Enhancing Student Engagement and Learning Outcomes Through Multimodal Robotic Interactions: A Study of non-Verbal Sound Recognition and Touch-Based Responses

Rezaul Tutul[1][a], Ilona Buchem[2] and André Jakob[2]
[1]Department of Mathematics and Natural Sciences, Humboldt University of Berlin, Berlin, Germany
[2]HARMONIK, Berliner Hochschule für Technik, Berlin, Germany

Keywords: Educational Robotics, Student Engagement, non-Verbal Interaction, Sound Recognition, Pepper Robot.

Abstract: This study examines the impact of multimodal robotic interactions on student engagement, motivation, and learning outcomes in an educational quiz-based setting using the Pepper robot. Two interaction modalities were compared: touch-based inputs (control group) and non-verbal sound-driven responses (experimental group), where students used coughing, laughing, whistling, and clapping to select answers. A novel quantitative metric was introduced to evaluate the effect of sound-driven interactions on engagement by analysing sound frequency, recognition accuracy, and response patterns. A between-subjects experiment with 40 undergraduate students enrolled in a C programming course was conducted. Motivation and engagement were assessed using the Intrinsic Motivation Inventory (IMI), while learning outcomes were measured through quiz performance (accuracy and response time). The results indicate that sound-driven interactions significantly improved quiz performance compared to touch-based inputs suggesting enhanced cognitive processing and active participation. However, no significant difference in motivation or engagement was observed between the groups (IMI subscale analysis, $p > 0.05$). These findings highlight the potential of sound-driven human-robot interactions to enhance learning experiences by activating alternative cognitive pathways.

1 INTRODUCTION

The integration of humanoid robots in education is transforming learning environments by offering interactive, personalized, and multimodal engagement (Belpaeme et al., 2018; Tutul et al., 2024; Buchem et al. 2024). Educational robots like Pepper provide students with new ways to interact with learning materials, shifting from traditional interfaces (e.g., keyboards, touchscreens) to more natural, intuitive communication methods, such as gesture, voice, and non-verbal sound-based interactions (Ouyang & Xu, 2024; Moraiti et al., 2022). While previous research has demonstrated the effectiveness of robot-assisted learning in enhancing motivation and engagement (Andić et al., 2024; Parola et al., 2021), there is limited empirical evidence on how non-verbal sound-driven responses influence learning outcomes and cognitive engagement in educational settings.

Student motivation and engagement are essential for academic success and knowledge retention (Ryan & Deci, 2000). Studies on intrinsic motivation suggest that active participation and novel interaction methods can foster deeper cognitive engagement (Huang & Hew, 2019; Wang et al., 2019). While traditional touch-based interactions remain widely used, they do not fully leverage multimodal capabilities in human-robot interaction (HRI). Non-verbal sound recognition, such as clapping, whistling, coughing, and laughing, offers an alternative hands-free, engaging interaction method (Li & Finch, 2021). However, research in this domain remains scarce, and the impact of non-verbal sound-driven interactions on learning outcomes has not been systematically studied (Fridin, 2014).

Additionally, while gamified quizzes and multimodal robotic interactions have been explored in educational robotics (Grover et al., 2016), the relationship between sound-driven engagement,

[a] https://orcid.org/ 0000-0002-8604-8501

Tutul, R., Buchem, I. and Jakob, A.
Enhancing Student Engagement and Learning Outcomes Through Multimodal Robotic Interactions: A Study of non-Verbal Sound Recognition and Touch-Based Responses.
DOI: 10.5220/0013353800003932
In *Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025) - Volume 2*, pages 131-138
ISBN: 978-989-758-746-7; ISSN: 2184-5026

motivation, and learning performance is still underexamined. Moreover, previous studies lack a structured quantitative metric for assessing the engagement impact of non-verbal interactions, making it difficult to determine their effectiveness compared to traditional methods.

To address this gap, this study investigates the impact of non-verbal sound-driven interactions on student motivation, engagement, and learning outcomes using the Pepper humanoid robot in a quiz-based learning environment. Two interaction modalities are compared:

1. Touch-based interaction (control group): Students select quiz answers using Pepper's tablet interface.
2. Non-verbal sound-driven interaction (experimental group): Students respond using predefined sounds (e.g., coughing, whistling, laughing, clapping), recognized via YAMNet-based sound recognition system.

2 RELATED WORKS

2.1 Sound Recognition in Educational Robotics

Recent advancements in human-robot interaction (HRI) have enabled robots to process non-verbal communication cues, such as gestures and sound-based interactions, to enhance student engagement and learning (Ouyang & Xu, 2024; Parola et al., 2021). Non-verbal sound cues including clapping, whistling, laughing, and coughing are widely recognized in speech and affective computing but remain underexplored in educational robotics (Fridin, 2014). Prior research has shown that sound-based interaction methods can improve social engagement in assistive robotics (Lea et al., 2022) and emotional responsiveness in child-robot interaction (Song et al., 2024). However, their application in formal learning environments remains limited.

The use of pre-trained deep learning models like YAMNet has significantly improved sound classification in robotic systems (Tutul et al., 2023). While studies have evaluated YAMNet's accuracy in detecting human-generated sounds, its impact on student engagement and learning outcomes in robot-assisted education has yet to be systematically examined. This study addresses this gap by exploring how non-verbal sound recognition influences motivation, engagement, and quiz performance.

2.2 Measuring Learning Outcomes and Engagement in Educational Robotics

Student engagement plays a crucial role in knowledge retention and active learning (Ryan & Deci, 2000). Several studies have explored how educational robotics can enhance student motivation through interactive and multimodal learning experiences (Belpaeme et al., 2018; Andić et al., 2024). However, engagement measurement in robotic learning environments remains a challenge, as traditional methods rely heavily on self-reported surveys rather than objective interaction metrics (Huang & Hew, 2019).

The Intrinsic Motivation Inventory (IMI) is one of the most widely used psychometric tools for evaluating student engagement and motivation (Ryan & Deci, 2000). While IMI has been successfully applied to robot-assisted learning (Mubin et al., 2013), it does not fully capture real-time engagement levels during interaction. To address this limitation, this study introduces a novel quantitative metric that evaluates engagement through sound frequency, recognition accuracy, and response patterns. This metric provides a more comprehensive assessment of active participation in multimodal learning environments.

2.3 Interaction Modalities in Learning Environments

Previous studies have investigated different interaction modalities in educational settings, including gesture-based, voice-based, and touch-based interactions (Huang et al., 2019; Wang et al., 2019). Touch-based interfaces, such as robotic tablets, remain the most commonly used method for student interaction (Ching & Hsu, 2023). However, recent research suggests that multimodal approaches, which combine touch, gesture, and speech-based input, can significantly enhance learning experiences by promoting active engagement and cognitive processing (Li & Finch, 2021).

Despite these advancements, there has been limited exploration of non-verbal sound-driven responses as a primary interaction method in learning environments (Han et al., 2008). The novelty of this study lies in its comparative analysis of touch-based and sound-driven responses, allowing for a better understanding of multimodal interaction benefits.

2.4 Research Gaps and Contributions

Existing studies on sound recognition in educational robotics primarily focus on technical accuracy rather than cognitive engagement and learning outcomes (Han et al., 2008; Moraiti et al., 2022). Additionally, while IMI has been used to measure motivation, few studies incorporate real-time behavioural engagement metrics (Ouyang & Xu, 2024). This study bridges these gaps by:

1. Evaluating the impact of non-verbal sound recognition on student engagement and learning outcomes.
2. Introducing a novel quantitative metric for assessing sound-driven engagement in robotic learning environments.
3. Providing an empirical comparison between sound-driven and touch-based interaction modalities in educational settings.

By addressing these research gaps, this study contributes to the design of adaptive, multimodal human-robot interaction systems that can be scaled across various learning disciplines.

3 METHODOLOGY

This study employed a between-subjects experimental design to evaluate the effects of touch-based and non-verbal sound-driven interactions on student motivation, engagement, and learning outcomes in a robot-assisted quiz-based learning environment. The experiment was conducted using the Pepper humanoid robot, and participants interacted with the system through one of two modalities: a traditional touch interface (control group) or a sound-driven interaction method (experimental group). The study aimed to investigate whether sound-based responses could enhance engagement and learning compared to traditional input methods.

3.1 Participants

A total of 40 undergraduate students (aged 18–29, M = 22.1, SD = 2.4) from a German university participated in this study. All participants were enrolled in an introductory C programming course and had prior exposure to educational robots through coursework, ensuring familiarity with human-robot interaction. They were randomly assigned to either the control group (n = 20), where they used Pepper's touchscreen interface to select quiz answers, or the experimental group (n = 20), where they responded using predefined non-verbal sounds (coughing, laughing, whistling, and clapping) recognized by YAMNet-based sound recognition system. Informed consent was obtained from all participants, and ethical approval was granted by the university's ethics committee.

3.2 Experimental Design

The experiment was structured into three phases (see figure 1 and 2). First, a 10-minute briefing and practice session was conducted, during which participants were introduced to the robot and the interaction modality assigned to their group. The main quiz session lasted 60 minutes, where each participant attempted 15 multiple-choice questions categorized into three difficulty levels (easy, medium, and hard). The control group answered via touch-based selection, while the experimental group provided answers using non-verbal sounds mapped to specific answer choices (e.g., coughing = Option A). To ensure fairness, the robot's verbal and visual feedback was kept consistent across both groups. The final phase consisted of a 20-minute post-experiment evaluation, where participants completed the Intrinsic Motivation Inventory (IMI) questionnaire and

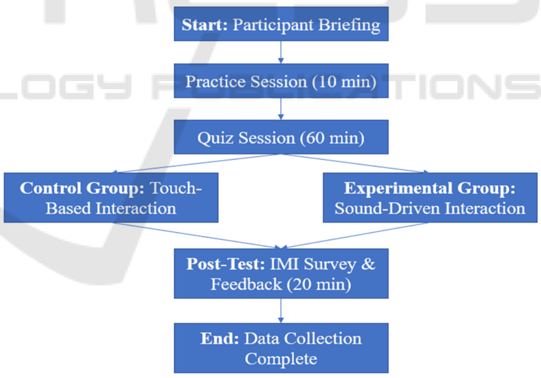

Figure 1: Experimental process.

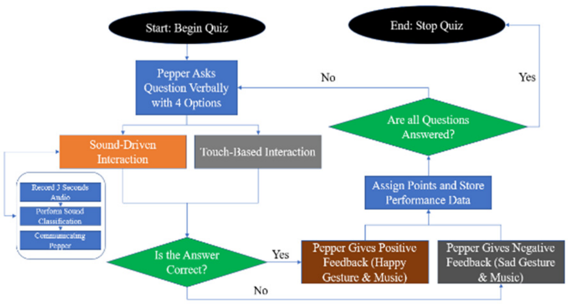

Figure 2: Quiz game algorithm flowchart.

provided qualitative feedback on their experience. The study took place in a quiet, controlled environment to minimize external distractions and ensure reliable sound recognition.

3.2.1 Interaction Modalities

Control Group (Touch-Based Interaction): Participants selected answers via Pepper's touchscreen (see figure 3). Upon selection, the robot provided verbal confirmation and visual feedback (head nod/shake).

Experimental Group (Sound-Driven Interaction): Participants responded by producing predefined non-verbal sounds (see figure 2). YAMNet is used in the server-client architecture to recognize the sound and mapped it to the corresponding answer choice (e.g., coughing = Option A, laughing = Option B, clapping = Option C, whistling = Option D). If recognition was successful, Pepper confirmed the answer verbally and displayed visual feedback.

Figure 3: Sound driven and touch based interaction.

3.3 Materials and Setup

The sound-driven quiz game system employed a client-server architecture (see figure 4). Pepper, acting as the client, was responsible for interacting with participants by asking quiz questions and providing verbal and gesture-based feedback. The server recognizing sound-driven responses (laughing, coughing, clapping, and whistling), by integrating YAMNet, a deep learning model for sound classification, communicating Pepper, and managing quiz questions and performance tracking.

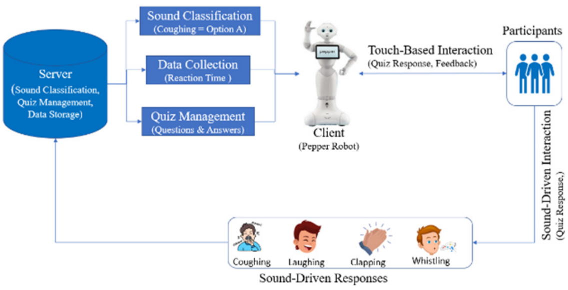

Figure 4: Sound-driven and touch-based interaction architecture.

An external microphone setup was used to enhance recognition accuracy by reducing background noise. The quiz content focused on C programming concepts, such as variables, loops, and conditions, ensuring alignment with the students' coursework. Each question had four answer choices, mapped to four distinct sound responses in the experimental group: Coughing (Option A), Whistling (Option B), Laughing (Option C), and Clapping (Option D). The system logged response accuracy, completion time, and recognition errors, which were later analyzed to evaluate interaction effectiveness.

3.4 Data Collection and Measures

To assess motivation and engagement, this study used the Intrinsic Motivation Inventory (IMI), which consists of six subscales: Interest/Enjoyment, Perceived Competence, Effort/Importance, Pressure/Tension, Perceived Choice, and Value/Usefulness. Each subscale was rated on a 5-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree). Responses were collected after the quiz session to capture changes in engagement levels. Learning outcomes were measured based on quiz performance, specifically through the number of correct answers and average response time per question. Additionally, a novel quantitative metric was introduced to analyse sound-driven engagement, considering sound frequency, recognition accuracy, and response patterns to evaluate active participation in multimodal interaction.

3.5 Data Analysis

For data analysis, descriptive statistics were used to calculate means and standard deviations for IMI subscales, quiz accuracy, and response times. Independent t-tests were conducted to compare IMI scores and quiz performance between the two groups, and effect sizes (Cohen's d) were computed to determine the magnitude of observed differences. The statistical hypotheses tested were (H1) sound-driven interactions improve quiz accuracy compared to touch-based interactions, and (H2) sound-driven interactions enhance engagement and motivation, as measured by IMI subscales. To control for potential confounds, the study ensured that all quiz questions, robot responses, and environmental conditions were identical across both groups.

Ethical considerations were strictly followed in this study. Participants were fully informed about the study's objectives, procedures, and their rights to withdraw at any time. All collected data were anonymized, and confidentiality was maintained throughout the research process. The findings from this study aim to contribute to advancing multimodal interaction designs in educational robotics, helping future research develop more adaptive and engaging human-robot learning environments.

4 RESULT

This section presents the findings of the study, including descriptive statistics (see table 1), inferential analyses, and insights into the effectiveness of sound-driven interactions compared to touch-based interactions. The results focus on motivation, engagement, and learning outcomes, measured through the Intrinsic Motivation Inventory (IMI) and quiz performance metrics. Additionally, recognition accuracy of sound-driven responses is analysed to understand its impact on student interaction.

Table 1: Descriptive Statistics for IMI subscales and quiz performance metrics between control and experimental groups.

Subscales	Control (N = 20)	Experimental (N = 20)
Interest/Enjoyment	4.44 (0.69)	4.41 (0.71)
Perceived Competence	4.28 (0.70)	4.38 (0.77)
Effort/Importance	4.39 (0.66)	4.32 (0.75)
Pressure/Tension	4.40 (0.59)	4.25 (0.70)
Perceived Choice	4.26 (0.66)	4.35 (0.82)
Value/Usefulness	4.38 (0.72)	4.23 (0.79)
Correct Answers (out of 15)	9.21 (1.55)	11.36 (1.0)
Completion Time (seconds)	78 (9.75)	83 (7.54)

The Intrinsic Motivation Inventory (IMI) subscales were analysed to assess differences in motivation and engagement between the control (touch-based) and experimental (sound-driven) groups. A paired t-test comparing post-session IMI scores revealed no statistically significant differences between the two groups across interest/enjoyment ($p=0.72$), perceived competence ($p=0.63$), effort/importance ($p = 0.58$), pressure/tension ($p = 0.81$), perceived choice ($p=0.74$), and value/usefulness ($p = 0.69$). These results suggest that while both interaction methods engaged students similarly, sound-driven interactions did not lead to a measurable improvement in self-reported motivation. Contrary to initial expectations, the use of non-verbal sound cues did not significantly enhance student engagement as measured by IMI, though qualitative feedback suggested that some participants found the experience more immersive.

In terms of learning outcomes, quiz performance was significantly higher in the experimental group compared to the control group. An independent samples t-test showed that the mean number of correct answers was significantly higher in the sound-driven group ($M = 11.36$, $SD = 1.00$) compared to the touch-based group ($M = 9.21$, $SD = 1.55$), $t = 5.47$, $p < 0.001$, Cohen's $d = 1.21$, indicating a large effect size in favour of the experimental group. These findings suggest that sound-driven interactions facilitated deeper cognitive engagement and improved accuracy in answering quiz questions. However, mean completion times per question were slightly longer for the experimental group ($M = 83s$, $SD = 7.54$) than for the control group ($M = 78s$, $SD = 9.75$). A t-test comparing response times revealed no statistically significant difference, $t = 1.37$, $p = 0.18$, suggesting that response efficiency was comparable between the two modalities. While sound-based responses took marginally longer, the difference was not substantial enough to indicate a cognitive load trade-off.

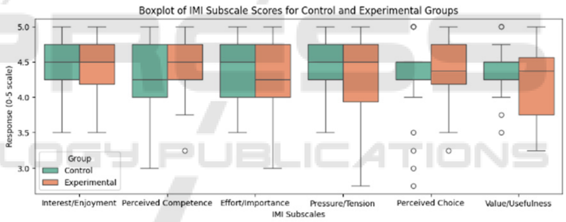

Figure 5: Boxplot of IMI Subscale Scores for Control and Experimental Groups.

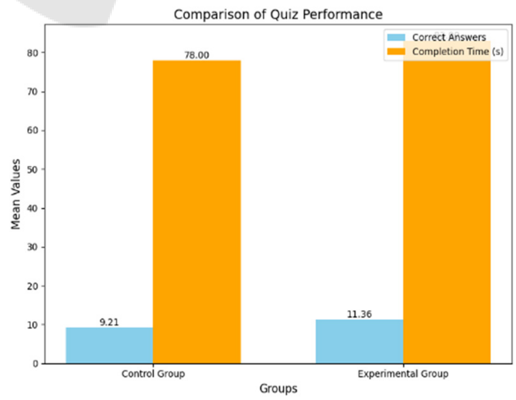

Figure 6: Quiz Performance Bar Chart.

Figure 5 presents a boxplot of IMI subscale scores, showing similar distributions across both groups. The boxplot confirms that motivation levels were comparable, as indicated by overlapping distributions across IMI subscales. The bar chart in Figure 6 illustrates the experimental group's higher average quiz scores and slightly longer completion times compared to the control group.

The recognition accuracy of sound-driven responses was also evaluated to understand its influence on student performance and interaction preferences. The overall recognition accuracy of the YAMNet model was 94%, with coughing (97%) and whistling (98%) achieving the highest accuracy, while laughing (90%) and clapping (92%) had slightly lower recognition rates. Analysis of response patterns indicated that students preferred sounds with higher recognition accuracy, suggesting that system reliability influenced interaction behaviour. Misclassification events were rare but occurred primarily in cases where laughing and clapping were confused. While these errors did not significantly affect overall quiz performance, they highlight potential technical limitations in real-time sound recognition systems.

To further explore engagement beyond IMI scores, a novel quantitative metric was introduced, analysing sound frequency, recognition accuracy, and response trends. Findings indicated that students in the experimental group used coughing and whistling more frequently due to higher recognition accuracy, while laughing and clapping were used less often. This suggests that students subconsciously adapted their interaction choices based on system reliability, reinforcing the importance of sound recognition accuracy in multimodal learning environments.

Overall, the results indicate that sound-driven interactions significantly improved learning outcomes, as evidenced by higher quiz accuracy in the experimental group. However, engagement levels remained comparable to the control group, suggesting that while non-verbal sound cues introduced novelty, they did not significantly enhance intrinsic motivation. These findings highlight the potential of sound-based human-robot interaction for learning environments, while also emphasizing the need for further refinements in recognition accuracy and system adaptability.

5 DISCUSSION

This study examined the impact of non-verbal sound-driven interactions on student engagement, motivation, and learning outcomes in a robot-assisted quiz-based learning environment. The results indicate that sound-driven interactions significantly improved quiz performance, as students in the experimental group outperformed those in the control group in terms of accuracy ($t = 5.47$, $p < 0.001$, Cohen's $d = 1.21$). However, self-reported motivation and engagement, measured through the Intrinsic Motivation Inventory (IMI), did not show significant differences between the two groups. This finding suggests that while sound-based interactions enhanced learning effectiveness, they did not intrinsically increase engagement or motivation beyond the level achieved through traditional touch-based interactions. This contradicts initial expectations that introducing non-verbal sound cues would lead to greater cognitive engagement and enjoyment. One possible explanation is that while sound-driven interactions required more active participation, they may not have been perceived as more enjoyable than touch-based selections, leading to similar IMI scores across groups.

A key insight from this study is that the cognitive processing required for sound-based responses may have contributed to improved quiz performance. The requirement to generate a non-verbal sound, wait for recognition, and receive feedback likely enhanced attention and retention, leading to higher accuracy in quiz responses. This aligns with prior research suggesting that multimodal interactions can stimulate deeper cognitive processing, thereby improving learning outcomes (Huang & Hew, 2019). However, it is also important to consider the impact of recognition accuracy on interaction efficiency. The sound recognition system (YAMNet) achieved an overall accuracy of 94%, but misclassification rates were higher for laughing and clapping. Interestingly, students in the experimental group naturally preferred sounds with higher recognition accuracy (coughing and whistling), indicating that interaction choices were subconsciously influenced by system reliability. This suggests that while sound-based interaction can be effective, its success depends on the accuracy and robustness of the recognition model and use case. Future work should explore adaptive recognition systems that can learn and optimize interactions based on user behaviour.

Another important finding concerns response time differences between the two groups. While the experimental group had slightly longer completion times per question ($M = 83$s vs. $M = 78$s), this difference was not statistically significant ($t = 1.37$, $p = 0.18$). This suggests that sound-driven interactions did not impose a major cognitive load penalty, making them a viable alternative to touch-based interactions in robot-assisted learning environments. However, it is worth considering whether the increased response time contributed to the higher quiz accuracy in the experimental group. Future studies should investigate whether the improved performance was a direct result of the interaction modality itself or simply a by-productB NM; of slower, more deliberate responses.

Although the findings provide strong support for sound-driven interactions in learning environments, there are several limitations to consider. First, the study was conducted in a controlled laboratory setting, which may not fully replicate real-world classroom conditions, where factors like peer influence, background noise, and social pressure could affect engagement and performance. Additionally, the sample size (N = 40) was relatively small, limiting the generalizability of the results. Future research should expand the sample size and conduct longitudinal studies to examine whether the benefits of sound-based interactions persist over time. Furthermore, while IMI provided useful self-reported insights, alternative engagement measures such as eye-tracking, physiological responses, or real-time interaction analytics could offer a more objective evaluation of engagement levels.

Overall, the findings highlight the potential of sound-driven human-robot interactions to enhance learning outcomes by promoting active participation and cognitive processing. While motivation levels remained comparable between the two groups, the higher quiz accuracy in the experimental group suggests that non-verbal sound interactions can be an effective alternative to traditional input methods in educational robotics. Future work should focus on improving recognition accuracy and use case, exploring multimodal adaptive learning systems, and testing these interactions in real-world educational settings to further validate their effectiveness.

6 CONCLUSIONS

This study investigated the impact of non-verbal sound-driven interactions on student engagement, motivation, and learning outcomes in a robot-assisted quiz-based learning environment. The findings reveal that students who interacted with the Pepper robot using sound-based responses achieved significantly higher quiz accuracy than those who used touch-based interactions ($t = 5.47$, $p < .001$, Cohen's $d = 1.21$), suggesting that non-verbal sound cues may enhance cognitive processing and knowledge retention. However, motivation and engagement levels, as measured by the Intrinsic Motivation Inventory (IMI), did not show significant differences between the two groups, indicating that while sound-based interactions improved learning outcomes, they did not intrinsically increase student motivation beyond traditional input methods. These results emphasize that while multimodal interactions can optimize learning efficiency, their ability to enhance motivation depends on additional factors such as user preference and system reliability.

Overall, this study highlights the potential of non-verbal sound-driven interactions in educational robotics, particularly in enhancing learning outcomes through increased cognitive engagement. While motivation levels remained similar across interaction methods, the findings suggest that sound-based responses can serve as an effective alternative to touch-based inputs in quiz-based learning. As educational technology advances, future research should aim to design scalable, adaptive human-robot interaction systems that cater to diverse learning needs and optimize multimodal engagement strategies in real-world educational settings.

REFERENCES

Andić, B., Maričić, M., Mumcu, F., & Flitcroft, D. (2024). *Direct and indirect instruction in educational robotics: A comparative study of task performance per cognitive level and student perception.* Smart Learning Environments, 11(12). https://doi.org/10.1186/s40561-024-00298-6

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics, 3*(21), eaat5954. https://doi.org/10.1126/scirobotics.aat5954

Buchem, I., Tutul, R., Bäcker, N. (2024). Same Task, Different Robot. Comparing Perceptions of Humanoid Robots Nao and Pepper as Facilitators of Empathy Mapping. In: Biele, C., et al. Digital Interaction and Machine Intelligence. MIDI 2023. Lecture Notes in Networks and Systems, vol 1076. Springer, Cham. https://doi.org/10.1007/978-3-031-66594-3_14

Ching, Y.-H., & Hsu, Y.-C. (2023). Educational Robotics for Developing Computational Thinking in Young Learners: A Systematic Review. In TechTrends (Vol. 68, Issue 3, pp. 423–434). Springer Science and Business Media LLC. https://doi.org/10.1007/s11528-023-00841-1

Fridin, M. (2014). Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education. Computers & Education, 70, 53–64. https://doi.org/10.1016/j.compedu.2013.07.043

Grover, S., Bienkowski, M., Tamrakar, A., Siddiquie, B., Salter, D., & Divakaran, A. (2016). Multimodal analytics to study collaborative problem solving in pair programming. In Proceedings of the Sixth International Conference on Learning Analytics & (pp. 516–

517). the Sixth International Conference. ACM Press. https://doi.org/10.1145/2883851.2883877

Han, J., Jo, M., Jones, V., & Jo, J. H. (2008). Comparative study on the educational use of home robots for children. *Journal of Information Processing Systems, 4*(4), 159–168.

Huang, B., & Hew, K. F. (2019). Effects of gamification on students' online interactive patterns and peer-feedback. *Distance Education, 40*(3), 350–376. https://doi.org/10.1080/01587919.2019.1632168

Lea, C., Huang, Z., Jain, D., Tooley, L., Liaghat, Z., Thelapurath, S., Findlater, L., & Bigham, J. P. (2022). Non-verbal sound detection for disordered speech. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (pp. 1-14).

Li, Y., & Finch, A. (2021). Exploring sound use in embodied interaction to facilitate learning: An experimental study. Journal of Applied Instructional Design, 11(4).

Moraiti, A., Moumoutzis, N., & Christodoulakis, S. (2022). Educational robotics and STEAM: A review. Frontiers in Education, 7, 1-18.

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J. J. (2013). A review of the applicability of robots in education. *Technology for Education and Learning, 1*, 1–7.

Ouyang, F., & Xu, W. (2024). The effects of educational robotics in STEM education: A multilevel meta-analysis. *International Journal of STEM Education, 11*(7). https://doi.org/10.1186/s40594-024-00469-4

Parola, A., Vitti, E. L., Sacco, M. M., & Trafeli, I. (2021). Educational Robotics: From Structured Game to Curricular Activity in Lower Secondary Schools. In Lecture Notes in Networks and Systems (pp. 223–228). Springer International Publishing.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Song, H., Huang, S., Barakova, E. I., Ham, J., & Markopoulos, P. (2024). How social robots can influence motivation as motivators in learning: A scoping review. In Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '23) (pp. 1-8). ACM. https://doi.org/10.1145/3594806.3604591

Tutul, R., Jakob, A., & Buchem, I. (2023). Sound recognition with a humanoid robot for a quiz game in an educational environment. Advances in Acoustics - DAGA 2023, 938–941.

Tutul, R., Buchem, I., Jakob, A., & Pinkwart, N. (2024). Enhancing Learner Motivation, Engagement, and Enjoyment Through Sound-Recognizing Humanoid Robots in Quiz-Based Educational Games. In Lecture Notes in Networks and Systems (pp. 123–132). Springer Nature Switzerland.

Wang, W., Li, R., Diekel, Z., & Jia, Y. (2019). Controlling object hand-over in human-robot collaboration via natural wearable sensing. IEEE Transactions on Human-Machine Systems, 49(1), 59-71.

Yolcu, A., & Demirer, V. (2023). The effect of educational robotics activities on students' programming success and transfer of learning. Education and Information Technologies, 28(6), 7189-7212.