Acta Informatica Pragensia
2026, Volume 15, Issue 1, Forthcoming article
https://doi.org/10.18267/j.aip.292

PRAGUE UNIVERSITY OF ECONOMICS AND BUSINESS

Article Open Access

Evaluating Intrinsic Motivation in Robot-Supported Quiz-Based Learning: A Comparative Study of Verbal-Only and Multimodal Feedback with Sound Input

Rezaul Tutul [1], Ilona Buchem [2], André Jakob [3], Niels Pinkwart [1]

[1] Faculty of Mathematics and Natural Sciences, Humboldt University of Berlin, Berlin, Germany
[2] Department of Economics and Social Sciences, Berliner University of Applied Science, Berlin, Germany
[3] Department of Electrical Engineering, Berliner University of Applied Science, Berlin, Germany

Corresponding author: Rezaul Tutul (tutulrez@student.hu-berlin.de)

Editorial Record

First submission received:
July 4, 2025

Revisions received:
September 4, 2025
October 13, 2025

Accepted for publication:
October 16, 2025

Academic Editor:
Zdenek Smutny
Prague University of Economics and Business, Czech Republic

This article was accepted for publication by the Academic Editor upon evaluation of the reviewers' comments.

How to cite this article:
Tutul, R., Buchem, I., Jakob, A., & Pinkwart, N. (2026). Evaluating Intrinsic Motivation in Robot-Supported Quiz-Based Learning: A Comparative Study of Verbal-Only and Multimodal Feedback with Sound Input. *Acta Informatica Pragensia*, 15(1), Forthcoming article. https://doi.org/10.18267/j.aip.292

Abstract

Background: Maintaining high motivation in robot-led educational activities is challenging when interactions rely solely on verbal communication. Incorporating multimodal feedback combining gestures, sounds and music may provide a richer and more engaging learning experience.

Objective: This study aims to examine whether integrating multimodal feedback with a real-time, fair first-responder detection system in a robot-led quiz game enhances students' intrinsic motivation and engagement compared to a verbal-only, sequential turn-taking interaction.

Methods: A two-group experiment was conducted with 48 university students randomly assigned to two groups. The experimental group interacted with a Pepper robot using buzzer-based competition and synchronized multimodal feedback (gestures, sounds and music), while the control group experienced verbal-only interaction with sequential turn-taking. After the session, participants completed the Intrinsic Motivation Inventory questionnaire covering five subscales: interest/enjoyment, perceived competence, effort/importance, perceived choice and pressure/tension. Statistical analyses included t-tests, 95% confidence intervals and effect sizes (Cohen's d) to compare group differences.

Results: The experimental group reported significantly higher scores in interest/enjoyment ($p < 0.001$, $d = 3.11$), perceived competence ($p < 0.001$, $d = 2.28$), effort/importance ($p < 0.001$, $d = 4.59$) and perceived choice ($p = 0.048$, $d = 0.93$). Pressure/tension scores were also higher ($p < 0.001$, $d = 1.95$), reflecting the excitement and mild stress of competitive gameplay.

Conclusion: Multimodal feedback combined with fair first-responder detection substantially enhances intrinsic motivation and engagement in robot-led learning environments. While competitive pressure increases tension, it also appears to stimulate focus and effort. These findings highlight the potential of multimodal, fair and interactive robot systems for creating more dynamic and emotionally engaging educational experiences.

Index Terms

Human–robot interaction; HRI; Multimodal feedback; Intrinsic motivation.

1 INTRODUCTION

Socially assistive robots (SARs) are increasingly entering classrooms as tutors, quiz masters and learning companions, offering new opportunities to foster engagement and motivation in education (Belpaeme et al., 2018; Kennedy et al., 2016). Their physical embodiment and ability to combine speech, gesture and affective expression provide advantages over screen-based systems, particularly in supporting active participation and social presence (Fung et al., 2024; Zhang et al., 2023).

However, many current educational robots rely predominantly on verbal-only interaction, which can limit their expressiveness, reduce their social appeal and fail to sustain learner motivation over time (Schiavo et al., 2024; Goldman et al., 2023; Schreiter et al., 2023). Recent advances in human–robot interaction highlight the value of multimodal feedback: coordinated use of gestures, music, speech and emotional expression to enhance attention, enjoyment and memory retention (Su et al., 2023; Alam, 2022). Within a motivational framework, these expressive features map directly onto self-determination theory (SDT), which posits that autonomy, competence and relatedness are core drivers of intrinsic motivation (Deci & Ryan, 2000). For instance, buzzer-based response mechanisms may enhance autonomy by giving learners control over when to participate, affective and multimodal feedback may foster competence by reinforcing mastery and self-efficacy, and shared, game-like experiences may promote relatedness through team-based dynamics (Fung et al., 2024; Yang et al., 2023, Tutul et al. 2025).

Recent developments further demonstrate how large-scale datasets and multimodal expressivity enhance robot-led education. For instance, Sorrentino et al. (2024) analysed emotional gesture synchronization in social robots for improved engagement, while Villegas-Ch et al. (2025) and Wang et al. (2023) reviewed adaptive affective feedback in humanoid tutoring systems. Similarly, Ackermann et al. (2025), Fernández-Herrero et al. (2024) and Duncan et al. (2024) highlighted multimodal expressivity as key for sustained motivation beyond novelty effects. These studies underscore that while expressive multimodal robots are gaining attention, systematic evidence on their motivational outcomes remains limited particularly in classroom contexts, which this study addresses.

Gamification further amplifies these effects. Elements such as competition, points and rewards have been shown to increase effort, persistence and enjoyment in learning technologies (Hamari et al., 2014; Zhang et al., 2023, Tutul et al., 2024). In robot-led quiz activities, first-responder detection adds an additional competitive dimension: the speed of response not only determines participation but also contributes to perceptions of fairness and recognition. According to arousal-based performance models (Yerkes & Dodson, 1908; Malhotra et al., 2010), moderate levels of competition-induced tension may actually enhance motivation and learning effort, although excessive stress may reduce engagement. Thus, designing robot-led systems that balance competition and emotional support is crucial for sustainable educational impact.

Despite these opportunities, empirical evidence on the motivational effects of multimodal versus verbal-only robot interaction remains limited. Previous research has examined engagement and usability (e.g., Kennedy et al., 2016; Louie & Nejat, 2020; Schreiter et al., 2023), but few studies have systematically evaluated how multimodal, gamified robot feedback affects intrinsic motivation using validated psychological frameworks such as the Intrinsic Motivation Inventory (IMI) (Nasir et al., 2020). Addressing this gap, the present study investigates whether multimodal feedback combined with buzzer-based first-responder detection enhances students' intrinsic motivation compared to verbal-only robot interaction with sequential turn-taking.

Building on self-determination theory (SDT) and prior research into human–robot interaction and gamification, this study examines whether multimodal feedback with buzzer-based first-responder detection enhances intrinsic motivation compared to sequential, verbal-only robot interaction. The central objective is to determine how design choices in robot-led quiz activities affect students' psychological needs for autonomy, competence and relatedness and, in turn, their intrinsic motivation. The Intrinsic Motivation Inventory (IMI) (Ryan, 1982) is a validated instrument derived from self-determination theory (Deci & Ryan, 1985, 2000) for assessing intrinsic motivation in task-based environments. In this study, five subscales were used:

- interest/enjoyment, reflecting intrinsic motivation and engagement in the activity.
- perceived competence, indicating the extent to which learners felt capable and effective.
- effort/importance, capturing the degree of personal investment and attention dedicated to the game.
- perceived choice, reflecting autonomy and volition in participation.
- pressure/tension, capturing stress levels during the interaction.

These subscales align closely with the SDT framework, where interest/enjoyment, perceived competence and perceived choice correspond to the satisfaction of autonomy and competence needs, while pressure/tension reflects the emotional arousal often associated with competitive tasks. Accordingly, we pose the following research questions:

- **RQ1.** How does multimodal robot feedback combined with first-responder detection improve students' intrinsic motivation compared to a verbal-only robot feedback with sequential turn-taking while two teams compete against each other in a quiz activity?
- **RQ2.** How does multimodal robot feedback combined with first-responder detection affect students' engagement, pressure and perceptions of fairness compared to a verbal-only robot feedback with sequential turn-taking while two teams compete against each other in a quiz activity?

Based on SDT and existing evidence on multimodal and gamified learning environments, we formulated the following hypotheses:

- **H1 (interest/enjoyment).** Students in the multimodal buzzer-based group will report higher interest/enjoyment than students in the verbal-only group, as expressive robot behaviour (music, gestures, affective speech) fosters greater engagement and enjoyment.
- **H2 (perceived competence).** Students in the multimodal buzzer-based group will report higher perceived competence than those in the verbal-only group, as multimodal praise and recognition reinforce mastery and self-efficacy.
- **H3 (effort/importance).** Students in the multimodal buzzer-based group will report higher effort/importance than those in the verbal-only group, since competition and expressive feedback increase task valuation and learning effort.
- **H4 (perceived choice).** Students in the multimodal buzzer-based group will report higher perceived choice than those in the verbal-only group, as the buzzer mechanism provides autonomy in deciding when to respond.
- **H5 (pressure/tension**). Students in the multimodal buzzer-based group will report higher pressure/tension than those in the verbal-only group, because competition can introduce moderate arousal and stress; however, this tension may also positively influence learning effort when kept at an optimal level.

2 RELATED WORK

2.1 Robots in education

Educational robots such as Pepper, NAO and Cozmo have been increasingly deployed to support learning across domains including language, science and programming (Belpaeme et al., 2018; Louie & Nejat, 2020). Their embodied presence and ability to use speech, gaze and gesture offer advantages over traditional digital tools by fostering social presence and interactive engagement (Tanaka et al., 2015; Goldman et al., 2023). Recent studies further suggest that physical embodiment can improve attention, enjoyment and learning persistence in both school and higher education contexts (Fung et al., 2024; Zhang et al., 2023; Tutul et al., 2024). However, much of this research has primarily evaluated usability or engagement outcomes rather than systematically examining intrinsic motivational processes. Moreover, while robots are often integrated into quiz-based or storytelling activities, many interactions remain limited to verbal-only exchanges, potentially reducing their effectiveness in sustaining long-term motivation (Schiavo et al., 2024).

2.2 Multimodal feedback and gamification

A growing body of work highlights the value of multimodal interaction in educational HRI. Robots capable of combining speech, gestures, music and affective expressions have been shown to elicit stronger enjoyment, emotional involvement and memory retention compared to verbal-only systems (Bagheri et al., 2020; Su et al., 2023). For example, Fung et al. (2024) demonstrated that humanoid robots supporting language learning with expressive multimodal behaviour significantly improved students' perceived competence and enjoyment, key constructs within self-determination theory (SDT). In addition, gamification elements such as points, badges and real-time competition have been widely recognized as effective strategies for sustaining motivation in technology-enhanced learning (Hamari et al., 2014; Yang et al., 2023).

In robot-led educational settings, first-responder detection introduces competitive dynamics by rewarding the fastest response. While competition can strengthen perceptions of competence and promote greater effort, it may also increase tension or stress if not perceived as fair (Kennedy et al., 2016; Malhotra et al., 2010). Fairness in response

detection is therefore critical, as biased or inaccurate recognition can reduce trust in the system and undermine motivational benefits (Fridin & Belokopytov, 2014). These findings suggest that integrating multimodal feedback and fair competition mechanisms into robot design may better support the autonomy, competence and relatedness needs central to SDT.

2.3 Measuring motivation in educational HRI

To evaluate motivational outcomes in HRI, validated psychological instruments such as the Intrinsic Motivation Inventory (IMI), the Godspeed questionnaire and the technology acceptance model (TAM) are frequently used (Bartneck et al., 2009; Nasir et al., 2020). IMI, in particular, has been widely applied in HRI and educational contexts because it captures multiple subdimensions of intrinsic motivation including interest/enjoyment, perceived competence, effort/importance, choice and pressure/tension (Baur et al., 2018). For example, studies in rehabilitation robotics and robot-assisted therapy have shown that music and multimodal interactions can significantly enhance enjoyment and engagement as measured by IMI (Baur et al., 2018). However, despite these advances, few controlled experiments have directly compared multimodal versus verbal-only robot feedback using IMI or other SDT-grounded frameworks. This lack of evidence leaves open questions about the motivational trade-offs between richer, competitive multimodal interactions and simpler, sequential verbal-only exchanges. The present study addresses this gap by conducting a controlled comparison of multimodal, buzzer-based robot feedback with verbal-only sequential interaction in a classroom quiz context, measured systematically using the IMI.

3 METHODS

3.1 Study design

A two-group design was selected to prevent learning or carry-over effects between conditions:

- **Experimental condition –** Pepper used buzzer-based first-responder detection combined with multimodal feedback (gestures, music, gamification and expressive verbal responses).
- **Control condition –** Pepper relied on sequential turn-taking with verbal-only feedback, without multimodal or gamification elements.

This design allowed us to isolate the effects of multimodal feedback and competitive dynamics on intrinsic motivation as shown in Table 1. The primary outcomes were the five Intrinsic Motivation Inventory (IMI) subscales: interest/enjoyment, perceived competence, effort/importance, perceived choice and pressure/tension. To complement quantitative data, participants also provided open-ended reflections, which were analysed thematically to capture perceptions of fairness, enjoyment and robot behaviour. The study was conducted during regular class sessions at the Berliner University of Applied Science, Germany, with each session lasting approximately 30 minutes (introduction, practice round, quiz and surveys). This naturalistic classroom setting ensured ecological validity.

Table 1. *Summary of interaction differences between control (verbal-only) and experimental (multimodal) systems.*

Feature	System A (experimental)	System B (control)
Input trigger	Buzzer sound	Robot calls each team
First-responder detection	Cross-correlation on buzzer	Predefined sequence
Answer input	Verbal (via QiSDK ASR)	Verbal (via QiSDK ASR)
Feedback type	Music + gesture + verbal	Verbal-only
Gamification	Yes (points, badges)	No

This two-group design allowed us to directly examine whether multimodal interaction with Pepper increased students' enjoyment, competence and effort (H1–H3), while also testing differences in perceived choice and pressure/tension (H4–H5) compared to a verbal-only control condition.

3.2 Participants

A total of 48 undergraduate students participated in the study, including 13 females. The average age was 22 years (SD = 1.7) and all were enrolled in undergraduate Programming in C course (Summer Semester 2025). None had prior hands-on experience with the Pepper robot. Recruitment occurred in scheduled class sessions. Participation was voluntary, without monetary or academic incentives, and informed consent was obtained from all the participants. The study protocol was approved by the Berliner University of Applied Science in Germany Review Board. Randomization was carried out using a computer-generated block randomization sequence at the session level, assigning students to either the experimental or control group. This ensured balanced group sizes and minimized allocation bias. A post hoc power analysis using G*Power indicated that, given our sample size of 48 participants (24 per group) and the observed effect sizes (Cohen's d = 0.93–4.59), the achieved statistical power was greater than 0.99 for all the subscales. This suggests that our study was sufficiently powered to detect even moderate effects.

3.3 Apparatus and system architecture

The system was designed to integrate hardware, audio processing and multimodal interaction to enable a fair and engaging quiz experience. Two wireless, recordable buzzer buttons were used by the two competing teams. These buttons transmitted audio through a single-channel microphone connected to a Bluetooth receiver attached to a desktop computer running the main detection application. The receiver sampled the audio at 16 kHz, providing sufficient temporal resolution for fast first-responder detection.

A Python-based desktop application handled the real-time detection process and communicated with Pepper using a WebSocket protocol, ensuring low-latency and bidirectional interaction. The Pepper robot's tablet, running Android 6.0, hosted a Kotlin-based quiz visualization interface, which displayed scores, badges and progress updates synchronized with the robot's gestures, speech and music cues.

Figure 1 illustrates the overall system architecture, from buzzer input capture to multimodal feedback execution. The detection mechanism responsible for ensuring accurate and fair identification of the first responder is detailed in Section 3.4.

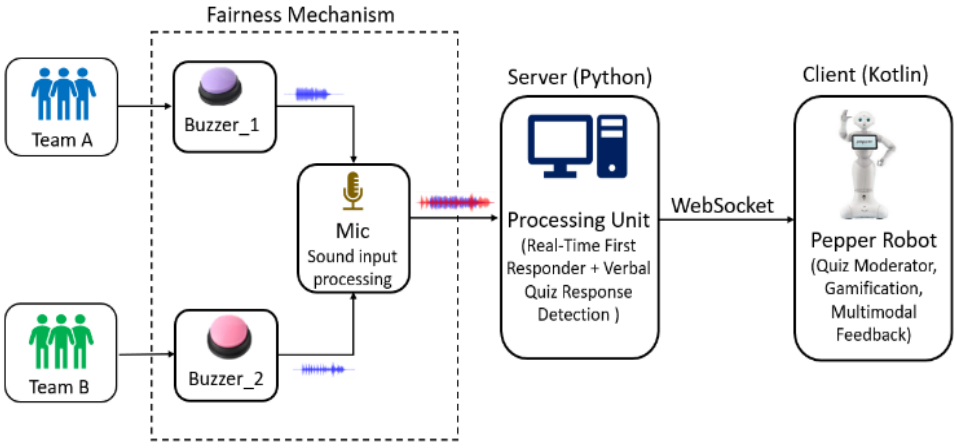

Figure 1. *Experimental (multimodal) system architecture.*

3.4 First-responder and verbal-answer detection module

The first-responder and verbal-answer detection module operationalized fairness by integrating a template-based audio matching approach for real-time buzzer detection with a QiSDK-powered automatic speech recognition (ASR) system for accurate verbal response recognition. Prior to each game session, participants recorded and tested their buzzer sounds using the Python application. This calibration step ensured that the unique sound signature of each button was captured under the current environmental noise conditions and stored as a template for later detection.

During gameplay, the system continuously monitored the incoming audio stream from the Bluetooth receiver. The following steps were executed in real time as shown in Figure 2:

- **Pre-processing:** A band-pass filter was applied to minimize environmental noise and enhance signal clarity.
- **Cross-correlation matching:** Cross-correlation was chosen due to its robustness in detecting time-lagged acoustic patterns under mild noise, ensuring fair identification of the first responder. The incoming audio was cross-correlated with the pre-recorded templates. The algorithm identified the first sound event with the highest correlation peak and the lowest time lag relative to the onset of the buzzer press.
- **First-responder assignment:** Based on the correlation lag, the system determined which team pressed their buzzer first and immediately communicated the result to the robot and tablet interface.

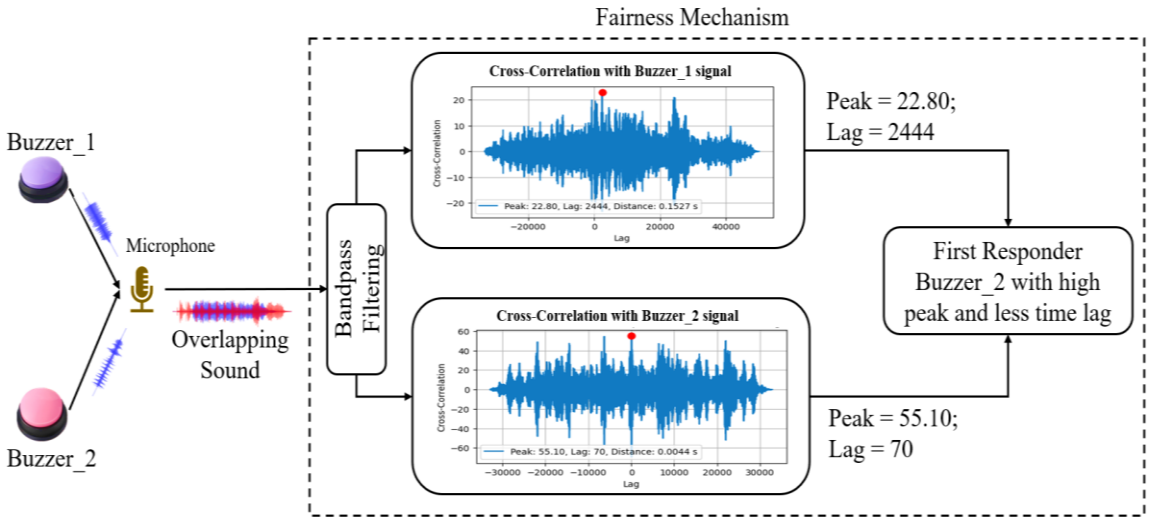

Figure 2. *Cross-correlation-based first-responder detection mechanism used in the multimodal condition.*

This approach ensured < 500 ms detection latency, which was imperceptible to participants and reinforced perceptions of fairness. Accuracy tests conducted prior to live sessions demonstrated 100% correct identification during controlled trials and 98% accuracy during pilot classroom sessions with mild background noise. The high precision of this detection process was repeatedly noted in participant feedback (e.g., *"Pepper detected the first buzzer very accurately"*), underscoring its importance for maintaining both fairness and trust during the quiz competition.

While the core interaction in the experimental condition relied on first-responder detection, the verbal-answer module used a speech-to-text system based on QiSDK ASR. Although this approach achieved high accuracy (95%) with latency < 500 ms in a controlled classroom environment, background noise occasionally caused minor recognition errors. To mitigate these issues, the system applied noise filtering and confirmation prompts from Pepper. These minor ASR inaccuracies did not significantly disrupt gameplay but may have contributed slightly to perceived tension during competitive rounds.

Together, the first-responder and ASR modules constituted the study fairness mechanism. Fairness here is defined as the equal opportunity for all participants to have their responses recognized without bias, consistent with procedural fairness frameworks (Colquitt, 2001). From a theoretical perspective, fairness is essential for sustaining motivation in competitive learning, as it reinforces competence (accurate recognition of mastery), autonomy (equal chance to compete) and relatedness (trust among peers and with the robot) as described in self-determination theory (Deci & Ryan, 2000). Thus, the fairness mechanism was not only a technical safeguard but also a psychological condition underpinning hypothesis related to motivation and engagement (H2, H3).

3.5 Quiz game flow

The quiz game was structured to create a fair, engaging and repeatable interaction sequence between the Pepper robot and the student teams. Each round consisted of five core stages:

1. **Question presentation:** Pepper verbally announced each multiple-choice question and simultaneously displayed the text and answer options on its tablet. The question data were provided by the desktop application and synchronized with Pepper via WebSocket.

2. **First-responder detection:** After the question prompt, the system activated the buzzer module. The team that pressed their buzzer first was determined using the first-responder detection module (Section 3.4). The result was transmitted in real time to Pepper, which immediately acknowledged the winning team.
3. **Response handling:**
    a. In the experimental condition, the first-responder provided their answer verbally. Pepper processed the speech using QiSDK automatic speech recognition and matched it against the correct answer stored in the quiz database.
    b. In the control condition, Pepper followed a sequential turn-taking procedure, inviting each team to respond in order until a correct or final attempt was reached.
4. **Feedback and multimodality:** Feedback was tailored to the assigned condition:
    a. In the experimental condition, Pepper combined gestures, expressive voice and music to celebrate correct answers or convey disappointment with incorrect answers.
    b. In the control condition, Pepper offered simple verbal confirmation without gestures or music.
5. **Scoring and progression:** Points were awarded to the team providing the correct answer. Scores were updated simultaneously on Pepper's tablet interface, ensuring transparency for all participants. In the multimodal condition, badges and celebratory animations were also shown to emphasize gamification.

The game proceeded until all the questions in the session were completed. A leaderboard was then displayed on Pepper's tablet and the robot verbally announced the winning team. This consistent five-stage loop ensured a smooth integration of technical detection, verbal interaction and multimodal feedback, while maintaining fairness across conditions. The flow also emphasized ecological validity, as it closely mirrored real quiz-based learning scenarios in classrooms.

3.6 Measures

The study employed a mixed-methods approach that combined standardized quantitative measures with open-ended qualitative questions to capture both statistical differences and deeper insights into students' perceptions of the robot quiz game. The five IMI subscales were selected as they most directly reflect autonomy, competence and relatedness within self-determination theory. Each subscale consisted of three slightly modified items that corresponded to one of our hypotheses: interest/enjoyment (H1), perceived competence (H2), effort/importance (H3), perceived choice (H4) and pressure/tension (H5) and rated on a 5-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree. The subscales and their items are shown in Table 2 below.

Table 2. *Intrinsic Motivation Inventory (IMI) subscales and corresponding questionnaire items used in this study.*

Subscales	Items
Interest/enjoyment	I found the quiz game with the robot enjoyable. I believe that interacting with the robot made the quiz more fun. Using the robot in the quiz game was exciting.
Perceived competence	I felt confident while participating in the quiz activity. I think I did pretty well in this quiz game. I believe I was successful in answering questions during the game.
Effort/importance	I put a lot of effort into this quiz activity. I tried hard to perform well in this game. I paid attention and stayed focused during the activity.
Perceived choice	I felt free to participate in the activity in my own way. I chose to engage with the robot quiz game because I wanted to. I felt that I had control over how I responded in the quiz.
Pressure/tension	I felt nervous while playing the game. I felt under pressure to perform well during the quiz. I felt tense when it was my turn to respond.

The IMI has been widely validated in technology-enhanced learning and human-robot interaction research, and it provides a reliable measure of intrinsic motivation dimensions relevant to gamified learning contexts.

To complement the quantitative data, participants responded to three open-ended questions at the end of the survey:

1. *What did you like most in this quiz game moderated by the Pepper robot?*
2. *How should we improve the quiz game moderated by the Pepper robot?*
3. *How motivating, engaging and enjoying was the game?*

These responses were analysed using thematic coding, focusing on recurring themes such as fun, fairness, pressure and human-like behaviour. This qualitative component enriched the interpretation of statistical findings by providing insights into students' subjective experiences.

3.7 Data analysis

The study employed both quantitative and qualitative analysis techniques to evaluate the effects of the experimental conditions on students' motivation and engagement. Responses from the five IMI subscales (interest/enjoyment, perceived competence, effort/importance, perceived choice, pressure/tension) were averaged across the three items of each subscale. Reliability was confirmed with Cronbach's alpha values greater than 0.70 for all the subscales. Statistical comparisons between the experimental (multimodal condition) and control (verbal-only condition) groups were conducted using independent-samples t-tests. Independent-samples t-tests were used as they provide a straightforward test of mean differences aligned with our hypotheses (H1–H5), which predicted differences between the multimodal and verbal-only groups. Effect sizes were reported using Cohen's d, with benchmarks of 0.2 = small, 0.5 = medium and 0.8 = large. The significance level was set at $p < 0.05$.

Qualitative responses to the three open-ended questions were analysed using qualitative content analysis (QCA) (Mayring, 2014), which involved systematic coding of text into categories. Two coders independently coded all the responses, then refined codes into five higher-level themes (enjoyment, motivation, fairness, pressure and human-like behaviour). Inter-rater reliability was assessed (Cohen's $\kappa = 0.82$), ensuring consistency. This method allows transparent, replicable categorization beyond descriptive thematic grouping.

The combination of quantitative comparisons and qualitative thematic insights allowed a more nuanced understanding of how multimodal robot feedback and buzzer-based first-responder detection influenced students' motivational experiences.

4 RESULTS

This section reports descriptive statistics, inferential comparisons between groups and distributional patterns of IMI subscales. We present means, standard deviations, interquartile ranges (IQR), t, p, 95% CIs for mean differences and Cohen's d.

4.1 Descriptive statistics

Table 3 presents the descriptive statistics for each IMI subscale across the control (verbal-only) and experimental (multimodal) conditions. Across all the subscales, the experimental group consistently reported higher mean scores compared to the control group, indicating a more positive motivational and engagement experience when multimodal feedback and sound-based first-responder detection were present.

H1 (Interest/enjoyment): For the interest/enjoyment subscale, the experimental group achieved a mean of 4.48 (SD = 0.34), notably higher than the control group's 3.39 (SD = 0.36). The median values and interquartile ranges (IQRs), 4.48 [0.46] for the experimental and 3.39 [0.49] for the control, suggest a tight clustering of scores around the group means, highlighting the consistent enjoyment experienced by participants in the multimodal setting.

Table 3. *Descriptive statistics for each Intrinsic Motivation Inventory (IMI) subscale comparing control (verbal-only, sequential interaction) and experimental (multimodal, buzzer-based) conditions.*

Subscales	Control mean (SD)	Experimental mean (SD)	Median [IQR] control	Median [IQR] experimental
Interest/enjoyment	3.39 (0.36)	4.48 (0.34)	3.39 [0.49]	4.48 [0.46]
Perceived competence	3.64 (0.38)	4.41 (0.28)	3.64 [0.51]	4.41 [0.38]
Effort/importance	3.70 (0.18)	4.70 (0.26)	3.70 [0.24]	4.70 [0.35]
Perceived choice	4.03 (0.41)	4.37 (0.31)	4.03 [0.55]	4.37 [0.42]

Subscales	Control mean (SD)	Experimental mean (SD)	Median [IQR] control	Median [IQR] experimental
Pressure/tension	3.76 (0.37)	4.44 (0.33)	3.76 [0.50]	4.44 [0.45]

H2 (Perceived competence): A similar trend was observed for perceived competence, where the experimental condition scored 4.41 (SD = 0.28; median = 4.41 [0.38]), compared to 3.64 (SD = 0.38; median = 3.64 [0.51]) in the control. This indicates that students in the experimental group felt more confident and capable during gameplay.

H3 (Effort/importance): For effort/importance, the experimental condition displayed the largest difference, with a mean of 4.70 (SD = 0.26; median = 4.70 [0.35]), compared to 3.70 (SD = 0.18; median = 3.70 [0.24]) in the control. This finding suggests that multimodal interaction encouraged students to invest more effort and prioritize performance during the activity.

H4 (Perceived choice): The perceived choice subscale also reflected higher autonomy perceptions in the experimental condition (M = 4.37, SD = 0.31; median = 4.37 [0.42]) compared to the control (M = 4.03, SD = 0.41; median = 4.03 [0.55]), indicating that the multimodal system fostered a greater sense of voluntary participation and control.

H5 (Pressure/tension): Finally, for pressure/tension, participants in the experimental group reported slightly higher scores (M = 4.44, SD = 0.33; median = 4.44 [0.45]) than those in the control condition (M = 3.76, SD = 0.37; median = 3.76 [0.50]). This reflects the competitive tension introduced by the fast-paced, first-responder quiz mechanics, which participants often described as "exciting" but occasionally "stressful".

Overall, the descriptive statistics reveal consistent advantages of the multimodal condition across all the motivational subscales, with small interquartile ranges indicating a high level of agreement among participants within each group.

4.2 Inferential statistics

Table 4 summarizes the results of the independent-samples t-tests comparing the control (verbal-only) and experimental (multimodal) conditions across all five IMI subscales. Significant differences were found in all the subscales, supporting the hypothesized advantages of the multimodal, sound-driven interaction.

For interest/enjoyment, the experimental group scored significantly higher ($t(46) = -6.96$, $p < 0.001$, Cohen's $d = -3.11$), with a mean difference of 1.09 points (95% CI [0.89, 1.29]). This represents a very large effect size, indicating that multimodal interaction with sound-driven first-responder detection substantially increased enjoyment and engagement compared to the verbal-only setup.

Table 4. *Independent-samples t-test results comparing control (verbal-only, sequential interaction) and experimental (multimodal, buzzer-based) conditions across five IMI subscales.*

Subscales	t(df)	p	Cohen's d	Mean diff (exp – ctrl)	95% CI
Interest/enjoyment	-6.96 (46)	< 0.001	-3.11	1.09	[0.89, 1.29]
Perceived competence	-5.24 (46)	< 0.001	-2.28	0.77	[0.58, 0.96]
Effort/importance	-9.83 (46)	< 0.001	-4.59	1.00	[0.87, 1.13]
Perceived choice	-2.12 (46)	0.048	-0.93	0.34	[0.13, 0.55]
Pressure/tension	-4.37 (46)	< 0.001	-1.95	0.68	[0.48, 0.88]

Similarly, perceived competence was significantly greater in the experimental condition ($t(46) = -5.24$, $p < 0.001$, Cohen's $d = -2.28$), with a mean difference of 0.77 points (95% CI [0.58, 0.96]). This suggests that the multimodal system made students feel more confident and successful during the quiz activities.

The effort/importance subscale revealed the largest difference ($t(46) = -9.83$, $p < 0.001$, Cohen's $d = -4.59$), with a mean difference of 1.00 point (95% CI [0.87, 1.13]). This extremely large effect demonstrates that the multimodal setting significantly motivated students to invest greater effort and remain highly focused during gameplay.

For perceived choice, the difference between groups was smaller but still significant ($t(46) = -2.12$, $p = 0.048$, Cohen's $d = -0.93$), with a mean difference of 0.34 points (95% CI [0.13, 0.55]). This indicates that the multimodal condition slightly enhanced participants' sense of autonomy and voluntary engagement.

Finally, the pressure/tension scores were also significantly higher in the experimental group ($t(46) = -4.37$, $p < 0.001$, Cohen's $d = -1.95$), with a mean difference of 0.68 points (95% CI [0.48, 0.88]). While this reflects increased competitive tension, qualitative feedback (section 4.4) highlighted that the pressure was perceived as positive and stimulating, contributing to the excitement of the quiz rather than reducing enjoyment.

Overall, the inferential analyses confirm that the multimodal, gamified system consistently outperformed the verbal-only version by fostering enjoyment, competence, effort and engagement, with particularly large effects in effort and enjoyment, directly supporting hypotheses H1 through H5.

4.3 Distribution patterns

The boxplot shown in Figure 5 illustrates the distribution of post-intervention IMI subscale scores for both the control and experimental conditions, highlighting differences in central tendency, variability and outliers across the five subscales.

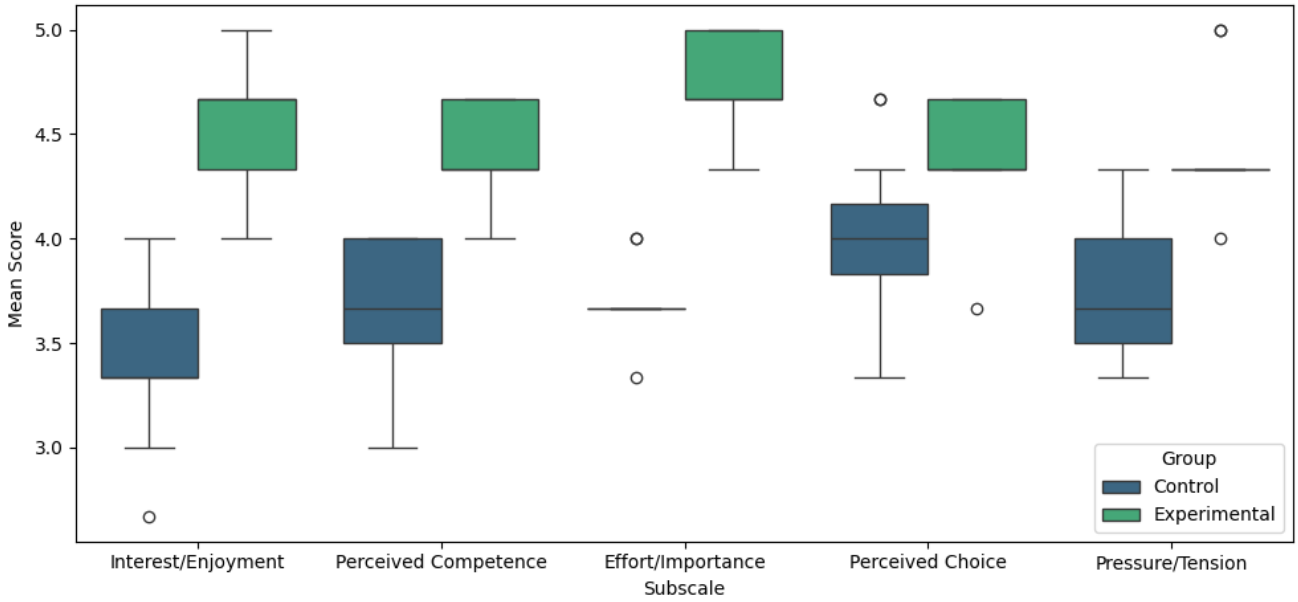

Figure 5. *Distribution of IMI subscale scores for control (verbal-only) and experimental (multimodal) conditions. Side-by-side boxplots display medians, interquartile ranges and outliers for each subscale.*

1. **Interest/enjoyment:**
    a. **Experimental group:** Scores are consistently high, clustering around 4.5, with a narrow interquartile range (IQR) indicating high agreement among participants. A few mild outliers are present but close to the upper whisker.
    b. **Control group:** Scores range more widely, centred around 3.5, with a lower median and a broader spread, indicating greater variability and lower overall enjoyment.
2. **Perceived competence:**
    a. **Experimental group:** Participants report high competence (median ~4.4), with a relatively tight distribution, signalling consistent perceptions of improved performance during the multimodal interaction.
    b. **Control group:** Medians are notably lower (~3.6) with a wider range, showing that participants in the control condition felt less confident and more varied in their competence ratings.
3. **Effort/importance:**
    a. **Experimental group:** Shows the highest clustering near the upper scale (4.7–5), with almost no variability, suggesting strong and unanimous engagement and effort.

   b. **Control group:** Scores centre around 3.7 with a narrow spread, indicating that participants tried but at a significantly lower level of commitment compared to the experimental group.
4. **Perceived choice:**
   a. **Experimental group:** Medians are higher (~4.4) with moderate variability, reflecting greater autonomy and perceived agency during the multimodal sessions.
   b. **Control group:** Slightly lower median (~4.0) and a wider range, suggesting more mixed experiences regarding autonomy and freedom of interaction.
5. **Pressure/tension:**
   a. **Experimental group:** Higher scores (~4.4) indicate that participants experienced more pressure and tension, likely due to the competitive and fast-paced nature of the multimodal condition. The IQR remains moderate, with a few outliers in the upper range.
   b. **Control group:** Lower tension (~3.7) with a tighter distribution, reflecting a more relaxed and less intense experience in the verbal-only condition.

4.4 Qualitative findings

The thematic analysis of open-ended responses provided further insights into participants' experiences with the robot quiz game. Five major themes emerged, each corresponding closely to the IMI subscales and hypotheses. Table 5 summarizes representative quotes for both the multimodal condition and the verbal-only control condition.

Enjoyment and fun (H1): Participants in the multimodal condition frequently highlighted the enjoyable and entertaining aspects of the robot quiz game. The integration of synchronized gestures, music and dance was described as enhancing the sense of fun. In contrast, control group participants acknowledged the activity as interesting but suggested adding multimodal elements (e.g., music, dance) to increase enjoyment.

Multimodal: *"The quiz game was very interesting and enjoyable."*

Control: *"Interesting way of learning, but adding music and dance would make it more fun."*

Table 5. *Thematic categories and representative participant quotes derived from qualitative feedback collected after gameplay.*

Theme	Verbal-only condition (control)	Multimodal condition (experimental)
Fun & enjoyment	"Interesting way of learning."	"The quiz game was very interesting and enjoyable."
Learning motivation	"Adding music and dance would make it more fun."	"Synchronized robot feedback was very motivating."
System fairness	"Very simple and fair interaction."	"Pepper detected the first buzzer very accurately."
Pressure & tension	"Felt no pressure at all, relaxed interaction."	"Pressing the buzzer on time is stressful but exciting."
Human-like behaviour	"Robot gave clear but simple feedback."	"The robot felt emotionally responsive, almost like a human."

Perceived competence and motivation (H2): The multimodal group reported feeling more competent and motivated, often citing the robot's expressive feedback as supportive and encouraging. Participants felt successful in answering questions and appreciated the reinforcement of correct responses. Control participants also expressed confidence but described the interaction as simpler and less stimulating, aligning with the lower competence scores observed quantitatively.

Multimodal: *"Synchronized robot feedback was very motivating."*

Control: *"Robot gave clear but simple feedback."*

Effort and engagement (H3): Responses indicated that students in the multimodal condition invested greater effort and concentration in the game. Many reported being eager to press the buzzer and compete. In the control condition, students engaged with the game but described their involvement as more relaxed and less competitive.

Multimodal: *"I tried hard to be the first to respond; it made me stay very focused."*

Control: *"The quiz was nice, but it felt more like just answering questions."*

Autonomy and choice (H4): Perceptions of autonomy were less salient in qualitative responses compared to other themes. However, some multimodal participants emphasized that the system allowed them to express themselves actively by pressing the buzzer and receiving tailored robot responses. Control participants described the quiz as straightforward and teacher-like, suggesting reduced perceptions of choice.

Multimodal: *"Pressing the buzzer made me feel more in control."*

Control: *"It was simple, like a teacher asking and we replying."*

Pressure and tension (H5): Finally, the theme of pressure and arousal emerged strongly in the multimodal group. Students described the buzzer competition as "stressful but exciting", reflecting the tension between performance anxiety and motivational arousal. In contrast, control participants reported a calmer and more relaxed interaction, with little sense of time pressure.

Multimodal: *"Pressing the buzzer on time is stressful but exciting."*

Control: *"Felt no pressure at all, relaxed interaction."*

5 DISCUSSION

This study investigated whether a sound-driven, multimodal robot quiz game can enhance students' motivation and engagement compared to a verbal-only condition. To address this objective, we formulated two research questions (RQs) supported by five hypotheses (H1–H5). The following sections integrate the quantitative and qualitative findings to answer each RQ.

RQ1. *How does multimodal robot feedback combined with first-responder detection improve students' intrinsic motivation compared to a verbal-only robot feedback with sequential turn-taking while two teams compete against each other in a quiz activity?*

H1 (Interest/enjoyment): Quantitative analyses revealed that interest/enjoyment scores were significantly higher in the multimodal condition ($M = 4.48$ vs 3.39, $t = -6.96$, $p < 0.001$, $d = -3.11$). This supports H1 and demonstrates a substantial motivational benefit. Qualitative responses echoed this, with students describing the game as *"very interesting and enjoyable"*, while control participants suggested that adding multimodal elements *"would make it more fun"*. These results align with prior findings that multimodal and gamified robot feedback increases enjoyment in learning contexts (Fung et al., 2024; Hamari et al., 2014).

H2 (Perceived competence): Students in the multimodal condition also reported higher competence ($M = 4.41$ vs 3.64, $t = -5.24$, $p < 0.001$, $d = -2.28$). Comments such as *"Synchronized robot feedback was very motivating"* illustrate how expressive reinforcement supported feelings of success. Control participants found the interaction clear but less stimulating, confirming H2.

H3 (Effort/importance): Effort/importance yielded the strongest effect ($M = 4.70$ vs 3.70, $t = -9.83$, $p < 0.001$, $d = -4.59$), showing that multimodal robot interaction motivated students to invest more energy. Participants described being *"focused"* and eager to press the buzzer, while control participants saw the activity as *"just answering questions"*. This strongly supports H3.

H4 (Perceived choice): Perceived choice improved moderately in the multimodal condition ($M = 4.37$ vs 4.03, $t = -2.12$, $p = 0.048$, $d = -0.93$). Qualitative data suggest that autonomy stemmed from pressing the buzzer ("made me feel more in control"), while the control condition felt more teacher-like. This provides partial support for H4.

Summary for RQ1: Taken together, H1–H4 confirm that multimodal robot feedback enhances enjoyment, competence, effort and autonomy. These outcomes directly answer RQ1, showing that multimodal interaction substantially increases motivation compared to a verbal-only condition.

RQ2. *How does multimodal robot feedback combined with first-responder detection affect students' engagement, pressure and perceptions of fairness compared to a verbal-only robot feedback with sequential turn-taking while two teams compete against each other in a quiz activity?*

> **H5 (Pressure/tension):** Quantitative results revealed higher pressure/tension in the multimodal condition ($M = 4.44$ vs 3.76, $t = -4.37$, $p < 0.001$, $d = -1.95$). Students described the buzzer competition as *"stressful but exciting"*, indicating positive arousal (eustress) rather than negative stress. By contrast, the control group reported being *"relaxed"* with little sense of time pressure. This supports H5 and suggests that multimodal interaction enhances competitive engagement through energizing tension. These findings align with arousal-based models of performance (Yerkes & Dodson, 1908; Malhotra, 2010), which suggest that moderate levels of tension can enhance focus, effort and engagement.
>
> **Fairness and human-likeness:** Although not part of the original hypotheses, two qualitative themes enrich the interpretation of RQ2. Students emphasized the fairness of the system, noting that *"Pepper detected the first buzzer very accurately"*, which validates the technical robustness of the cross-correlation module. Additionally, some participants described Pepper as *"emotionally responsive, almost like a human"*, reflecting heightened social presence in the multimodal condition. Perceptions of fairness are consistent with studies showing that students value transparency and unbiased responses in robot-led activities (Fridin & Belokopytov, 2014; Belpaeme et al., 2018).

Summary for RQ2: Findings show that multimodal robot moderation increases pressure in a constructive way, while also being perceived as fair and socially engaging. These insights extend the scope of RQ2 by linking technical system design to educational engagement outcomes.

While the strong motivational effects observed in the multimodal condition support the proposed hypotheses, alternative explanations such as novelty effects or social desirability bias cannot be fully ruled out. Students' initial excitement about interacting with a humanoid robot for the first time may have inflated enjoyment and effort ratings. Moreover, group-level dynamics could have amplified competitive arousal independently of multimodal feedback. Future longitudinal studies should disentangle these factors by controlling for repeated exposure and group competition levels.

5.1 Implications

The findings of this study have important implications for both educational practice and the design of robot-mediated learning systems.

5.1.1 Educational implications

By demonstrating that multimodal robot feedback significantly improves students' enjoyment, competence and effort, this study provides evidence that robotic systems can serve not only as instructional tools but also as motivational facilitators. The presence of beneficial pressure (eustress) suggests that competitive quiz-based interactions can foster active participation and sustained engagement. For educators, this indicates that deploying social robots such as Pepper in classroom activities may enhance students' intrinsic motivation and engagement, particularly when multimodal cues (gestures, music, expressive feedback) are integrated into the learning process.

It is also important to acknowledge the potential novelty effect associated with the use of Pepper. Participants' high enjoyment and motivation could partially stem from the excitement of interacting with the robot for the first time. While this novelty-driven engagement is common in educational HRI settings, longitudinal studies are needed to examine whether these motivational gains persist once the novelty diminishes over time.

5.1.2 Technical implications

The positive perception of fairness highlights the importance of robust detection mechanisms in educational robotics. Our cross-correlation-based first-responder module ensured accurate recognition of buzzer inputs and was explicitly recognized by participants as *"accurate and fair"*. This underlines that technical reliability is not only a performance requirement but also a social factor influencing trust and acceptance. Future HRI systems should prioritize detection accuracy and latency minimization to maintain perceived fairness in competitive learning contexts.

The integration of multimodal feedback in the robot-led quiz game is grounded in the principles of self-determination theory (SDT; Deci & Ryan, 2000). By providing gesture-based and musical feedback, the system supports autonomy by allowing participants to engage actively in a dynamic interaction rather than passive listening. Synchronized gestures, verbal affirmations and musical cues enhance competence by reinforcing a sense of mastery and successful performance. Additionally, the robot's expressive responses create a sense of relatedness, fostering social presence and shared enjoyment within the group setting.

5.1.3 Design implications

From a design perspective, this study shows the value of combining gamification mechanics with multimodal interaction. Features such as synchronized dance, music and expressive gestures strengthened students' motivation and enhanced the robot's social presence. However, designers should also be mindful of balancing pressure and autonomy. While competition was motivating, excessive stress could hinder learning in some contexts. Beyond the current implementation, the system could benefit from incorporating adaptive and personalized feedback mechanisms. For example, the robot could adjust its verbal feedback, gestures and musical responses based on learners' prior performance or engagement profiles. Such personalization could enhance feelings of autonomy and competence, aligning with principles of self-determination theory and further boosting motivation during repeated interactions.

Taken together, these implications suggest that educational robots should not only deliver content but also function as motivational partners, with fairness, multimodality and adaptive engagement mechanisms at their core.

5.2 Limitations and future work

This study has some limitations that should be acknowledged. Firstly, while the results demonstrate clear motivational and engagement benefits of the multimodal condition, we did not employ a pre/post IMI design, which limits our ability to track baseline changes over time. Future experiments should incorporate repeated measurements to capture longitudinal trends. Secondly, the study was limited to a single university and a relatively small participant group, which constrains the generalizability of the findings. Additional experiments in varied cultural and educational contexts are needed to examine broader applicability. Thirdly, while our system achieved near-perfect accuracy in first-responder detection, residual delays in noisy conditions could influence perceived fairness in larger or more acoustically complex environments. Lastly, a longitudinal approach is recommended to assess whether the observed motivational and engagement effects are sustained over multiple sessions as the novelty of robot-assisted quizzes diminishes.

Future studies could integrate physiological or behavioural measures, such as heart rate variability, facial expression analysis, or gaze tracking, to capture real-time indicators of stress, engagement and emotional states. These multimodal data would complement self-reported measures and offer deeper insights into learners' experiences. Future studies may also expand the design space by introducing adaptive feedback mechanisms that calibrate competition intensity and multimodal cues according to learners' individual profiles. Furthermore, integrating learning outcome measures (e.g., retention tests, concept mastery) with motivational and engagement indicators would provide a more holistic evaluation of robot-mediated education. Finally, cross-cultural studies could investigate how perceptions of fairness, pressure and enjoyment vary in different educational contexts, thereby broadening the relevance of research into multimodal human–robot interaction.

6 CONCLUSION

This study examined how a sound-driven, multimodal quiz game moderated by the Pepper robot influences students' motivation, engagement and perceptions of fairness compared to a verbal-only condition. Guided by two research questions and five hypotheses, the investigation combined quantitative analyses of Intrinsic Motivation Inventory (IMI) subscales with qualitative feedback from participants.

The findings demonstrate that multimodal robot interaction significantly enhanced students' enjoyment, perceived competence, effort and sense of choice, while also fostering higher engagement through competitive tension. Participants perceived the system as fair due to the reliable cross-correlation-based first-responder detection module and described the robot's expressive behaviour as emotionally engaging and human-like. These results confirm that

fairness and multimodal feedback are critical design elements for sustaining motivation and engagement in robot-mediated education.

From an educational perspective, the study highlights that social robots can act not only as facilitators of learning but also as motivational partners capable of creating enjoyable and stimulating environments. Technically, the integration of precise sound-based first-responder detection with multimodal feedback establishes a scalable foundation for fair and engaging classroom applications.

Nevertheless, the scope of the study was limited by its moderate sample size and short-term focus. Future work should validate these results with larger and more diverse participant groups, investigate long-term motivational effects and implement adaptive feedback systems that dynamically adjust competition intensity to learner profiles.

Overall, this research contributes to the growing evidence that educational robots designed with fairness, multimodality and gamification at their core can meaningfully enhance student motivation and engagement. By linking technical accuracy with educational impact, the study advances both theoretical understanding of human-robot interaction and practical deployment of intelligent robotic systems in classroom contexts. Future research should further explore adaptive multimodal responses informed by learners' real-time affective states and examine the system scalability across diverse cultural and educational settings.

ADDITIONAL INFORMATION AND DECLARATIONS

Acknowledgments: The authors would like to thank the participating students and faculty members of the Berliner University of Applied Sciences for their valuable time and cooperation during the experimental sessions. The authors also appreciate the constructive feedback provided by anonymous reviewers, which helped improve the quality and clarity of this article.

Funding: This research received no external funding.

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: R.T.: Conceptualization, Methodology, Software, Data curation, Writing – Original draft preparation. I.B.: Supervision, Validation, Writing – Review & Editing. A.J.: Resources, Technical Support, Software. N.P.: Supervision, Conceptualization, Review & Editing.

Institutional Review Board Statement: This study was conducted in accordance with the ethical guidelines of the Berliner Hochschule für Technik (BHT; or Berliner University of Applied Science) and the General Data Protection Regulation (GDPR). As the study involved anonymous, voluntary participation of adult students and no collection of personal or sensitive data, formal ethics committee approval was not required under §1(2)–(3) of the BHT Ethics Commission Statute (April 2025).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the research.

Statement on the Use of Artificial Intelligence Tools: The authors declare that they didn't use artificial intelligence tools for text or other media generation in this article.

Data Availability: The data that support the findings of this study are available from the corresponding author.

REFERENCES

Ackermann, H., Henke, A., Chevalère, J., Yun, H. S., Hafner, V. V., Pinkwart, N., & Lazarides, R. (2025). Physical embodiment and anthropomorphism of AI tutors and their role in student enjoyment and performance. *Npj Science of Learning*, 10(1), Article 1. https://doi.org/10.1038/s41539-024-00293-z

Alam, A. (2022). Social Robots in Education for Long-Term Human-Robot Interaction: Socially Supportive Behaviour of Robotic Tutor for Creating Robo-Tangible Learning Environment in a Guided Discovery Learning Interaction. *ECS Transactions*, 107(1), 12389–12403. https://doi.org/10.1149/10701.12389ecst

Bacula, A., & Knight, H. (2024). Dancing with Robots at a Science Museum: Coherent Motions Got More People to Dance, Incoherent Sends Weaker Signal. In *Proceedings of the 2024 International Symposium on Technological Advances in Human-Robot Interaction* ,(pp. 83–91). ACM. https://doi.org/10.1145/3648536.3648546

Bagheri, E., Vanderborght, B., Roesler, O., & Cao, H.-L. (2020). A Reinforcement Learning Based Cognitive Empathy Framework for Social Robots. *International Journal of Social Robotics*, 13(5), 1079–1093. https://doi.org/10.1007/s12369-020-00683-4

Bartneck, C., Kulic, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Baur, K., Speth, F., Nagle, A., Riener, R., & Klamroth-Marganska, V. (2018). Music meets robotics: a prospective randomized study on motivation during robot aided therapy. *Journal of NeuroEngineering and Rehabilitation*, 15, Article 79. https://doi.org/10.1186/s12984-018-0413-8

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

Colquitt, J. A. (2001). On the dimensionality of organizational justice: A construct validation of a measure. *Journal of Applied Psychology*, *86*(3), 386–400. https://doi.org/10.1037/0021-9010.86.3.386

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behaviour*. Springer.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behaviour. *Psychological Inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Duncan, J. A., Alambeigi, F., & Pryor, M. W. (2024). A Survey of Multimodal Perception Methods for Human–Robot Interaction in Social Environments. *ACM Transactions on Human-Robot Interaction*, 13(4), 1–50. https://doi.org/10.1145/3657030

Fernández-Herrero, J. (2024). Evaluating Recent Advances in Affective Intelligent Tutoring Systems: A Scoping Review of Educational Impacts and Future Prospects. *Education Sciences*, 14(8), Article 839. https://doi.org/10.3390/educsci14080839

Fridin, M., & Belokopytov, M. (2014). Acceptance of socially assistive humanoid robot by preschool and elementary school teachers. *Computers in Human Behavior*, 33, 23–31. https://doi.org/10.1016/j.chb.2013.12.016

Fung, K. Y., Lee, L. H., Sin, K. F., Song, S., & Qu, H. (2024). Humanoid robot-empowered language learning based on self-determination theory. *Education and Information Technologies*, 29(14), 18927–18957. https://doi.org/10.1007/s10639-024-12570-w

Goldman, E.J., Baumann, A., & Poulin-Dubois, D. (2023). Pre-schoolers' anthropomorphizing of robots: Do human-like properties matter? *Frontiers in Psychology*, 13, 1102370. https://doi.org/10.3389/fpsyg.2022.1102370

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? – A literature review of empirical studies on gamification. In *2014 47th Hawaii International Conference on System Sciences*, (pp. 3025–3034). IEEE. https://doi.org/10.1109/HICSS.2014.377

Hirschmanner, M., Gross, S., Krenn, B., Neubarth, F., Trapp, M., & Vincze, M. (2018). Grounded Word Learning on a Pepper Robot. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents,* (pp. 351–352). ACM. https://doi.org/10.1145/3267851.3267903

Huang, P., Hu, Y., Nechyporenko, N., Kim, D., Talbott, W., & Zhang, J. (2024). EMOTION: Expressive Motion Sequence Generation for Humanoid Robots with In-Context Learning. *IEEE Robotics and Automation Letters*, 10, 7699–7706. https://doi.org/10.1109/LRA.2025.3575983

Jacobs, E., Garbrecht, O., Kneer, R., & Rohlfs, W. (2023). Game-based learning apps in engineering education: requirements, design and reception among students. *European Journal of Engineering Education*, 48(3), 448–481. https://doi.org/10.1080/03043797.2023.2169106

Kennedy, J., Baxter, P., & Belpaeme, T. (2016). The robot who tried too hard: Social behavior of a robot tutor can negatively affect child learning. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction,* (pp. 67–74). ACM. https://doi.org/10.1145/2696454.2696457

Kim, J. (2017). Transforming Music Education for the Next Generation: Planting 'Four Cs' Through Children's Songs. *International Journal of Early Childhood*, 49(2), 181–193. https://doi.org/10.1007/s13158-017-0187-3

Kim, S.-H., Nam, H., Choi, S.-M., & Park, Y.-H. (2024). Real-Time Sound Recognition System for Human Care Robot Considering Custom Sound Events. *IEEE Access*, 12, 42279–42294. https://doi.org/10.1109/access.2024.3378096

Krogsager, A., Segato, N., & Rehm, M. (2014). Backchannel Head Nods in Danish First Meeting Encounters with a Humanoid Robot: The Role of Physical Embodiment. In *Human-Computer Interaction. Advanced Interaction Modalities and Techniques*, (pp. 651–662). Springer. https://doi.org/10.1007/978-3-319-07230-2_62

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: A survey. *International Journal of Social Robotics*, 5, 291–308. https://doi.org/10.1007/s12369-013-0178-y

Louie, W.-Y. G., & Nejat, G. (2020). A Social Robot Learning to Facilitate an Assistive Group-Based Activity from Non-expert Caregivers. *International Journal of Social Robotics*, 12(5), 1159–1176. https://doi.org/10.1007/s12369-020-00621-4

Malhotra, D. (2010). The desire to win: The effects of competitive arousal on motivation and behavior. *Organizational Behavior and Human Decision Processes*, 111, 139–146. https://doi.org/10.1016/j.obhdp.2009.11.005

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution*. SSOAR.

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. https://doi.org/10.2316/Journal.209.2013.1.209-0015

Nasir, J., Dillenbourg, P., Norman, U., & Bruno, B. (2020). When Positive Perception of the Robot Has No Effect on Learning. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication*, (pp. 313–320). IEEE. https://doi.org/10.1109/RO-MAN47096.2020.9223343

Schiavo, F., Campitiello, L., Todino, M. D., & Di Tore, P. A. (2024). Educational Robots, Emotion Recognition and ASD: New Horizon in Special Education. *Education Sciences*, 14(3), 258. https://doi.org/10.3390/educsci14030258

Schreiter, T., Morillo-Méndez, L., Chadalavada, R.T., Rudenko, A., Billing, E.A., Magnusson, M., Arras, K.O., & Lilienthal, A.J. (2023). Advantages of Multimodal versus Verbal-Only Robot-to-Human Communication with an Anthropomorphic Robotic Mock Driver. In *2023*

32nd IEEE International Conference on Robot and Human Interactive Communication, (pp. 293–300). IEEE. https://doi.org/10.1109/RO-MAN57019.2023.10309629

Sorrentino, A., Fiorini, L., & Cavallo, F. (2024). From the Definition to the Automatic Assessment of Engagement in Human–Robot Interaction: A Systematic Review. *International Journal of Social Robotics*, 16(7), 1641–1663. https://doi.org/10.1007/s12369-024-01146-w

Sripathy, A., Bobu, A., Li, Z., Sreenath, K., Brown, D.S., & Dragan, A.D. (2022). Teaching Robots to Span the Space of Functional Expressive Motion. In *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems,* (pp. 13406-13413). IEEE. https://doi.org/10.1109/IROS47612.2022.9981964

Su, H., Qi, W., Chen, J., Yang, C., Sandoval, J., & Laribi, M.A. (2023). Recent advancements in multimodal human–robot interaction. *Frontiers in Neurorobotics*, 17, 1084000. https://doi.org/10.3389/fnbot.2023.1084000

Tanaka, F., Isshiki, K., Takahashi, F., Uekusa, M., Sei, R., & Hayashi, K. (2015). Pepper learns together with children: Development of an educational application. In *2015 IEEE-RAS 15th International Conference on Humanoid Robots*, (pp. 270–275). IEEE. https://doi.org/10.1109/HUMANOIDS.2015.7363546

Theodotou, E. (2025). Dancing With children or dancing for children? Measuring the effects of a dance intervention in children's confidence and agency. *Early Child Development and Care*, 195(1–2), 64–73. https://doi.org/10.1080/03004430.2025.2452587

Tutul, R., Buchem, I., & Jakob, A. (2024). Towards a Gamified Human-Robot Interaction Framework for Enhanced Learning. In *17th annual International Conference of Education, Research and Innovation,* (pp. 5412–5421). IATED. https://doi.org/10.21125/iceri.2024.1322

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 *Digital Interaction and Machine Intelligence*, (pp. 123–132). Springer. https://doi.org/10.1007/978-3-031-66594-3_13

Tutul, R., Jakob, A., & Buchem, I. (2025). A dynamic template adaptation approach for noise-robust sound classification and distance determination in single-channel audio. *International Journal of Science Technology Engineering and Mathematics*, 5(1), 22–41. https://doi.org/10.53378/ijstem.353154

Tutul, R., & Pinkwart, N. (2025). Design and Evaluation of a Sound-Driven Robot Quiz System with Fair First-Responder Detection and Gamified Multimodal Feedback. *Robotics*, 14(9), 123. https://doi.org/10.3390/robotics14090123

Umbrico, A., Cesta, A., Cortellessa, G., & Orlandini, A. (2020). A Holistic Approach to Behavior Adaptation for Socially Assistive Robots. *International Journal of Social Robotics*, 12(3), 617–637. https://doi.org/10.1007/s12369-019-00617-9

Villegas-Ch, W., Buenano-Fernandez, D., Navarro, A. M., & Mera-Navarrete, A. (2025). Adaptive intelligent tutoring systems for STEM education: analysis of the learning impact and effectiveness of personalized feedback. *Smart Learning Environments*, 12(1), Article 41. https://doi.org/10.1186/s40561-025-00389-y

Wang, H., Tlili, A., Huang, R., Cai, Z., Li, M., Cheng, Z., Yang, D., Li, M., Zhu, X., & Fei, C. (2023). Examining the applications of intelligent tutoring systems in real educational contexts: A systematic literature review from the social experiment perspective. E*ducation and Information Technologies*, 28(7), 9113–9148. https://doi.org/10.1007/s10639-022-11555-x

Whittaker, S., Rogers, Y., Petrovskaya, E., & Zhuang, H. (2021). Designing Personas for Expressive Robots. *ACM Transactions on Human-Robot Interaction*, 10, Article 8. https://doi.org/10.1145/3424153

Yang, Q.-F., Lian, L.-W., & Zhao, J.-H. (2023). Developing a gamified artificial intelligence educational robot to promote learning effectiveness and behaviour in laboratory safety courses for undergraduate students. *International Journal of Educational Technology in Higher Education*, 20(1), Article 18. https://doi.org/10.1186/s41239-023-00391-9

Yerkes, R. M., & Dodson, J. D. (1908). The Relation of Strength of Stimulus to Rapidity of Habit-Formation. *Journal of Comparative Neurology and Psychology*, 18, 459-482. http://dx.doi.org/10.1002/cne.920180503

Zhang, Y., & Zhu, Y. (2022). Effects of educational robotics on the creativity and problem-solving skills of K-12 students: a meta-analysis. *Educational Studies*, 50(6), 1539–1557. https://doi.org/10.1080/03055698.2022.2107873

Zhang, X., Li, D., Tu, Y.-F., Hwang, G.-J., Hu, L., & Chen, Y. (2023). Engaging Young Students in Effective Robotics Education: An Embodied Learning-Based Computer Programming Approach. *Journal of Educational Computing Research*, 62(2), 532–558. https://doi.org/10.1177/07356331231213548

Acta Informatica Pragensia is published by the Prague University of Economics and Business, Czech Republic | eISSN: 1805-4951