sounds over time. An additional measure of the game's impact can be derived from the number of correctly recognized answers, as awareness of proper pronunciation also potentially implies that a child can accurately hear both the word and their own speech. To assess this impact, we compared the results of two groups of sessions—those conducted with a tutorial (indicating the initial stage) and those conducted without a tutorial (representing repeated game-play). Indeed, valuable insights can be gained by examining what has changed and what has remained the same.

TABLE I

COMPARISON OF GAMES WITH(OUT) TUTORIAL AND AGE GROUP. DIFFERENCES DISCUSSED IN THE TEXT ARE HIGHLIGHTED IN BOLD.

Measure	**With Tutorial (Sess. 1–5, 12)**	**No Tutorial (Sessions 6–11)**	**Older (Sess. 2)**	**Younger (Sess. 5)**
Time	26:54	12:00	27:10	37:40
Task Time	100.7 s	95.3 s	90.6 s	94.2 s
Right Ans.	74.1%	70.1%	66.7%	75.0%
Recog. Ans.	17.2%	14.6%	16.7%	16.7%
Support	**0.44**	**0.15**	0.44	0.54
Explanation	0.32	0.43	**0.39**	**0.25**
Hint	0.19	0.19	**0.11**	**0.25**
Reactions	**1.50** per task	**1.00** per task	**2.11**	**1.08**
Mult. Voices	71.4%	80.6%	100%	87.5%
One Voice	15.4%	16.7%	0%	8.3%
No Voice	13.2%	2.8%	0%	4.2%

Table I shows that both answer accuracy (74.1% vs. 70.1%) and recognition (17.2% vs. 14.6%) levels remain nearly unchanged. It is possible that different age groups improved at different rates. For example, older children contributed more at the initial stage and maintained their level (as they were already familiar with and could recognize many sounds), while younger children improved over time. However, since we only have the average results of the age-mixed groups, we cannot confirm this. The idea generation rate remains high (Multiple Voices: 71.4% vs. 80.6%), indicating that children continue to eagerly share their thoughts about sounds and enjoy guessing as an activity. At the same time, they have become more confident and accustomed to Pepper's imperfections, reducing the need for teacher support when they feel disappointed by Pepper's misunderstandings (0.44 vs. 0.15 per task). Interestingly, while children remain engaged in generating ideas, their reactions to game events have decreased (1.50 vs. 1.00 per task), which may also indicate an increased awareness of the game's nuances.

C. RQ3 on the Long-term Potential

1) Novelty: Fig. 6 shows that sounds are explored only up to 38% after about 4 hours of gameplay, and no child would have encountered all sounds even if they played all the games themselves. The novelty of the sounds remains steady ($M = 77.1\%$, $SD = 18.5$), indicating that new sounds continue to be introduced over time.

2) Variety: Unique sounds and categories within a game were high ($M = 92.7\%$, $SD = 8.5\%$), providing evidence

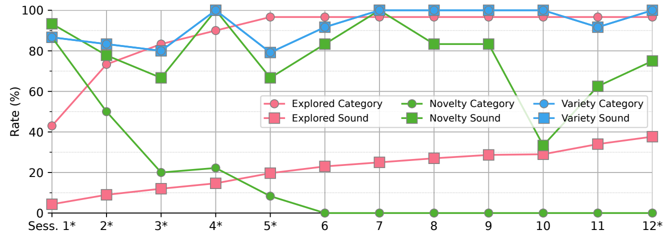

Fig. 6. The evolution of longevity metrics per category and sound.

that the game offers sufficient variety.

3) Exploration: Categories are explored up to 97%, with the clapping category (only in the random theme) not picked by random sampling. This suggests that while most categories are well-represented, some may need further attention to ensure balanced variety.

We conclude that the game offers long-term potential, with novel sounds introduced for extended gameplay. Since the game focuses on sounds, the diminishing novelty of categories is not a concern. However, there is high potential to extend the dataset, possibly via an online connection with mechanisms for appropriateness in terms of matching the sound and being suitable for the children.

D. RQ4 on the Integration into Kindergarten

1) Teacher's role in the game and need for intervention: For children aged 4–6, teachers primarily took on the role of mediators, highlighting key aspects of the game and guiding children's attention (Fig. 3, Q9). With younger children, however, teachers frequently had to participate actively as equal players. For 3-year-olds they also often had to intervene to maintain the game (e.g. giving hint, encouraging children to guess or to answer). The need for intervention was occasional for 4–5-year-olds and rare for older children. Teachers reported that the most common reason for intervention was when Pepper failed to recognize children's correct answers. In such cases, teachers reassured the children that their response was correct and that Pepper had misunderstood. Notably, one of the games on the third week (Session 11, see e.g., Fig. 5) was played by children entirely on their own without teacher support (though the teachers were present in the room).

2) Children's group behavior during the game: Children across all age groups displayed cooperative behaviors, such as taking turns, discussing guesses and choices, and reacting collectively to game situations (Fig. 3, Q10). Teachers observed that older children functioned as a team, while younger children also communicated with each other but were more focused on themselves and the teachers. They showed strong interest in physical interaction and often wanted to touch Pepper. Interestingly, 6-year-olds showed interest in playing individually with Pepper and even competing against it or other children. This could represent the next stage of game difficulty—children taking turns but playing for themselves, engaging one-on-one with Pepper.

3) Game appropriateness considering age: Fig. 3 clearly demonstrates that children of different ages engage with the game in different ways, potentially finding various interests for themselves. We can also observe this in Table I, which compares two game sessions of two groups of children. Unlike other sessions, the children in these sessions were divided by age—3–4 (Session 5, as on Fig. 3) years and 5–6 years (Session 2, Fig. 3), with five children in each group. The data shows that although both groups provided nearly the same number of correct answers (67% vs. 75%), the younger children received twice as many hints from the teachers (hints rate 0.11 vs. 0.25 per task). They were less concerned with the game rules (explanation rate 0.39 vs. 0.25 per task) and half as engaged in reacting to game events (reaction rate 2.11 vs. 1.08). Nevertheless, they were just as eager to generate ideas about the sounds (multiple voice rate 100% vs. 88%). We see it as a positive aspect that the game can be beneficial for children of different ages in different ways, allowing them to participate year after year in kindergarten, gradually developing from initial hesitation and shyness toward the robot to eventually competing with it.

4) Game usability and integration into kindergarten routine: Initially, Pepper's deployment in the kindergarten was agreed upon for two weeks, but we left it for an additional two weeks without any extra requests to the teachers. Robot logs revealed that teachers continued playing during this extended period, and even new teachers joined the experience. During deployment, two technical issues required developer visits—one for network reconfiguration, another to reboot the robot. Separately, interaction errors occurred when the robot responded to teachers instead of children, due to dialogue sequencing that failed not distinguish adult from child input.

Teachers suggested that in their daily practice, the game could be used once or twice a week, offering valuable benefits for children's future interactions with both people and robots. These benefits, among others, include concentration training, speech improvement, adapting speech when addressing others, building confidence, and practicing social skills (e.g., collaborative behavior).

Teachers confirmed their interest in continuing the activity with Pepper, as evidenced by their identification of a few areas for potential improvement in the game for future use. They also noted: "The weeks with the robot were an exciting and enriching experience for both us and the children, and we really liked the overall concept of the game. We see the potential of the project and appreciate that children can engage with robotics and technology in a playful way".

VI. Conclusions and Future Work

This paper aims to explore long-term child-robot interaction (CRI) as it unfolds in real-world settings. We present the results of designing, deploying for four weeks, and evaluating a robot-mediated activity ("Sound Game"), envisioned to be the first step in a planned series of activities for practicing phonological awareness (PA) with kindergarten children aged 3–6. The activity was used exclusively by kindergarten teachers in natural environments.

We adopted a child-centered perspective and applied longevity-focused CRI design principles to create an activity that is long-lasting, engaging for children, usable by teachers, and effective for PA development. To assess whether the designed activity met these goals, we analyzed data on children's behavior and teachers' experiences during their game sessions, gathered through teacher expertise (questionnaires and feedback), audio data coding, and robot logs.

Overall, a high level of children engagement was evident. Only for 3-year-old children the game appeared quite challenging—both in terms of completing tasks and following the game procedure. Children's performance dynamics indicated that they remained eager to share their thoughts about sounds, enjoyed guessing as an activity, and demonstrated a high rate of sound identification. Additionally, our study confirmed the long-term potential of the game, as well as its usability for teachers. The teachers' role was mainly that of a mediator, with minimal to moderate intervention needed.

We hope that insights from this work will contribute to refining and advancing the design of PA practice with robots.

References

[1] J. E. Michaelis, B. Cagiltay, R. Ibtasar, and B. Mutlu, ""Off script:" design opportunities emerging from long-term social robot interactions in-the-wild," in *HRI '23*, 2023, pp. 378–387.

[2] A. Ahtinen and K. Kaipainen, "Learning and teaching experiences with a persuasive social robot in primary school–findings and implications from a 4-month field study," in *PERSUASIVE 2025*. Springer, 2020.

[3] D. P. Davison, F. M. Wijnen, V. Charisi, J. Van Der Meij, V. Evers, and D. Reidsma, "Working with a social robot in school: a long-term real-world unsupervised deployment," in *HRI '20*, 2020, pp. 63–72.

[4] V. Charisi, D. Davison, D. Reidsma, and V. Evers, "Evaluation methods for user-centered child-robot interaction," in *RO-MAN 2016*. IEEE, 2016, pp. 545–550.

[5] I. Rudenko, A. Rudenko, A. J. Lilienthal, K. O. Arras, and B. Bruno, "The child factor in child–robot interaction: discovering the impact of developmental stage and individual characteristics," *Int. J. Soc. Robot*, Aug. 2024.

[6] L. P. E. Toh, A. Causo, P.-W. Tzuo, I.-M. Chen, and S. H. Yeo, "A review on the use of robots in education and young children," *Educ. Technol. Soc*, vol. 19, no. 2, pp. 148–163, 2016.

[7] J. Kanero, V. Geçkin, C. Oranç, E. Mamus, A. C. Küntay, and T. Göksun, "Social robots for early language learning: Current evidence and future directions," *Child Development Perspectives*, vol. 12, no. 3, pp. 146–151, 2018.

[8] M. M. Neumann, "Social robots and young children's early language and literacy learning," *Early Child. Educ. J.*, vol. 48, no. 2, Mar. 2020.

[9] C. E. Snow, M. S. Burns, and P. Griffin, Eds., *Preventing Reading Difficulties in Young Children*. Washington, D.C.: National Academies Press, 1998.

[10] M. E. Reynolds, K. Callihan, and E. Browning, "Effect of instruction on the development of rhyming skills in young children," *Contemporary Issues in Communication Science and Disorders*, vol. 30, no. Spring, pp. 41–46, Mar. 2003.

[11] M. Alemi, A. Ghanbarzadeh, A. Meghdari, and L. J. Moghadam, "Clinical application of a humanoid robot in pediatric cancer interventions," *Int. J. Soc. Robot*, vol. 8, pp. 743–759, 2016.

[12] L. Cañamero and M. Lewis, "Making new "new AI" friends: Designing a social robot for diabetic children from an embodied AI perspective," *Int. J. Soc. Robot*, vol. 8, no. 4, pp. 523–537, 2016.

[13] B. Scassellati, L. Boccanfuso, C.-M. Huang, M. Mademtzi, M. Qin, N. Salomons, P. Ventola, and F. Shic, "Improving social skills in children with asd using a long-term, in-home social robot," *Science Robotics*, vol. 3, no. 21, p. eaat7544, 2018.

[14] E. Coronado, X. Indurkhya, and G. Venture, "Robots meet children, development of semi-autonomous control systems for children-robot interaction in the wild," in *ICARM 2019*. IEEE, 2019, pp. 360–365.

[15] M. Jung and P. Hinds, "Robots in the wild: A time for more robust theories of human-robot interaction," in *THRI*, vol. 7. ACM, 2018.

[16] F. Tanaka, J. R. Movellan, B. Fortenberry, and K. Aisaka, "Daily HRI evaluation at a classroom environment: Reports from dance interaction experiments," in *HRI '06*, 2006, pp. 3–9.

[17] M. Tonkin, J. Vitale, S. Ojha, M.-A. Williams, P. Fuller, W. Judge, and X. Wang, "Would you like to sample? Robot engagement in a shopping centre," in *RO-MAN 2017*. IEEE, 2017, pp. 42–49.

[18] H. Kidokoro, T. Kanda, D. Bršćić, and M. Shiomi, "Will I bother here? - A robot anticipating its influence on pedestrian walking comfort," in *HRI '13*, Mar. 2013, pp. 259–266.

[19] M. K. Lee, S. Kiesler, J. Forlizzi, and P. Rybski, "Ripple effects of an embedded social agent: a field study of a social robot in the workplace," in *CHI '12*, 2012, pp. 695–704.

[20] M. Hellou, J. Lim, N. Gasteiger, M. Jang, and H. S. Ahn, "Technical methods for social robots in museum settings: An overview of the literature," *Int. J. Soc. Robot*, vol. 14, no. 8, pp. 1767–1786, 2022.

[21] Z. Zhao and R. McEwen, ""Let's read a book together": A long-term study on the usage of pre-school children with their home companion robot," in *HRI '22*. IEEE, 2022, pp. 24–32.

[22] G. Venture, B. Indurkhya, and T. Izui, "Dance with me! Child-robot interaction in the wild," in *ICSR 2017*. Springer, 2017, pp. 375–382.

[23] E. Hyun and H. Yoon, "Characteristics of young children's utilization of a robot during play time: A case study," in *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 2009, pp. 675–680.

[24] M. M. De Graaf, S. Ben Allouch, and J. A. Van Dijk, "Long-term evaluation of a social robot in real homes," *Interaction Studies*, vol. 17, no. 3, pp. 462–491, 2016.

[25] A. Weiss, A. Pillinger, and C. Tsiourti, "Merely a conventional 'diffusion' problem? On the adoption process of Anki Vector," in *RO-MAN 2021*. IEEE, 2021, pp. 712–719.

[26] G. Pérez, G. A. García, M. Castro, M. Castaño, M. J. L.-G. De Quevedo, R. Durán, J. G. Amores, G. Álvarez, L. Levinson, S. Šabanović, *et al.*, "Child-robot conversation in the wild wild home: A language processing user study," in *RO-MAN 2023*. IEEE, 2023.

[27] S. Grofčíková and M. Máčajová, "Rhyming in the context of the phonological awareness of pre-school children," *Center for Educational Policy Studies Journal*, vol. 11, no. 1, pp. 115–138, 2021.

[28] S. A. Stahl and B. A. Murray, "Defining phonological awareness and its relationship to early reading." *J. Educ. Psychol.*, vol. 86, 1994.

[29] P. Küspert and W. Schneider, *Hören, lauschen, lernen: Sprachspiele für Kinder im Vorschulalter. Würzburger Trainingsprogramm zur Vorbereitung auf den Erwerb der Schriftsprache [Hear, Eavesdrop, Learn]*. Göttingen, Germany: Hogrefe, 2018.

[30] G. T. Gillon, *Phonological awareness: From research to practice*. Guilford Publications, 2017.

[31] H. Skowronek and H. Marx, "The Bielefeld longitudinal study on early identification of risks in learning to read and write: Theoretical background and first results," *Children at risk: Assessment, longitudinal research, and intervention*, pp. 268–294, 1989.

[32] B. A. Blachman, E. W. Ball, R. Black, and D. M. Tangel, *Road to the Code: A Phonological Awareness Program for Young Children.* ERIC, 2000.

[33] F. A. Fernández-Otoya, M. Raposo-Rivas, and A. X. Halabi-Echeverry, "A qualitative systematic literature review on phonological awareness in preschoolers supported by information and communication technologies," *Education Sciences*, vol. 12, no. 6, p. 382, 2022.

[34] E. J. Mun, Y. T. Kim, M. K. Kang, and S. J. Yeon, "Effect of robot-assisted phonological awareness training on invented spelling for children with reading disabilities," *Communication Sciences & Disorders*, vol. 22, no. 1, pp. 76–89, 2017.

[35] K. J. Piczak, "ESC: Dataset for environmental sound classification," in *Proceedings of the 23rd ACM international conference on Multimedia*, ser. MM '15. New York, NY, USA: ACM, Oct. 2015, pp. 1015–1018.

[36] A. Andriella, R. Ros, Y. Ellinson, S. Gannot, and S. Lemaignan, "Dataset and evaluation of automatic speech recognition for multilingual intent recognition on social robots," in *HRI '24*. ACM, Mar. 2024, pp. 865–869.

[37] I. Rudenko, U. Norman, R. Maure, A. Rudenko, N. Weinberger, F. Krebs, F. Peller-Konrad, F. Reister, T. Asfour, and B. Bruno, "Drawings for insight on preschoolers' perception of robots," in *HRI '24*, 2024, pp. 920–924.