Impact of Multi-Modal Robot Interaction on Children's Math Performance and Emotional Expression in Online Learning

Devasena Pasupuleti[1,a)] Hamed Mahzoon[2,b)] Kazuki Sakai[1,c)] Hiroshi Ishiguro[1,d)]
Yuichiro Yoshikawa[1,e)]

概要：Incorporating robots into educational settings such as e-learning platforms offers opportunities to enhance children's learning experiences. This study investigates how mixed-modality robots, consisting of physically and virtually embodied robots, influence children's learning outcomes, focus, and emotional responses within a Math e-learning environment. We conducted a pilot study with Japanese children aged 9-10 years, comparing their interaction with autonomous multi-modal embodied agents versus voice-only, non-embodied agents. While both conditions were perceived as helpful for maintaining focus and improving Math performance, no significant differences were observed between them. However, the embodied agents elicited significantly more positive emotional reactions, such as excitement and smiles, whereas the non-embodied agents resulted in negative expressions such as frowns and boredom. These findings highlight the potential of utilizing the advantages of combining physical and virtual agents to improve the quality of education in e-learning platforms.

1. Introduction

Academic self-esteem (ASE), defined as how well individuals perceive their performance in subjects like Math and English, develops at age 6 and becomes difficult to influence after age 12 [1]. Poor ASE leads to school dropouts, low attendance, and weak academic skills [2]. In non-western nations emphasizing academic achievement, children face increased pressure and low self-esteem [1]. While improving Math grades increases confidence, technologies to positively affect ASE remain under-explored. Beyond performance gains, maintaining children's enjoyment and positive emotions is crucial for sustained ASE enhancement. This study: 1) designs a multi-modal robot system integrated into Math e-learning, 2) explores robot behaviours improving Math performance and emotional expression, and 3) evaluates the technology's effects through a pilot study.

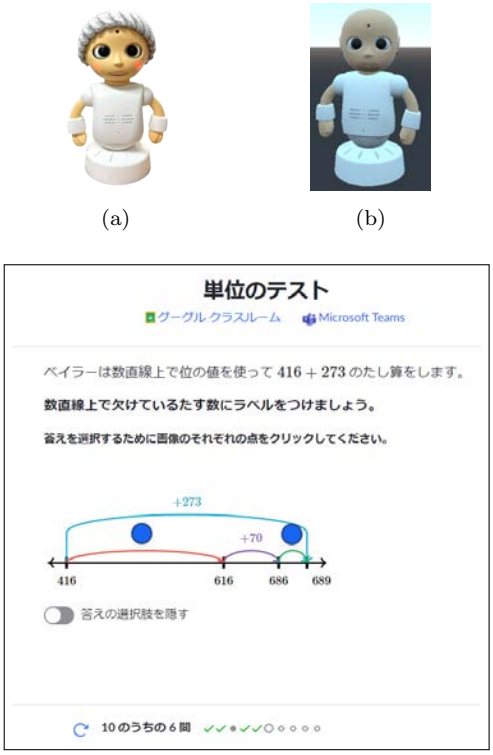
2. Related Work

2.1 Math E-learning and Enhancement Strategies

Post-COVID-19, online learning has challenged Math instruction [3]. Math apps produce short-term gains but have long-term limitations such as struggling to understand the material, completing tasks, and staying motivated [4]. Studies show that while game-based apps engage children, they often shift focus from Math concepts to tool mechanics, inhibiting learning [4], [5]. We selected Khan Academy (Fig.1c) for its simplicity, curricular alignment, and free access [6].

Breaks, defined as brief pauses to engage in different activities, benefit high-cognitive tasks like learning [7]. Ramachandran et al. tested dynamically timed breaks: breaks as rewards when performance improved >20%, and breaks as rest when performance declined >20% [7]. Both dynamic conditions significantly improved Math perfor-

[1] Graduate School of Engineering Science, The University of Osaka, Osaka, Japan
[2] Institute for Open and Transdisciplinary Research Initiatives, The University of Osaka, Osaka, Japan
a) devasena.pasupuleti@irl.sys.es.osaka-u.ac.jp
b) mahzoon@irl.sys.es.osaka-u.ac.jp
c) sakai.kazuki@irl.sys.es.osaka-u.ac.jp
d) ishiguro@sys.es.osaka-u.ac.j
e) y.yoshikawa.es@osaka-u.ac.jp



(a) (b) (c)

図 1: The tabletop physical *CommU* Robot (top left), the virtual *CommU* Robot (top right) and a Math problem displayed on the Khan Academy e-learning website (bottom).

mance versus breaks given at fixed intervals of time. We adopted this formula for triggering breaks.

Praise intrinsically motivates children, increasing task completion and confidence [8]. Person-oriented praise at 25% frequency most effectively improves performance on complex tasks, while 41% of children respond to non-verbal praise [9], [10]. For high-cognitive activities, mixing 25% verbal and 75% non-verbal praise optimally improves children's confidence and performance.

2.2 Social Robots in Education

Social robots, whether physically or virtually embodied, demonstrate advantages over human counterparts in education [11]. As customizable tutoring agents or peer learners, they improve cognitive outcomes, boost enjoyment, and offer social-emotional support [12], [13]. Robot modality significantly influences engagement, with physical robots offering tangible interactions, eliciting positive social behaviours and emotions [14]. However, educational studies often find no significant learning outcome differences between physical and virtual robots [15]. During e-learning, physical robots may distract, while virtual

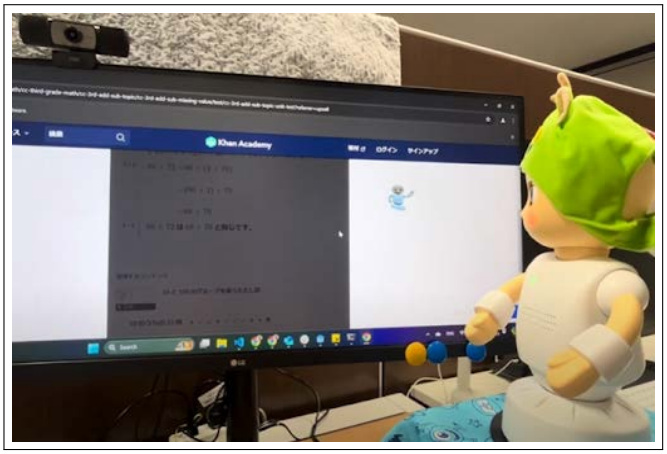


図 2: Physical and Virtual *CommU* interacting with each other following a break during learning.

robots redirect focus using pointing, gesturing, or appearing beside screen content. Additionally, children are naturally drawn to dynamic elements and interactive social cues that refocus attention [16]. This hybrid approach leverages both modalities: physical robots handle tasks requiring physical interaction (encouragement, demonstrations, tactile feedback), while customizable virtual agents guide attention to digital content.

2.3 Hypotheses

In this study, we investigate differences in children's learning, focus, and emotions when interacting with mixed-modality agents (physical and virtual) versus voice-only characters (Fig.2). Two conditions are tested: (1) "With Embodiment" - physical and virtual robot integrated into an online Math platform, and (2) "Without Embodiment" - two voice-only characters integrated into the same platform. We propose the following hypotheses:

H1: Children in condition 1 will demonstrate improved refocusing towards the learning material, due to the virtual robot's visual and social cues. **H2**: Children in condition 1 will demonstrate enhanced positive emotional expressions due to the combined social presence and tangibility of multi-modal agents. **H3**: Children in condition 1 will demonstrate greater improvements in Math performance, due to their enhanced emotional state and better attention refocusing.

3. Multi-modal Robot Design

We used *CommU*, a tabletop, non-mobile, humanoid robot with 14 DOF [17] (Fig.1a). *CommU* was developed by Osaka University researchers in collaboration with Vstone Co., Ltd., Osaka [17]. We employed the physical

and virtual versions (Fig.1b), named “Aoi” and “Ren” respectively—popular gender-neutral Japanese names to avoid gender bias. Both conditions used autonomous prescripted programs.

3.1 Physical Robot Functionality

The physical *CommU* provided timely praise, motivation, and breaks during learning. The robot calculated correctly and incorrectly answered Math questions in real-time using a Python script interfacing with Khan Academy via the Selenium library, automating a web browser (Edge) to navigate exercises and monitor button status. An HTTP-Client class controlled the physical robot.

The material contained 10 grade-3 addition/subtraction questions. First correct answers triggered verbal praise; second correct answers triggered non-verbal gestures (dance, clapping, high-five). An identical strategy applied to incorrect answers. We implemented 25% verbal and 75% non-verbal praise frequency. For consecutive incorrect answers, *CommU* switched to verbal motivation. *CommU* maintained eye contact throughout, shifting gaze when participants focused on the screen, simulating joint examination behavior. Using the formula from [7], participants received 0-2 breaks based on efficiency and accuracy. If no break occurred by 60% completion, one was provided. Break activities occurred in order: two-minute deep breathing exercise, then two-minute stretching exercise. *CommU* demonstrated, encouraged participation, and guided through gestures, then congratulated participants upon completion. When breaks triggered, a grey overlay covered the Math problem to help participants relax without pressure. Upon completion, *CommU* thanked participants and entered sleep mode.

3.2 Virtual Robot Functionality

Virtual *CommU* served as a refocusing agent. We imported *CommU*’s 3D model from Blender into the Godot 3D game engine to design behaviors and enable real-time system interaction [18]. It appeared in reduced size at bottom screen corners to avoid distracting from Math problems. Virtual *CommU* maintains the same 14 DOF as the physical robot. Its voice was generated using text-to-speech software, ensuring clear distinction from the physical robot’s voice. Its position and movement were customized for Khan Academy. Given known x-y display limits where Math problems appear, we programmed virtual *CommU*’s movements to avoid overlapping the

図 3: Virtual *CommU* trying to attract children’s attention to the e-learning material through various gestures.

problems. Godot’s transparent screen functionality made virtual *CommU* appear screen-integrated. Communication between physical and virtual *CommU*s occurred via Python script: the physical robot looped commands to Godot’s directory.

At study start, virtual *CommU* appeared in the bottom left corner and introduced itself. During idle periods, it maintained eye contact with math problems through blinking and subtle body movements. After breaks, virtual *CommU* enlarged, appeared beside problems, and encouraged refocusing. Following the “Monalisa effect,” *CommU* was designed to appear making eye contact with participants during interactions [19]. Animations included appearing through an infinity door, transforming into blue light, or using thought bubbles (Fig.3). After guiding participants, virtual *CommU* briefly interacted with physical *CommU*. Physical *CommU* responded with nods or acknowledgements while tracking virtual *CommU*’s movements. Upon completion, virtual *CommU* thanked participants and faded.

3.3 Without-Embodiment Condition

This condition maintained identical e-learning material, encouragement formulas, break algorithms, and sequences. Physical and virtual *CommU* were replaced with voice-only characters delivering identical dialogues in the same voices. Two hidden speakers, positioned on either side of the display, replicated agent positioning. A Python script managed text-to-speech and interacted with Khan

Academy via Selenium. Two hidden speakers positioned on either side of the display replicated the physical and virtual agent positioning.

4. Methodology

4.1 Participants

We conducted a pilot study with 12 elementary school children (6 girls, 6 boys) aged 9-10 from Osaka, Japan, via convenience sampling. Six participants were assigned to each condition. All self-reported average Math ability. The study was conducted entirely in Japanese. Khan Academy materials and robot interactions were also in Japanese. Standard English questionnaires were translated to Japanese by native translators, then back-translated for verification. The study received approval from our university's Institutional Ethics Committee with written and verbal consent from participants and parents. Confidentiality and anonymity of the data collected were maintained throughout the study.

4.2 Measures

Math Performance: We administered a Math test comprising 10 questions at a Japanese Grade 3 level, covering addition, subtraction, multiplication, and division problems. Similar difficulty level tests were administered pre-and post-intervention. We also recorded the time taken by each participant to complete the test.

Refocusing Ability: We utilized the "Usefulness" sub-scale from the Intrinsic Motivation Inventory (IMI) that contained 4 items. Each item was evaluated using a 7-point Likert scale, which ranged from "not at all true" to "very true."

Emotional Expression: (1) We utilized the "Enjoyment" sub-scale of the IMI index, which included three items. Each item was evaluated using a 7-point Likert scale, which ranged from "not at all true" to "very true." (2) We conducted a video analysis using the ELAN software, to capture the frequency of positive and negative facial expressions, gestures, and verbal utterances, utilizing a valence expression formula ([20], Equation (1)).

$$\begin{aligned} Valence = Smile + 2 \times (Laughter + Excitement + \\ Positive\ Verbal) - (Startle + Frown) - 2 \\ \times (Shrugging + Negative\ Verbal) \end{aligned} \quad (1)$$

Parent Questionnaires: To assess whether participants had any pre-existing challenges with attention, concentration, hyperactivity, reading, or Math, parents were asked to complete two questionnaires based on their children. (1) "Attention/Concentration" and "Reading/Writing/Arithmetic" sub-scales of the 5-15R questionnaire, which comprise 9 and 13 items respectively; (2) "Inattention" sub-scale of the Strength and Difficulties Questionnaire (SDQ), consisting of 5 items; Both assessed using a 3-point Likert scale ranging from "does not apply" to "applies".

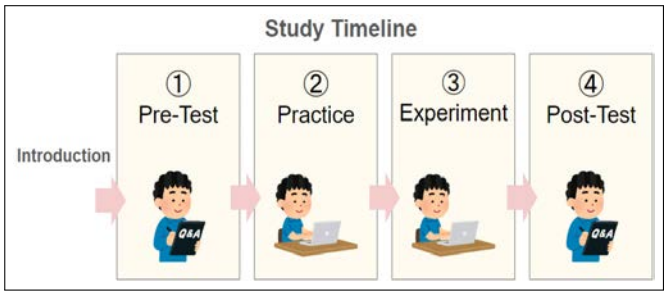


図 4: The experiment began with an introduction, followed by a pre-test, a practice session led by a researcher to help children familiarize themselves with the e-learning material, the main experiment, and concluded with a post-test.

4.3 User Study

The study was conducted inside an empty room at The University of Osaka, Japan. Parents accompanied their children to the venue, provided written consent, and were seated in a nearby soundproof room to avoid influencing their children's behaviour. Two cameras were installed: one near the questionnaire desk for pre- and post-test completion, and another near the experiment desk. Camera feeds were streamed to a monitor in the soundproof room, allowing parents to observe the session without the children's awareness. Children interacted with the math e-learning material on a desktop computer using a keyboard and mouse. Following the pre-test, a researcher conducted a practice session to ensure familiarity with the system, including navigation and interaction with robots and voice characters. The researcher did not intervene during the experiment but remained nearby to address technical issues, ensuring independent engagement and minimizing bias. Each child completed a single session (Fig.4) lasting up to 1 hour. Parents completed questionnaires during the session, and each child received a sticker featuring $CommU$ as a token of appreciation.

5. Results

All statistical analyses used a significance level of $\alpha = 0.05$. Data normality was assessed using the Shapiro–Wilk

test; parametric or non-parametric tests were applied accordingly.

5.1 Parent Questionnaire Analysis

With-Embodiment: On the 5–15R *Attention and Concentration* sub-scale, 4 children showed no difficulties, 1 showed a few problems, and 1 showed significant difficulties. On the *Reading, Writing, Arithmetic* sub-scale, 5 children had no difficulties and 1 showed minor challenges. SDQ *Inattention* scores indicated no issues for 5 children and borderline hyperactivity for 1 child.

Without-Embodiment: On the 5–15R *Attention and Concentration* sub-scale, 2 children showed no difficulties and 4 showed minor problems. For *Reading, Writing, Arithmetic*, 4 children had no difficulties and 2 showed minor challenges. All children scored within the normal range on the SDQ *Inattention* sub-scale.

5.2 Concentration and Refocusing Ability

Scores from the IMI *usefulness* sub-scale were averaged (1–2.9 = low, 3–4.9 = moderate, 5–7 = high). Children rated the activity as highly useful for refocusing in both the With-Embodiment (M = 6.04, SD = 1.36) and Without-Embodiment (M = 5.85, SD = 0.70) conditions. A Mann–Whitney U test revealed no statistically significant difference between conditions (U = 14.50, p = .568).

5.3 Emotional Expression

IMI *enjoyment* scores indicated high enjoyment in both conditions (With-Embodiment: M = 6.28, SD = 0.70; Without-Embodiment: M = 6.05, SD = 0.69), with no statistically significant difference between conditions (t(10) = 0.555, p = .591).

Video analysis of emotional expressions and verbal interactions (Fig.5) showed significantly higher positive emotional valence in the With-Embodiment condition (M = 30.83, SD = 14.89) compared to the Without-Embodiment condition, where most children exhibited negative valence (M = 3.50, SD = 0.65). This difference was statistically significant (t(10) = 3.109, p = .011).

5.4 Math Performance

Math Scores: In the With-Embodiment condition, post-test scores (M = 7.83, SD = 2.32) were slightly higher than pre-test scores (M = 7.67, SD = 1.63), but the difference was not statistically significant (t(5) = -0.542, p = .611). Scores in the Without-Embodiment condition remained unchanged (M = 8.83, SD = 2.04). No significant between-condition differences were found at pre-test (U = 8.50, p = .117) or post-test (U = 14.50, p = .56).

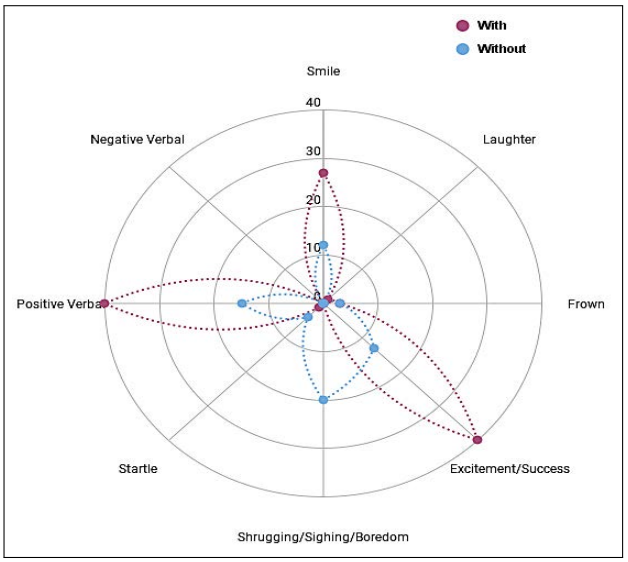


図 5: A radar graph signifying the frequency of valence emotions in the “With-Embodiment” and “Without-Embodiment” conditions.

Math Efficiency: Completion time decreased statistically significantly post-intervention in both the With-Embodiment condition (t(5) = 4.609, p = .006) and the Without-Embodiment condition (z = -2.201, p = .028). Post-intervention completion times did not significantly differ between conditions (With-Embodiment: M = 3.83 min, SD = 1.37; Without-Embodiment: M = 3.86 min, SD = 1.36; t(10) = -0.034, p = .974).

6. Discussion

Refocusing Ability: Hypothesis 1 was not supported. Both conditions were rated as highly useful for refocusing, with no significant differences observed. This may be due to the small sample size and the nature of the break activities, as breathing and stretching exercises may not have been sufficiently engaging to create a strong need for refocusing. Future work should explore break activities that are both relaxing and intrinsically engaging for children. Overall, the high ratings across conditions suggest that the system design effectively supported sustained attention during learning.

Emotional Expression: Hypothesis 2 was supported. Children in the With-Embodiment condition exhibited significantly more positive emotional expressions (e.g., smiles and excitement), whereas boredom and negative expressions were more frequent in the Without-Embodiment condition (Fig.6). These findings are con-

sistent with prior work showing that physically embodied agents elicit stronger emotional responses and higher engagement. The significant difference in emotional valence highlights the role of embodiment in enhancing user experience, even in the absence of performance gains.

Math Performance: Hypothesis 3 was partially supported. Both conditions led to reduced task completion time, indicating improved efficiency. The With-Embodiment condition showed a slight, non-significant increase in accuracy, suggesting a potential benefit of embodiment that was not statistically confirmed. Individual motivation and interest may have influenced performance outcomes, and the small sample size remains a key limitation, underscoring the need for larger-scale studies.

7. Conclusion and Future Work

This paper investigated the effects of combining physical and virtual robot modalities on children's learning outcomes, concentration, and emotional experience in a Math e-learning environment, compared to voice-only, non-embodied agents. We motivated the need for mixed-modality educational systems, highlighted the benefits of embodied agents, and described the design of physical and virtual robot behaviours integrated with Khan Academy Math content. A pilot study with Japanese children aged 9–10 compared two conditions: With-Embodiment and Without-Embodiment. Results showed that both conditions improved concentration and reduced Math task completion time, with no significant differences between them. Although the With-Embodiment condition yielded a small, non-significant improvement in Math performance, children interacting with embodied agents displayed significantly more positive emotional expressions (e.g., smiles and excitement), whereas voice-only interactions were associated with boredom and frustration. These findings suggest that mixed-modality embodiment primarily enhances the emotional quality of learning experiences rather than immediate performance gains. Future work will extend this research through a larger, long-term study in India with a more diverse participant pool to examine cultural effects on engagement with embodied versus non-embodied agents. We also plan to refine the system by varying break types and duration and by isolating the individual effects of physical and virtual robots on children's learning, self-esteem, and emotional expression.

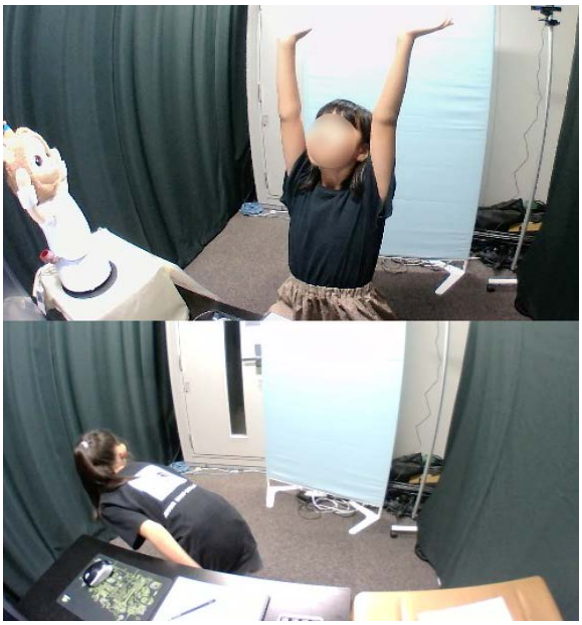


図 6: Participant interacting positively with the system in the "With-Embodiment" condition (top) and participant appearing distracted from the system in the "Without-Embodiment" condition (bottom).

参考文献

[1] X. Gong, J. Zheng, J. Zhou, E. S. Huebner, and L. Tian, "Global and domain-specific self-esteem from middle childhood to early adolescence: Co-developmental trajectories and directional relations," *Journal of Personality*, vol. 92, pp. 1356–1374, Nov. 2023.

[2] M. Hosogi, A. Okada, C. Fujii, K. Noguchi, and K. Watanabe, "Importance and usefulness of evaluating self-esteem in children," *BioPsychoSocial Medicine*, vol. 6, p. 9, Mar. 2012.

[3] D. A. Pulungan, H. Retnawati, and A. Jaedun, "Students' difficulties in online math learning during pandemic covid 19," *AKSIOMA: Jurnal Program Studi Pendidikan Matematika*, vol. 11, no. 1, p. 305, 2022.

[4] R. Kay and J. Kwak, "Do math apps help elementary school students? it depends," in *Proceedings of EdMedia + Innovate Learning 2017* (J. P. Johnston, ed.), (Washington, DC), pp. 27–32, Association for the Advancement of Computing in Education (AACE), June 2017.

[5] S. Kim and M. Chang, "Computer games for the math achievement of diverse students," *Educational Technology & Society*, vol. 13, no. 3, pp. 224–232, 2010.

[6] "Khan academy." `https://www.khanacademy.org/`, 2008. [Accessed 18-09-2024].

[7] A. Ramachandran, C.-M. Huang, and B. Scassellati, "Give me a break! personalized timing strategies to promote learning in robot-child tutoring," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '17, (New York, NY, USA), p. 146–155, Association for Computing Machinery, 2017.

[8] D. E. Kanouse, P. Gumpert, and D. Canavan-Gumpert, "The semantics of praise," in *New directions in attribution research* (J. H. Harvey, W. Ickes, and R. F. Kidd, eds.), vol. 3, pp. 97–115, Hillsdale, NJ: Erlbaum, 1981.

[9] M. Bani, "The use and frequency of verbal and non-verbal praise in nurture groups," *Emot. Behav. Diffic.*, vol. 16, pp. 47–67, Feb. 2011.

[10] R. M. Baron, A. R. Bass, and P. M. Vietze, "Type and frequency of praise as determinants of favorability of self - image: An experiment in a field setting[1]," *J. Pers.*, vol. 39, pp. 493–511, Dec. 1971.

[11] J. Wainer, D. J. Feil-Seifer, D. A. Shell, and M. J. Mataric, "Embodiment and human-robot interaction: A task-based perspective," in *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 872–877, 2007.

[12] D. Leyzberg, S. Spaulding, M. Toneva, and B. Scassellati, "The physical presence of a robot tutor increases cognitive learning gains," in *Building Bridges Across Cognitive Sciences Around the World - Proceedings of the 34th Annual Meeting of the Cognitive Science Society, CogSci 2012* (N. Miyake, D. Peebles, and R. Cooper, eds.), Building Bridges Across Cognitive Sciences Around the World - Proceedings of the 34th Annual Meeting of the Cognitive Science Society, CogSci 2012, pp. 1882–1887, The Cognitive Science Society, 2012.

[13] I. Leite, A. Pereira, G. Castellano, S. Mascarenhas, C. Martinho, and A. Paiva, "Modelling empathy in social robotic companions," in *Proceedings of the 19th International Conference on Advances in User Modeling*, UMAP'11, (Berlin, Heidelberg), p. 135–147, Springer-Verlag, 2011.

[14] A. Vrins, E. Pruss, J. Prinsen, C. Ceccato, and M. Alimardani, "Are you paying attention? the effect of embodied interaction with an adaptive robot tutor on user engagement and learning performance," in *Social Robotics*, (Cham), pp. 135–145, Springer Nature Switzerland, 2022.

[15] J. Kennedy, P. Baxter, and T. Belpaeme, "Comparing robot embodiments in a guided discovery learning interaction with children," *Int. J. Soc. Robot.*, vol. 7, pp. 293–308, Apr. 2015.

[16] Rhonda N. McEwen and Adam K. Dubé, "Engaging or distracting: Children's tablet computer use in education," *J. Educ. Techno. Soc.*, vol. 18, no. 4, pp. 9–23, 2015.

[17] T. Kanda, M. Shiomi, Z. Miyashita, H. Ishiguro, and N. Hagita, "A communication robot in a shopping mall," *IEEE Transactions on Robotics*, vol. 26, no. 5, pp. 897–913, 2010.

[18] Godot, "Godot Engine - Free and open source 2D and 3D game engine." `https://godotengine.org/`, 2014. [Accessed 19-09-2024].

[19] M. Gamer and H. Hecht, "Are you looking at me? measuring the cone of gaze," *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 33, pp. 705–715, June 2007.

[20] M. Tielman, M. Neerincx, J.-J. Meyer, and R. Looije, "Adaptive emotional expression in robot-child interaction," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '14, (New York, NY, USA), p. 407–414, Association for Computing Machinery, 2014.