Teachable Robots for Education: Bridging Children Learning and Cognitive Modelling

TARAKLI, Imene

Teachable Robots for Education: Bridging Children Learning and Cognitive Modelling

Imène Tarakli

A thesis submitted in partial fulfilment of the requirements of
Sheffield Hallam University
for the degree of *Doctor of Philosophy*

September 2025

To all those who live under oppression yet still dare to dream of freedom,
and to all who stand beside them in solidarity.

Abstract

This thesis investigates how social robots can serve as peers within the learning-by-teaching paradigm. Taking inspiration from developmental psychology, it validates a cognitive peer-model architecture grounded in Interactive Reinforcement Learning, where children teach a robot through evaluative feedback and observe its progress.

The model was evaluated in a large-scale study with primary school pupils. To maximise ecological validity, the study preserved the classroom environment and enabled multiple one-on-one child–robot interactions to occur in parallel. Findings showed that children engaged more with the robot than with independent practice, which in some cases translated into higher retention of the taught material.

Because human–robot interaction extends beyond cognitive outcomes, the thesis also examines psychological and social dimensions of teachable robots. A controlled study explored how varying levels of involvement in teaching influenced perceptions of trust, ownership, and anthropomorphism. Results indicated that while deeper involvement increased self-investment, positive perceptions depended strongly on the robot's transparency and behaviour.

Finally, the thesis proposes two new frameworks that align robot learning more closely with natural human teaching. ECLAIR enables robots to interpret the diverse verbal feedback children naturally provide, while INFORM, inspired by memory consolidation, allows robots to replay experiences offline to recover structure and generalise beyond immediate tasks.

Taken together, these contributions present a path toward robots that learn as peers: responsive, imperfect, and capable of growing alongside children. The thesis argues that making robots adaptive learners also improves the learning experience of the children who teach them.
Keywords: Teachable Robots, Reinforcement Learning, Robots in Education, Human-Robot Interaction.

I hereby declare that:

1. I have not been enrolled for another award of the University, or other academic or professional organisation, whilst undertaking my research degree.

2. None of the material contained in the thesis has been used in any other submission for an academic award.

3. I certify that this thesis is my own work. The use of all published or other sources of material consulted have been properly and fully acknowledged. I confirm that I have sought and obtained copyright permission for any third-party materials included in this thesis.

   The data presented in Chapter 5 was obtained in an experiment carried out in Bielefeld University, in collaboration with Dimitri Lacroix. I designed and implemented the reinforcement learning model, and I independently analysed and interpreted the resulting data.

   I used AI at Level 2 (AI for Shaping) of the Artificial Intelligence Transparency Scale (AITS). Specifically, I acknowledge the use of ChatGPT (https://chatgpt.com/) to assist with spelling, grammar checking, or rephrasing throughout all sections of this thesis.

4. The work undertaken towards the thesis has been conducted in accordance with the SHU Principles of Integrity in Research and the SHU Research Ethics Policy, and ethics approval has been granted for all research studies in the thesis.

Ethics review reference number	**Title of research study**	**Approval date**	**Post-approval amendments**
ER50127814	Interactive Reinforcement Learning in Simulation	16/12/2022	–
ER56422859	User Perspectives on Interactive Reinforcement Learning (Online Study)	12/07/2023	–
ER63794452	Interactive Learning with Social Robots (University Students)	24/02/2024	–
ER65839196	Learning by Teaching with Peer Robots (Primary School Children)	31/05/2024	17/07/2024

5. The word count of the thesis is 41,900.

Name	Imène Tarakli
Award	Doctor of Philosophy
Date of Submission	September 2025
College	Business, Technology and Engineering
Director(s) of Studies	Professor Alessandro Di Nuovo

Aknowledgement

Education has always held a special place in my life. The joy of learning new things has always empowered and excited me, I remember how I could hardly sleep the night before school started, eager to return to the classroom and discover something new. Pursuing a PhD always felt like a natural path, yet nothing truly prepares you for how challenging, exhausting, and at times lonely this journey can be. I have, however, been fortunate to be surrounded by fantastic people who lifted me up, encouraged me, and reminded me of why I chose this path. What follows are my sincere thanks to them.

First, I thank my Director of Study and mentor, Alessandro Di Nuovo. When I first applied for this position, it was almost on a whim. I simply did not want to regret missing the chance to pursue a subject that felt perfect for me. I never truly believed I would be accepted. Yet not only did you choose my profile, you also believed in me and supported me throughout this journey. You guided me with patience and wisdom, answering the countless questions I brought to each meeting. You never imposed your views, but instead equipped me with the tools I needed to develop my own critical thinking and research skills. When I began, I knew very little about research; and although I still have much to learn, these four years have already taught me more than I could have imagined. Thank you for believing in me, for opening doors to opportunities I never dreamt of, and for shaping the researcher I am becoming.

I also want to thank the other members of my supervisory committee, Richard Moore and Samuele Vinanzi. Richard, thank you for always being there to support me, even when circumstances changed. You made sure I had everything I needed to

carry out my studies, drawing on your network, your expertise, and your experience with human studies. I am also grateful for your sharp eye as a reviewer, your attention to detail meant that nothing ever went unnoticed. Samuele, although you joined later in my PhD, you quickly became one of my strongest supporters. You encouraged me at every step and reminded me to celebrate each achievement, no matter how small. Your positivity and constant cheering made the final stretch of this journey lighter and more joyful.

My gratitude also extends to my external supervisor, Friederike Eyssel. Thank you for your warm encouragement and support, for bringing a psychological lens to my work, and for so kindly welcoming me into your lab. Through your guidance, I came to truly appreciate the value of interdisciplinarity and the importance of being meticulous when designing a study, where every detail can influence the outcome.

I also want to thank all the teachers I have had throughout my life, who each left their mark on my journey. In particular, I am grateful to my high school mathematics and physics teachers, Mr. Bousseksou, Mr. Slimani, and Mr. Moussaoui. You always challenged me to give my best, raising the bar high and showing me that through hard work and persistence, I could achieve anything I set my mind to. Thanks to you, I learnt that effort always pays off and that there are no true limits to what I can accomplish. I am equally thankful to my engineering school teachers, Mr. Ouadjaout, Prof. Kebli, and Prof. Belouchrani. I was constantly inspired by the passion and excitement you brought to your work, and I am grateful for how you showed me that research can be not only empowering, but also deeply enjoyable.

This thesis was written during a time when the world was far from peaceful, and it was easy to feel overwhelmed by sadness and anger. In those moments, I came to realise even more deeply the importance of friendship; the people who remind you that hope is still present and who lift you up when everything feels heavy.

First, I want to thank my closest friends: Sara, Khadidja, and Amira. Even though life has taken us miles apart, to three different countries, you have always been there for me. Thank you for the joy, laughter, and lightness you continue to bring into my life. I cherish all the memories we share, and I look forward to creating many more together in the years to come.

I was fortunate to pursue my PhD within the European Marie Curie PERSEO network, surrounded by brilliant and generous colleagues, many of whom I am grateful to call friends. Thank you for creating a joyful and motivating environment. A special thanks goes to Dimitri Lacroix, with whom I collaborated most closely. Thank you for sharing your knowledge of psychology and statistics, for always taking the time to answer my questions, for reviewing parts of my thesis, and for making me feel so welcome in Bielefeld.

When I first arrived in Sheffield, I didn't know anyone, but I was fortunate to meet many kind people along the way, both among the staff at Hallam and the friends I made outside of work. Most of all, I feel lucky to be part of the most welcoming lab environment I could have wished for, one that made waking up and going to work exciting every day. When I started, our lab counted only five people; today, it has grown to more than twenty, and I feel honoured to have been part of this evolution. I want to thank every member of the lab for making daily life fun, whether it was over lunch, during coffee breaks, or while pulling through deadlines together, even in our grumpy, sleep-deprived state.

Among my labmates, I especially want to thank Aitsam Muhammad, my "lab bestie." Thank you for helping me create a home away from home, for always being there whenever I needed support, and for welcoming me into your house with your family and your warm dishes. When we first started in the lab, we were both still finding our way, and I am glad we shared those early challenges and discoveries together. I am proud of the researchers we have become.

I especially want to thank Kavyan Zoughalian, who began as a close colleague and has since come to hold a much more important place in my life. You have been my source of laughter when stress and anxiety felt overwhelming, and you have cared for me in countless ways, making sure I ate, exercised, and looked after myself. You motivated me to never settle in my research and encouraged me to always strive for more. Your ambition and energy, especially during this most challenging year, gave me the strength to keep going. Thank you for always being ready to help, even when it meant waking up at 6 a.m. to conduct a study. Most of all, you made me feel special, supported, and loved, making Sheffield feel even more like home.

I am endlessly thankful that our paths crossed, and I look forward with love and excitement to the next chapter of our life together.

Finally, I want to express my deepest gratitude to my family, who have been my constant source of love and strength.

First, to my sister, Nesrine, who never stopped asking when I would finally finish this thesis, as though she were the one writing it herself. Thank you for your constant support, for always being there when I needed you, and for making the effort to visit me despite the challenges of traveling with two kids. To my brother-in-law, thank you for patiently reviewing my motivation letters and teaching me how to write them so well that I no longer need to bother you with it. To my beautiful and amazing niece, Leila: accepting the PhD offer was the hardest choice I had to make, because it meant being away from you. While it was difficult to watch you grow from afar, your video calls and messages always brought sunshine to my days (even during the grey winters in the UK). I hope your joyfulness and sparkle will always remain with you, and that you continue to be the most wonderful princess unicorn this world could ever need. To my little baby nephew, Malik: although you are not yet a year old, you already fill my heart with so much love. Every one of your smiles melts my heart, and I cannot wait to see you grow and to spend endless moments of fun together.

And finally, to my parents, my greatest pillars and the loudest cheerleaders of my success. Thank you for always supporting me, for the sacrifices you made to ensure I had everything I needed to thrive. Thank you for the nights you stayed up with me to make sure I was ready for my exams, and for driving me across the city so I could have access to the best education. Thank you for instilling in me the values of generosity and the importance of sharing knowledge. And above all, thank you for always believing in me, even in the moments when I doubted myself.

Research Contribution

The research contributions presented in this thesis have resulted in the following peer-reviewed publications:

- **Tarakli, Imene**, Samuele Vinanzi, Richard Moore, Alessandro Di Nuovo. **"Robots and Children that Learn Together: Improving Knowledge Retention by Teaching Peer-Like Interactive Robots"**, *Computers & Education* (under review).
- **Tarakli, Imene**, Samuele Vinanzi, Alessandro Di Nuovo. **"Gender Differences in Learning-by-Teaching a Social Robot: Insights from a Primary School Study"**, *International Conference on Robot and Human Interactive Communication (RO-MAN), 2025.*
- **Tarakli, Imene** and Alessandro Di Nuovo. **"Investigating the Impact of Learning-by-Teaching a Social Robot on Students with Different Prior Knowledge Levels"**, *18th AAATE Conference, 2025.*
- **Tarakli, Imene**, Samuele Vinanzi, Alessandro Di Nuovo. **"Interactive Reinforcement Learning from Natural Language Feedback"**, *International Conference on Intelligent Robots and Systems (IROS), 2024.*
- **Tarakli, Imene**, Alessandro Di Nuovo. **"Personalised Interactive Reinforcement Learning with Multi-Task Pre-training"**, *Human Friendly Robotics Workshop, 2024.*

- **Tarakli, Imene**, Alessandro Di Nuovo. **"Learning while Sleeping: Integrating Sleep-Inspired Consolidation with Human Feedback Learning"**, *International Conference on Developmental Learning (ICDL), 2024.*.
- **Tarakli, Imene**, Alessandro Di Nuovo. **"User Perception of Teachable Robots: A Comparative Study of Teaching Strategies, Task Complexity and User Characteristics"**, *International Conference on Social Robotics, 2023.*

Table of contents

List of figures

List of tables

Chapter 1

Introduction

"I'll teach you how to jump on the wind's back, and then away we go."

— J.M. Barrie, *Peter Pan*

Learning is a continuous and layered process that shapes the human experience [95, 290]. From infancy, coordinating movement or producing simple sounds, to complex milestones such as language acquisition and abstract reasoning, children develop through ongoing interaction with their environment and with others [209, 34]. Early development draws on associative learning, imitation, and reinforcement [253, 16, 168]. As cognitive capacities expand, children increasingly reflect on their own thinking [91], generate independent ideas [99], and consider the perspectives of others [287]. These skills support collaborative knowledge construction and the regulation of one's understanding, capabilities that are regarded as distinctively human.

Social interaction plays a fundamental role in this developmental process. Through engaging with others, children are exposed to new perspectives, they test and refine their ideas, and they reorganise their understanding in response to what others know, believe, or expect. Sometimes they learn directly from others, and sometimes they learn by trying to explain, negotiate, or make sense of what others might be thinking. In this way, interaction becomes not just a setting for learning, but a mechanism through which it unfolds [280, 37, 243].

Peer-to-peer learning builds directly on this idea by placing learners in symmetrical roles, where knowledge is developed collaboratively rather than transmitted from expert to novice. When children learn with peers, especially when they are asked to explain their reasoning or engage with differing views, they often interact more deeply with the material, restructuring their thinking in the process [232]. One powerful form of this is learning-by-teaching, where the act of explaining to another becomes an opportunity to consolidate and extend one's own understanding [79].

While peer-to-peer learning holds great educational value, it also presents practical challenges. Not all children benefit equally from peer interactions, and the dynamics of who teaches whom, how support is exchanged, and whether the interaction is productive can be difficult to control in classroom settings [62, 92]. In this context, social robots have emerged as promising tools to support and structure peer-like interactions. Unlike traditional digital tools, social robots bring physical embodiment into the classroom, allowing them to engage children more directly through gaze, gesture, and shared space, and offering a responsive and socially engaging partner for learning [24, 135, 140] . When designed as novice peers, these robots invite children to take on the role of teacher, engaging in learning-by-teaching scenarios where the robot's behaviour is shaped by the child's input. In doing so, they harness the cognitive benefits of teaching while providing a peer interaction that is both adaptive and consistent, qualities that are difficult to achieve in traditional classroom peer learning [263, 158, 82].

This thesis investigates how interactive learning models can be used to endow social robots with realistic peer-like behaviours that evolve through teaching interactions with children. It explores how these models can support natural, multimodal communication, adapt to individual learners, and generalise knowledge beyond specific tasks. By grounding the robot's learning in children's input, this work aims to create more authentic and cognitively plausible learning companions. In doing so, the thesis bridges insights from developmental psychology and cognitive robotics and contributes novel architectures, empirical findings, and ethical reflections on how teachable robots can be integrated into real educational environments. This includes the design of interactive reinforcement learning models, large-scale classroom deployments, and the study of psychological outcomes such as trust and psychological ownership.

1.1 Challenges

Designing a robot that can act as a realistic, teachable peer is a complex task. It requires more than just replicating correct answers or using engaging animations. For a peer robot to be meaningful in a classroom setting, it must be able to learn from each individual child, interact naturally, respond appropriately, and adapt its behaviour over time. This section outlines the core challenges addressed in this thesis.

1.1.1 Realistic Cognitive Modelling and Natural Testing

Building a learning model that truly reflects what each child teaches the robot is not straightforward. One-size-fits-all approaches do not work in real classrooms; each child brings their own pace, strategies, and way of expressing feedback [235]. For a robot to act as a meaningful peer, it needs to personalise its learning trajectory to each child's input [246, 24]. At the same time, running studies in real school environments adds another layer of complexity. Instead of isolated lab settings, the studies in this thesis span a range of contexts, including real classroom deployments where several children interacted with different robots at the same time. Managing multiple autonomous, personalised robots in parallel, while preserving the natural flow of the classroom, is a core challenge addressed in this work [132, 135, 246].

1.1.2 Psychological Implications

Children do not just treat robots as machines. They often form emotional connections, similar to the bonds they build with peers [131, 149]. This can be beneficial; it makes the interaction more engaging, but it also carries risks. Over-trust may lead to over-reliance, where children defer too readily to the robot's answers or guidance. Emotional attachment may also create expectations the robot cannot fulfil [173, 250], and in some cases children may even prefer interacting with the robot over peers [59]. In the context of peer robots, striking a balance is therefore crucial: the robot should be engaging and trustworthy enough to sustain interaction, but not so realistic that children attribute human-like authority or understanding [130, 249]. This thesis examines how involvement in teaching affects how participants perceive and trust the robot, and whether this leads to deeper psychological engagement or misplaced confidence.

1.1.3 Natural Interaction

One major limitation in many robot learning studies is that they constrain how users are allowed to interact. Children often have to press buttons, follow strict prompts, or stick to predefined answers [216]. But real communication doesn't work that way, especially with children, who naturally use language to guide, correct, or evaluate others [230]. To make the learning experience more authentic, the robot must be able to interpret different types of spoken input and respond in flexible, context-aware ways [149, 135]. This thesis tackles this challenge by enabling the robot to learn from three forms of natural language feedback (evaluative, corrective, and guidance) so that learning can happen through genuine dialogue, not just structured input.

1.1.4 Lack of Generalisation

Most interactive robot learning models focus on learning one task at a time, optimising for immediate feedback. But human learning is rarely that narrow. We build generalisable knowledge: we abstract patterns, infer rules, and apply what we learn in one situation to new ones [271, 88]. To build more child-like peer robots, we need to move beyond short-sighted, task-specific learning. This thesis addresses this by proposing a new model that enables the robot to consolidate its experiences over time and generalise beyond the immediate teaching task, drawing inspiration from the way children abstract and transfer knowledge across contexts, though the underlying mechanisms remain distinct.

1.2 Research Goals and Questions

This thesis aims to develop teachable robots that can act as adaptive, responsive, and cognitively plausible peers in classroom settings. By drawing on developmental psychology and interactive reinforcement learning, the goal is to create systems that do not simply execute predefined behaviour but instead learn from children's input, mirroring how real peer interactions support learning.

Based on the challenges identified, this thesis pursues five central research goals.

First, it seeks to design a robot learning model that adapts in real time to children's feedback, addressing the challenge of creating peer-like robots that are genuinely responsive. This leads to the first research question:

RQ1: *How can interactive reinforcement learning be used to enable children to effectively teach a robot using evaluative feedback?*

Second, the thesis examines how children respond to such a system in an ecologically valid classroom setting, addressing the challenge of moving beyond lab-based studies.

RQ2: *What learning and retention outcomes emerge when children teach a robot compared to when they practise alone?*

Third, the work considers the psychological and social dynamics of interaction, focusing on involvement, trust, and psychological ownership.

RQ3: *How does a user's level of involvement in the robot's learning process affect psychological factors such as trust and psychological ownership?*

Fourth, the thesis addresses the challenge of enabling natural interaction by exploring different kinds of feedback.

RQ4: *How do evaluative, corrective, and guidance feedback influence the interaction and the robot's learning?*

Finally, the work tackles the challenge of long-term generalisation by developing a consolidation-inspired model.

RQ5: *Can a robot consolidate and generalise its learning beyond the specific task, in a way that mirrors human learning?*

1.3 Organisation of the Thesis

The remaining chapters of this thesis are organised as follows:

- **Chapter 2:** Provides a multidisciplinary foundation, drawing on developmental psychology, human-robot interaction, and reinforcement learning. It discusses the theoretical basis for peer learning, the role of social robots in education, and existing models of teachable robots.
- **Chapter 3:** Introduces a novel interactive reinforcement learning framework that enables a robot to learn from child-provided feedback. This chapter details the architecture design and an online study comparing evaluative and preference-based learning approaches.
- **Chapter 4:** Describes an ecologically valid classroom study where children taught a robot across two learning tasks. It compares the learning outcomes and engagement levels of children in a learning-by-teaching condition versus a tablet-based self-practice condition.
- **Chapter 5:** Investigates how involvement in teaching a robot influences user's trust, psychological ownership, and engagement. A lab-based study explores how different levels of control affect user perception and epistemic trust.
- **Chapter 6:** Extends the robot's capabilities by allowing it to learn from different types of natural language feedback (evaluative, corrective, and guidance) using large language models. This chapter explores how this improves the naturalness and effectiveness of the interaction.
- **Chapter 7:** Proposes a sleep-inspired reinforcement learning model that enables the robot to consolidate and generalise its learning beyond the immediate task. This chapter evaluates the approach in simulated environments to demonstrate its long-term potential.

- **Chapter 8:** Reflects on the findings across all studies, discusses broader implications for educational robotics and developmental theory, and addresses key limitations and ethical considerations.

- **Chapter 9:** Summarises the main contributions of the thesis and offers final reflections on the future of teachable robots in education.

Chapter 2

Background

[Epigraph removed for copyright reasons. Original source: J.R.R. Tolkien, The Hobbit*, 1937]*

2.1 Developmental Models for Learning

Understanding how children learn has long been a central concern in developmental psychology and education. Over time, theoretical and empirical work has converged on the idea that learning is not solely an individual cognitive process, but one shaped by rich social, linguistic, and cultural interaction. This subsection presents a chronological overview of influential developmental theories, with a focus on how social interaction, dialogue, and peer conflict contribute to cognitive growth.

2.1.1 From Individual Construction to Social Interaction

In the early 20th century, Jean Piaget advanced one of the most influential theories of cognitive development [208]. He argued that children are active constructors of knowledge who progress through stages by interacting with their environment. His framework emphasised individual exploration, where cognitive growth occurs through *equilibration*; children's efforts to maintain cognitive balance between their internal schemas and external experiences. Although Piaget's theory centred on the solitary child, later research expanded his ideas by showing that social contexts such as peer discussion, especially with partners who hold different views, can also stimulate cognitive conflict and conceptual change [181, 73].

In parallel, Lev Vygotsky offered a contrasting perspective that placed social interaction at the heart of development [280]. His concept of the Zone of Proximal Development (ZPD) describes the gap between what a child can achieve alone and what becomes possible with guidance from a more knowledgeable partner. Importantly, this "other" need not be an adult; peers can play an equally significant role, particularly when collaboration is structured to enable dialogue and mutual support. Here the emphasis shifts from individual discovery to socially scaffolded participation, a perspective that continues to shape contemporary views of collaborative learning.

2.1.2 Learning through Language and Scaffolding

In the decades that followed, Jerome Bruner helped formalise the notion of *scaffolding*; the idea that learning occurs most effectively when support is tailored to a child's current level and gradually withdrawn as competence grows [293]. Bruner's work also reinforced the central role of language in cognitive development, emphasising that interaction does not merely transmit knowledge but actively shapes the learner's reasoning. Educational settings that encourage verbal explanation, questioning, and shared meaning-making became particularly valuable for promoting conceptual understanding [53]. This line of research contributed to a broader shift in how educational theory conceptualises the learner: not as an isolated thinker but as a participant in a community of discourse. Collaborative learning frameworks built on these ideas have highlighted that learning happens with others, through shared activity and the co-construction of meaning [180].

2.1.3 Toward Peer-Based Learning

Building on earlier research into social interaction and co-construction, later work turned to the specific cognitive benefits of peer teaching. In this context, children are not only recipients of guidance but also take on the role of teacher. The idea that teaching others enhances one's own understanding dates back to classical thinkers such as Seneca - *"Teaching is learning"*, and Comenius - *"He who teaches, learns"*, but it was not until the mid-twentieth century that empirical research began to examine its educational value systematically. Early studies on peer tutoring revealed that student tutors often demonstrated greater learning gains than their tutees, suggesting that the act of teaching actively consolidates knowledge [60].

The underlying mechanisms of this effect have been examined in more detail. Teaching requires the tutor to retrieve and organise information, monitor the learner's understanding, and adapt explanations accordingly [232]. These demands engage meta-cognitive functions such as planning, self-assessment, and reflection, all of which are known to support deeper learning [79]. The impact is further reinforced by what has become known as the *protégé effect*, whereby tutors experience a heightened sense of responsibility for the learner's success, increasing their own effort and investment in the task [46].

Taken together, these studies suggest that peer teaching benefits both tutees and tutors by stimulating explanation, monitoring, and reflective processes linked to deeper understanding. This body of work has laid the groundwork for extending learning-by-teaching paradigms into new domains. In particular, researchers have begun to investigate whether similar dynamics can be elicited in interactions with technology, and more recently, with social robots.

2.2 Social Robotics in Education

As educational technologies have evolved, increasing attention has been paid to how they might support not only cognitive development but also the social and emotional dimensions of learning. Within this landscape, social robots have emerged as a distinctive tool, designed to engage learners through speech, gesture, gaze, and other social cues. Unlike screen-based systems or virtual agents, robots share physical space with learners and can be framed as peers or tutors, enabling more natural and socially grounded interaction. A growing body of research suggests that such embodied agents can enrich children's learning experiences, supporting cognitive gains while also fostering motivation, confidence, and social skills [24]. This section reviews recent findings on the use of social robots in education, with a focus on four key dimensions: the role of physical embodiment and social presence, their impact on learning outcomes, their contribution to social–emotional engagement, and their capacity for personalised interaction.

2.2.1 The Role of Physical Embodiment and Social Presence

One defining feature of social robots is their physical embodiment. Unlike virtual agents or screen-based characters, robots can share physical space with learners, make eye contact, use gestures, and respond in real time [164]. Prior work suggests that such affordances can enhance perceptions of social presence, influencing how children attend to and engage with the robot [24, 164]. Empirical studies have reported that physically embodied robots are often perceived as more engaging and enjoyable than screen-based characters, and can lead to higher learning gains in some contexts [138, 150]. Moreover, physical presence has been associated with more positive perceptions and longer interaction times compared to virtual agents [215], supporting the idea that embodied robots can function as social partners. Such opportunities for shared activity resonate with Vygotskian perspectives on the centrality of joint interaction in learning [280].

Recent conceptual models extend this view by emphasising the role of embodiment in sustaining affective as well as cognitive engagement. Chen et al. [47] propose that physical co-presence enables a robot to maintain a socially responsive presence and to adapt its role and behaviour in ways that support the learner's flow during interaction. Empirical findings provide converging evidence on the cognitive dimension of this claim. For instance, Leyzberg et al. [162] found that children tutored by a physically present robot solved puzzles more efficiently and improved more rapidly than those working with a disembodied voice or video. Similarly, Vrins et al. [279] showed that children who interacted with an embodied robot achieved significantly higher vocabulary scores and reported greater engagement than peers in a video-based condition. Together, these results suggest that physical embodiment can capture attention and facilitate sustained cognitive performance, while conceptual models highlight its potential to also support affective engagement. In a geography task, Jones et al. [128] reported that children who received feedback from a physically embodied robot expressed greater enjoyment, trust, and perceived understanding than those in screen-only conditions. Kennedy et al. [134] similarly found that children working with a physically present robot displayed more frequent and sustained gaze, indicating higher social attention and responsiveness. Although immediate learning gains did not differ significantly, the embodied robot proved more effective at establishing joint attention, which is often regarded as an important foundation for collaborative learning. In a broader synthesis, Li [164] concluded that physical co-presence, rather than embodiment alone, tends to produce stronger effects on users' attentiveness, compliance, and affective responses across educational and social contexts.

Embodiment also plays a role in non-verbal scaffolding. For instance, de Wit et al. [69] showed that iconic gestures—movements that visually represent concepts—can enhance vocabulary learning in second-language instruction. Children who interacted with a gesturing robot demonstrated greater vocabulary gains and higher engagement throughout the task. These findings resonate with Bruner's concept of scaffolding, in which multimodal cues help bridge new knowledge within a child's developmental zone [293]. At the same time, the design of embodied cues must be developmentally appropriate. Kennedy et al. [135] caution that mismatches between gesture and speech may confuse learners, and excessive movement can become distracting. This concern is particularly relevant for younger children, who are still developing attentional control and are more sensitive to irrelevant stimuli [292]. Poorly timed or unnecessary gestures may therefore hinder, rather than support, learning.

Overall, physical embodiment enhances the quality of child–robot interaction by enabling joint attention, non-verbal responsiveness, and emotionally grounded communication. When designed thoughtfully, it can create opportunities for socially meaningful and developmentally aligned learning, supporting both cognitive and interpersonal dimensions of education in ways that resonate with Vygotskian and Brunerian perspectives.

2.2.2 Learning Outcomes and Cognitive Benefits

Beyond fostering engagement, social robots can activate a range of cognitive processes that support meaningful and lasting learning. Rather than merely increasing attention or enjoyment, well-designed robot interactions have been shown to promote conceptual understanding, strategic reasoning, and reflective habits [24]. A growing body of evidence suggests that robot-assisted learning can support cognitive development through mechanisms such as scaffolded challenge, metacognitive prompting, and self-regulated learning [20, 202].

Social robots can guide learners through increasingly complex tasks by introducing stepwise challenges and providing real-time support. For example, in a mathematics tutoring study, Ligthart et al. [165] found that scaffolded dialogue with a socially grounded robot significantly improved both accuracy and speed in solving multiplication problems. The robot gradually adapted the difficulty through narrative-based tasks, guiding children from factual recall to structured application. Similarly, Ong et al. [195] demonstrated that a robot peer tutor helped first-grade students solve math word problems using prompts based on Polya's four-step framework: understanding the problem, devising a plan, executing the plan, and evaluating the result. Most children improved their performance, and teachers valued the robot's ability to structure the problem-solving process in real time. Beyond task-specific scaffolding, Resing et al. [224] used a dynamic testing design to examine children's performance on the Tower of Hanoi puzzle. A peer robot delivered graduated prompts that helped learners reduce errors and improve planning efficiency in post-tests, suggesting that such systems may strengthen executive functioning through interactive guidance.

Preliminary neuroscientific evidence also points to the cognitive impact of robot-mediated learning. Alimardani et al. [7] found that children learning with a physically present robot showed stronger neural activation in brain regions associated with memory and attention than peers using a screen-based interface. This neural activity was positively correlated with individual learning gains, suggesting that physical embodiment and social responsiveness can deepen mental engagement at a biological level.

Beyond task performance, social robots can encourage metacognitive awareness and learning autonomy. Jones et al. [127] investigated how varying levels of scaffolding affected self-regulated learning (SRL) during a geography task. Using an Open Learner Model, the robot supported learners in setting goals, monitoring progress, and choosing strategies. Children who received adaptive SRL feedback showed the highest learning gains and engagement, particularly those with lower prior ability, indicating that timely, personalised prompts can foster independence as well as achievement. Metacognitive prompting also plays a valuable role in language learning. In a second-language acquisition study, Schodde et al. [241] found that a robot tutor combining adaptive feedback with explanation improved vocabulary acquisition in slower learners. The robot not only adjusted its instructional behaviour based on learner responses but also used re-engagement strategies when children appeared distracted. These interactions helped learners reflect on their progress and stay focused, reinforcing understanding through dialogue. Similarly, Edwards et al. [80] showed that prompting children to justify their answers and engage in discussion, rather than relying on rote memorisation, supported the development of critical thinking and argumentation skills. These findings suggest that robot-mediated dialogue may encourage deeper reasoning through collaborative meaning-making.

Taken together, this evidence suggests that social robots can serve as tools for enacting developmental principles in practice, operationalising scaffolding, guided participation, and metacognitive prompting within real learning contexts.

2.2.3 Social and Emotional Development

Children's learning is shaped not only by what they think, but also by how they feel: about themselves, the task, and those around them [121]. Emotions such as curiosity, confidence, and anxiety can influence whether a child embraces a challenge or avoids it, and whether they persist or give up [207]. As Bandura [16] explains, children learn not only through direct experience but also by observing others and interpreting social cues. Through such social modelling and reinforcement, they develop emotional tendencies and motivational beliefs. In this context, social robots, with their physical presence and interactive behaviours, can serve as social models that support learning in ways extending beyond traditional instruction.

Motivation is one area where robots have shown particular promise. Self-Determination Theory (SDT) [70] posits that fulfilling learners' needs for autonomy, competence, and relatedness enhances intrinsic motivation and engagement. This framework has been applied to social robots; for example, van Minkelen et al. [273] found that preschoolers who interacted with a robot addressing all three needs maintained higher engagement over time compared to control conditions. Beyond SDT, other research highlights the motivational value of socially responsive behaviours. Donnermann et al. [75] reported that university students who worked with a robot that personalised interaction through names, small talk, and encouragement showed higher motivation and achieved better academic outcomes than those who did not.

Robots have also been shown to support perseverance and resilience. In a study by Park et al. [203], children who interacted with a peer-like robot that modelled a growth mindset displayed greater persistence on difficult tasks and reported stronger growth mindset beliefs than those in a neutral condition. Importantly, children recognised the robot's mindset, suggesting that robots may act as role models shaping motivational dispositions.

Beyond perseverance, robots may encourage more expansive forms of expression such as creativity and perspective-taking. Ali et al. [5] found that children working with a creatively expressive robot generated ideas with greater fluency, novelty, and originality. These effects illustrate the modelling capacity of robots, where children adopt not only content but also expressive styles and emotional tone. Wang et al. [281] extend this argument in their framework of "creative accessibility," describing how multimodal interaction with social robots can broaden children's opportunities for imaginative and exploratory learning.

Interpersonal growth is another domain where robots show potential. Derakhshan et al. [72] found that students who practiced medical English with an empathetic robot exhibited higher levels of confidence and empathy in communication tasks. Similarly, Park et al. [202] showed that robots capable of backchanneling—offering social cues like nods and affirmations—encouraged richer storytelling and greater verbal output in children, supporting both confidence and communicative fluency.

Robots may also influence how learners seek help and manage uncertainty. Hei et al. [113] found that students were more likely to ask for assistance from a robot tutor than from a human teacher, particularly when the robot had previously been introduced in a friendly manner. This effect appears linked to reduced social anxiety and a greater sense of psychological safety, enabling learners to admit difficulty without fear of judgment.

However, the emotional influence of robots is not always beneficial. Bruzzo et al. [38] reported that participants interacting with a socially expressive robot—using eye contact, personalised greetings, and warmth—were more likely to accept incorrect suggestions than those working with a neutral robot. While social presence increased trust, it also appeared to reduce critical evaluation, underscoring the need to balance emotional engagement with instructional accuracy.

Overall, these findings suggest that social robots can shape not only task performance but also broader affective and interpersonal development, including resilience, creativity, empathy, and help-seeking behaviours. At the same time, the risks of over-trust highlight the importance of careful design to ensure that emotional engagement supports, rather than undermines, independent judgment. Thoughtfully designed, socially responsive robots may create conditions that foster motivation, emotional confidence, and openness to collaboration, contributing to children's social and emotional growth.

2.2.4 Personalisation and Adaptive Interaction

Personalised interaction is central to effective education, enabling instruction to meet learners where they are intellectually, emotionally, and socially [267]. In human teaching, this personalisation arises from the teacher's ability to interpret student responses and adapt instruction in real time. Through adaptive capabilities, social robots can approximate the responsiveness of human teaching [24]. In this section, personalisation strategies are reviewed with a focus on cognitive adaptation (adjusting the teaching content), affective adaptation (adjusting the robot's affect), behavioural adaptation (responding to non-verbal cues), and role dynamics.

Cognitive personalisation refers to the robot's ability to track the learner's evolving knowledge and adjust content difficulty accordingly. Studies have implemented this in different ways. In a second-language vocabulary task, Schodde et al. [240] used Bayesian Knowledge Tracing (BKT) to enable a robot to select tasks based on learners' mastery levels, which led to significantly higher in-session performance compared to a non-personalised baseline. Ramachandran et al. [219] extended this approach by using a Partially Observable Markov Decision Process (POMDP) model that considered both knowledge and engagement states when determining the robot's next help action, resulting in higher mathematics learning gains. The benefits of cognitive adaptation have also been shown in longer-term studies. Leyzberg et al. [160] demonstrated that when robots adapted lesson sequences to target individual learners' weakest skills across sessions, learning gains reached the 98th percentile. Even simple mechanisms, such as ordering lessons based on skill level, can yield substantial improvements, as reported in earlier work by Leyzberg et al. [161]. In a classroom setting, Baxter et al. [20] found that a personalised peer robot that adapted task difficulty, speech style, and use of names significantly improved children's learning, especially on material not covered in class. More recently, Maaz et al. [175] combined cognitive and emotional assessment to place students into performance bands and tailor mathematical instruction, yielding an 8% improvement over control.

Beyond knowledge and skills, personalisation must also align with the learner's emotional state. Affective adaptation enables the robot to interpret and respond to emotions such as frustration, boredom, or curiosity. Several studies show that emotional adjustment can strongly influence motivation and engagement. In a mathematics study with the Ms. An robot, for instance, adapting instructional strategies to students' emotions led to deeper reflection and sustained attention, even though final scores were comparable across conditions [166]. Donnermann et al. [76] likewise reported that socially adaptive robots, which modulated tone and encouragement in response to learner feedback, improved students' perceptions of the interaction. At the same time, the increased verbosity, such as detailed reviews of mistakes and performance, sometimes caused frustration and stress, highlighting the delicate balance between responsiveness and user comfort. A different approach has explored reinforcement learning (RL) as a mechanism for affective adaptation. Gordon et al. [100] trained an RL-based robot to adjust its affective behaviour to children's emotional states during second-language instruction. While children's responses revealed clear individual differences, and the learned policies adapted accordingly, convergence was slow even after multiple sessions, underscoring the challenge of affective personalisation in sparse-data environments.

Personalisation also extends to the behavioural level, where robots adjust their social signalling, gestures, gaze, and prosody in response to the learner's non-verbal cues. This form of adaptation supports engagement by maintaining a sense of attunement and interpersonal synchrony. Park et al. [202] showed that when a robot adapted its backchanneling behaviour, using nods and gaze based on children's storytelling rhythms, children perceived it as more attentive and remained more engaged. Bourguet et al. [29] extended this idea by integrating facial expression and posture recognition to trigger motivational behaviours; affect-responsive robots were consistently rated as more supportive. However, as Molenaar et al. [183] found, not all behavioural adaptations translate to improved learning: a robot matching children's speech pitch had no effect on vocabulary recall, pointing to the need for careful alignment between adaptive cues and pedagogical goals.

A further dimension of personalisation involves flexible role dynamics. Rather than remaining fixed as either a tutor or peer, a robot may adapt its social role to the learner's needs. In a study by Chen et al. [48], a robot that alternated between peer- and tutor-like roles improved word retention and promoted richer emotional engagement compared to fixed-role counterparts. These findings suggest that personalisation is not only about tailoring content or behaviour, but also about enabling the robot to shift its relational stance in ways that sustain both learning and motivation.

While these studies demonstrate the promise of personalisation, they also highlight its limits. Over-personalisation can backfire when adaptation becomes too narrow or repetitive. Gao et al. [96] observed that robots adapting too closely to user preferences reduced engagement, particularly when behaviours converged into predictable patterns. In contrast, children responded more positively to varied and flexible behaviours. This distinction underscores the importance of designing adaptive systems that remain context-sensitive and socially natural, rather than mechanically formulaic.

Ultimately, the role of personalisation is not just to optimise task performance, but to cultivate meaningful, individualised learning relationships. Robots that flexibly adjust to learners' cognitive states, emotional needs, and social cues are more likely to foster engagement, motivation, and trust—factors that are important in children's learning.

2.3 Learning-by-Teaching Social Robots

Social robots can take on different roles in education, including that of teacher, tutor, peer, or learner [24]. Each role fundamentally shapes the nature of child–robot interaction, as it frames the robot's authority, competence, and relational stance. These role assignments, in turn, influence how children engage with the robot, how much trust they place in its input, and how they perceive their own responsibilities in the learning process. This section first reviews how such roles are conceptualised, implemented, and perceived in educational HRI. It then turns to empirical work on learning-by-teaching paradigms, before examining the computational models that support role-appropriate behaviours and autonomy.

2.3.1 Social Roles of Robots in Education

Early work in educational HRI often positioned robots as teachers or tutors, emphasising their instructional authority. Teacher roles are typically framed as authoritative and content-driven. Robots in this role often deliver structured lessons to groups of learners in a one-on-many format, mimicking traditional classroom instruction [275, 3]. The interaction is generally unidirectional, with limited responsiveness to individual learner needs. While this setup can be effective for conveying information, it provides limited opportunities for personalised feedback or learner agency.

Tutor roles, in contrast, are more commonly used in one-on-one scenarios. Tutor robots support learners through scaffolded dialogue, personalised prompts, and adaptive feedback. This framing fosters more interactive, learner-centred engagement and is by far the most widely used in the field, accounting for nearly half of all robot roles in educational HRI studies [24].

More recent approaches have explored peer and learner roles, where interaction is more symmetrical and collaborative. In these roles, the robot is framed as a socially equal or cognitively relatable partner, which can reduce hierarchical distance and encourage more open learner behaviour. For example, Zaga et al. [298] found that children were more engaged and performed better when the robot was framed as a peer rather than a tutor. Similarly, Salomons et al. [238] showed that adult learners with low prior knowledge achieved greater learning gains when interacting with a peer robot that used inclusive language (e.g., "we" and "us") compared to a tutor robot that used directive language (e.g., "you"). Participants also rated the peer robot as more friendly, respectful, and intelligent.

This growing interest in peer roles reflects a broader shift toward reciprocal and socially grounded learning interactions. Evidence suggests that different roles emphasise different strengths: tutor roles often yield greater instructional efficiency and knowledge gains, while peer roles are particularly effective at fostering social involvement, motivation, and emotional expressiveness. For instance, Chen et al. [49] found that children interacting with a tutor robot learned more vocabulary than those working with a novice tutee, but displayed fewer facial expressions and less affective engagement. This highlights the importance of aligning role design not only with cognitive goals but also with socio-emotional learning outcomes.

Peer roles themselves encompass distinct pedagogical configurations. Educational theory typically differentiates between *peer learning*, described as a "two-way reciprocal learning activity" in which participants share knowledge and ideas on equal terms, and *peer tutoring*, which "implies an unequal partnership" due to one learner assuming a position of responsibility as the tutor [137]. Social robots can be designed to support either format. In peer learning contexts, the robot and child collaborate toward shared objectives, modelling cooperative reasoning and mutual exploration. For instance, Ishida et al. [122] found that children interacting with a co-learning robot adapted their vocabulary and speech to align with the robot's utterances, indicating that peer-like configurations can promote linguistic alignment and the co-construction of meaning.

In peer tutoring scenarios, however, the robot is often presented as a novice learner, making intentional errors that prompt the child to take on the teaching role. This configuration is especially well suited for implementing learning-by-teaching approaches, where children consolidate their own understanding by guiding the robot's learning process. By exhibiting fallibility (asking for help, making mistakes, or gradually improving over time), robots can prompt children to adopt the role of teacher [263]. While these dynamics are explored in more detail in the next section, their pedagogical value is already evident. Sudo et al. [261] showed that children who taught a robot that repeatedly made grammatical errors reported feeling more comfortable making mistakes themselves, even though learning gains did not significantly improve. These findings suggest that robot fallibility can reduce performance pressure and encourage persistence, particularly in emotionally sensitive learning contexts.

The specific role assigned to a robot fundamentally shapes the learning dynamic, influencing the learner's behaviour, perception of the interaction, and the pedagogical possibilities it affords. While teacher and tutor roles provide structure, expertise, and instructional efficiency, peer and novice roles offer greater potential for fostering learner autonomy, emotional engagement, and shared meaning-making. By positioning the child as an active participant rather than a passive recipient, these roles can support richer, more reflective interactions. As educational HRI continues to evolve, peer-based approaches, and particularly those involving learning-by-teaching with novice robots, are emerging as a promising direction for creating interactive, relational, and developmentally inspired learning experiences.

2.3.2 Empirical Insights into Learning-by-Teaching

The learning-by-teaching (LbT) paradigm, long recognised in educational theory, has only recently begun to gain traction in child–robot interaction research. While still less common than tutor-based approaches, a growing body of studies has begun to explore how children's role as teachers can be modelled in interactions with robots.

Early work demonstrated that children are capable of guiding a robot's learning and that this process benefits their own understanding. Tanaka and Matsuzoe [263] showed that children who taught a robot English vocabulary through corrections and explanations achieved higher retention and engagement than controls. Similarly, Lemaignan et al. [158] used handwriting as a teaching medium, with children providing demonstrations and feedback that helped the robot's writing improve over time. This setup prompted children to reflect more carefully on stroke formation and spatial organisation, illustrating how LbT can encourage meta-cognitive monitoring. Later studies explored how LbT unfolds in richer instructional contexts. Verhoeven et al.[276] found that teaching vocabulary through storytelling with a robot encouraged children to provide elaborated explanations, gestures, and contextualisation. Pareto et al.[201] compared tutoring a robot to tutoring a younger peer, showing that children offered similar levels of scaffolding and feedback in both conditions, suggesting that appropriately framed robot tutees can elicit authentic teaching behaviours comparable to those evoked by human partners. Moving beyond one-on-one interaction, El Hamamsy et al. [82] showed that when groups of children collaborated to teach handwriting to a robot, they engaged in joint reflection regardless of whether the robot prompted task-based or social engagement.

A central theme in more recent research concerns the robot's fallibility and its effects on teaching strategies. In a study conducted by Chandra et al. [45], children interacted with a robot that either learned continuously, learned based on personalization, or did not learn at all. Results showed that children's handwriting improved significantly when the robot exhibited learning behavior, and children were more engaged when the robot demonstrated progress. Similarly, Bowman-Smith et al. [30] demonstrated that when robots made either typical or atypical errors, children provided more elaborate and persistent teaching than when no errors occurred, with atypical mistakes prompting the deepest engagement.

The framing of the robot's behaviour also shapes children's emotional and motivational responses. Muldner et al. [188] showed that attribution style matters: children responded more positively when the robot explained success in terms of effort and ability, and when it used inclusive language such as "we" rather than "I." In contrast, self-deprecating explanations ("I'm not smart") elicited frustration. These results suggest that how a robot communicates its learning trajectory influences not only how children teach but also how they feel about teaching.

Finally, studies have highlighted the role of interaction design in supporting LbT. Johal et al. [126] found that a side-by-side seating arrangement encouraged more positive feedback than face-to-face positioning. Lindberg et al. [169] compared teaching robots with virtual agents, showing that the physical robot captured more attention early in the session, though some children's belief that the robot already knew the material reduced their willingness to teach, emphasising the role of children's prior assumptions in shaping motivation.

Across these strands, evidence shows that LbT with robots can elicit instructional behaviours such as explaining, correcting, and encouraging, which in turn foster children's own learning and reflection. However, the effectiveness of this paradigm depends not only on the robot's fallibility and perceived responsiveness, but also on how its role and behaviour are framed.

2.3.3 Computational models for peer-to-peer tutoring

In LbT scenarios, the success of the interaction depends not only on the child tutor's ability to convey knowledge, but also on the tutee's ability to behave like a credible learner [232]. The tutee, here, the social robot, must exhibit novice-like behaviour to trigger the Protégé Effect, a psychological phenomenon that increases the tutor's sense of responsibility for the tutee's learning [46]. Over time, the robot should demonstrate measurable progress, fostering reflective knowledge-building through interaction [27]. For effective peer-to-peer learning, the robot must follow a learning trajectory that begins with novice-level performance and gradually advances, encouraging the child to deepen their own understanding [233].

To support such interactions, various computational models have been developed to simulate novice-like behaviour in social robots. These models aim to present the robot as a peer who is learning from the child, thereby encouraging tutoring behaviours and cognitive engagement. Across the literature, three primary approaches can be identified: Wizard-of-Oz (WoZ) control, scripted simulation, and learning-based adaptation. Each offers different affordances and limitations in creating robots that behave like credible, teachable agents.

Wizard-of-Oz systems rely on human operators to control the robot's behaviour behind the scenes. These systems offer high flexibility and responsiveness, enabling smooth and engaging interactions. For example, studies by Tanaka et al. [263], Verhoeven et al. [276], and Das et al. [67] used WoZ setups to explore how children teach robots through natural language, and collaborative dialogue. Although effective for early-stage experimentation, WoZ models lack real autonomy and cannot support scalable deployment in classroom environments. They provide no actual learning or memory on the robot's side, and their reliance on human intervention limits their use in autonomous educational agents.

Scripted models go a step further by simulating learning through pre-defined behaviours and staged timing, creating the appearance of progression without underlying adaptation. In Zhexenova et al. [300], for example, children were asked to teach a NAO robot a new Latin-based Kazakh script. Although the robot was framed as a novice learner, its responses followed a fixed sequence of "improvements," giving the impression of learning while being entirely pre-programmed. Yadollahi et al. [295] used a similar approach in a reading task: the robot's errors and gestures (pointing at or looking at the text) were fully scripted, allowing the researchers to test how specific deictic behaviours shaped children's comprehension. Likewise, in Lindberg et al. [169], a NAO robot acted as a teachable agent in a math task, with all mistakes and progression predetermined. While this setup initially captured children's attention, engagement often declined as the session progressed, and children differed in how they perceived the robot—some treating it as a genuine learner, others as a computer with unlimited knowledge. These findings suggest that while scripted behaviours can elicit teaching, they may struggle to sustain engagement over time, especially when children question the robot's authenticity as a learner.

In contrast, learning-based models allow robots to adapt dynamically to the child's input rather than following pre-scripted behaviours. A landmark example is Lemaignan et al. [158], where a NAO robot learned to write letters by updating its internal representation using a Principal Component Analysis (PCA) model of handwriting. As children corrected its strokes, the robot adjusted its motor commands and visibly improved over time. This approach was later extended in follow-up studies by El Hamamsy et al. [82] and Chandra et al. [45], which explored collaborative settings and domain-specific applications, such as school classrooms or psychologist-led sessions. Beyond handwriting, Chen et al. [49] developed a learning-based model where children taught a robot to recognise stages of cell division through image classification, leading to greater conceptual gains and higher motivation. Such systems generate more authentic tutee behaviours and elicit stronger teaching responses from children. Yet they also highlight a core limitation: they are usually tailored to a single domain. Lemaignan's handwriting model, for instance, cannot be directly transferred to grammar instruction or visual reasoning without redesigning the entire learning pipeline. As a result, the current landscape of learning-based systems remains powerful but fragmented, with each model tied closely to its specific task.

Despite these advances, existing approaches still fall short of emulating a realistic learning trajectory. What is needed are models that move beyond narrow, task-specific adaptations to represent internal learning states and show genuine improvement through interaction with children.

2.4 Teachable Robots

The previous sections highlighted the importance of designing robot peers that can emulate the developmental path of a child learner: starting from novice performance, showing progress, and supporting reciprocal teaching dynamics. A promising route toward this goal is the notion of teachable robots: agents explicitly designed to learn from human input, whether through feedback, demonstration, or evaluative guidance. This section reviews computational approaches developed to support such robots, focusing on how human-guided learning frameworks, originally applied in domains such as manipulation and control, can be adapted to create social robots that act as credible tutees in educational contexts.

2.4.1 Technical Terminology

Teachable robots are increasingly implemented using Reinforcement Learning (RL), a framework well-suited to interactive, adaptive behavior. In RL, an agent learns to make decisions through feedback and rewards [262]. To clarify the terminology used in this thesis, the following section defines the foundational concepts.

Markov Decision Process (MDP)

A Markov Decision Process provides the formal foundation for modelling sequential decision-making problems under uncertainty [262]. It is defined by the tuple $(\mathcal{S}, \mathcal{A}, P, R, \gamma)$ where:

- $\mathcal{S}$ is the set of all possible *states* of the agent and environment,
- $\mathcal{A}$ is the set of *actions* available to the agent,
- $P(s' \mid s, a)$ is the *transition probability* of moving to state s' after taking action a in state s,
- $R(s, a)$ is the *reward function*, providing the environmental feedback upon taking action a in state s,
- $\gamma \in [0, 1]$ is the *discount factor*, that determines the sensitivity of the agent to future decisions.

At each time step t, the agent observes state s_t, selects action a_t, receives reward r_{t+1}, and transitions to state s_{t+1}. The objective is to learn a *policy* $\pi : \mathcal{S} \to \mathcal{A}$, or more generally $\pi(a \mid s)$, which defines the probability of taking action a in state s, such that the expected *discounted cumulative reward* is maximised:

$$\pi^* = \arg\max_\pi \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \right].$$

Two common functions are used to evaluate a given policy π:

- The *state-value function*, $V^\pi(s)$, defines the expected return when starting in state s and following policy π:

$$V^\pi(s) = \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s \right].$$

- The *action-value function*, or *Q-value*, $Q^\pi(s,a)$, defines the expected return when taking action a in state s and then following policy π:

$$Q^\pi(s,a) = \mathbb{E}_\pi \left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid s_t = s, a_t = a \right].$$

Algorithms

The methods presented in this thesis are built upon two fundamental RL algorithms: *Q-learning* and *Actor–Critic*. Both aim to find optimal policies, but differ in how they estimate value and update behaviour.

Q-learning. Q-learning [284] is a model-free, off-policy algorithm that learns the optimal action-value function $Q^*(s,a)$ by iteratively updating Q-values using the Bellman optimality equation:

$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right],$$

where α is the learning rate, r is the received reward, and γ is the discount factor. The optimal policy π^* is derived by selecting the action with the highest Q-value in each state:

$$\pi^*(s) = \arg\max_a Q^*(s,a).$$

Q-learning is simple to implement and well-suited for discrete state–action spaces.

Actor–Critic. Actor–Critic is a hybrid approach combining the advantages of value-based and policy-based methods [262]. The *actor* is responsible for selecting actions via a policy $\pi_\theta(a \mid s)$, while the *critic* estimates the value function $V_w(s)$ (or action-value $Q_w(s,a)$) to evaluate the actor's performance. The policy parameters θ are updated by ascending the gradient of expected reward:

$$\nabla_\theta J(\theta) = \mathbb{E}_{s\sim d^\pi, a\sim\pi_\theta}\left[\nabla_\theta \log \pi_\theta(a \mid s)\, A(s,a)\right],$$

where $A(s,a)$ is the advantage function, representing how much better an action is compared to the baseline. The critic provides this advantage estimate using temporal-difference (TD) learning:

$$\delta_t = r + \gamma V_w(s') - V_w(s).$$

Actor–Critic algorithms are especially useful in continuous or high-dimensional action spaces due to their efficient policy updates and stability in learning.

2.4.2 Learning from Humans

Interactive reinforcement learning allows agents to learn from human guidance through various teaching modalities. Chetouani et al. [50] identify four principal forms of human input: *demonstration*, *instruction*, *evaluative feedback*, and *preference feedback*. Each modality differs in the nature of the signal, integration mechanism, and computational demands.

Demonstration. Human demonstrations are typically represented as a sequence of state–action pairs: $D = \{s_t, a_t^*\}$, where a_t^* is the optimal action performed by the human at time t given the state s_t [11]. These demonstrations are collected via three general methods:

- **Kinesthetic teaching:** The human physically guides the robot by manipulating its joints to demonstrate the desired behavior.

- **Teleoperation:** The human remotely controls the robot using external devices such as joysticks or motion capture systems.

- **Observation:** The robot passively observes a human performing a task, typically through vision or sensor data, without explicit control over the robot's actions.

Each modality has trade-offs. Kinesthetic teaching allows precise control and is intuitive but requires physical effort and suitable hardware [221]. Teleoperation provides remote flexibility but may introduce latency, reduce demonstration fidelity, and require specific interfaces [221]. Observation scales well but often requires complex inference to translate observed behavior into actionable robot policies [51]. Learning from these demonstrations can proceed via two main approaches:

- **Behavioral cloning:** This method uses supervised learning to imitate expert behavior by training a policy $\pi_\theta(a \mid s)$ that maps states to actions [15]. The learning objective is to minimise the discrepancy between the expert's demonstrated actions a_t^* and the actions predicted by the policy for the same states. This is done by minimising a loss function:

$$\mathcal{L}(\theta) = \sum_{(s_t, a_t^*) \in D} \|a_t^* - \pi_\theta(s_t)\|^2$$

- **Inverse Reinforcement Learning (IRL):** Instead of directly learning a policy, IRL seeks to infer a reward function $R(s, a)$ that makes the expert's behavior appear near-optimal [192]. A simple formulation ensures that the expert policy π^* has higher value under the recovered reward than any other policy π:

$$E_{\pi^*}\left[\sum_t \gamma^t R(s_t, a_t)\right] \geq E_\pi\left[\sum_t \gamma^t R(s_t, a_t)\right] \quad \forall \pi.$$

  Once R is estimated, standard reinforcement learning can be used to derive a policy consistent with the expert's behavior.

Instruction. Human instructions are typically represented as a probability distribution over the action space, expressed as:$\Pr_t(a \mid i)$, where $a \in \mathcal{A}$ and i denotes the instruction issued at time t [50].

Two common modes of instruction are often distinguished: *action advice* and *verbal critique* [152]. In action advice, the agent receives explicit guidance such as *"move right"* or *"go left"*, which directly suggests an optimal action. In contrast, critique-based instruction involves scalar feedback, typically in the form of binary evaluative phrases like *"good job"* or *"don't do that"*.

Learning from instructions is generally formulated as the task of mapping human-issued command, often in natural language, into executable action sequences [174]. This process assumes some degree of mutual understanding between the human and the agent regarding the semantics of the instruction, which is often engineered through pre-defined mappings or intermediate symbolic representations.

However, some challenges might arise. Pre-defined mappings between language and actions require extensive engineering and often lack flexibility across domains [189]. Recent advances in Large Language Models (LLMs) offer more generalisable alternatives, yet they must be carefully grounded in the environment's actual state and action space to prevent hallucinated or invalid commands. In small or discrete environments, LLMs may also introduce unnecessary latency and computational overhead, making simpler, domain-specific models a more practical choice.

Evaluative Feedback. Evaluative feedback refers to scalar signals provided by a human teacher to indicate approval or disapproval of an agent's action. At each time step, the human observes the state transition (s_t, a_t, s_{t+1}) and delivers a feedback value $H(s_t, a_t)$ that reflects their judgment of the agent's behaviour [50]. Unlike the environment reward $r(s, a)$, which is predefined by the task designer, evaluative feedback originates from the human teacher. It can either complement the reward signal as an additional source of guidance in sparse or delayed reward settings, or replace it entirely, allowing the agent to learn solely from human-provided evaluations within the MDP framework. Teaching with evaluative feedback is considered a low-cost strategy: the agent explores autonomously, and the human intervenes only when necessary to provide evaluations. However, constant intervention is cognitively demanding, and most practical implementations must account for sparse and delayed human input.

A common approach to integrating human evaluations is through value shaping [189]. The human feedback is treated as a proxy for a latent value function $\hat{H}(s, a)$, which is used to augment the agent's Q-function:

$$Q'(s, a) = Q(s, a) + \beta \cdot \hat{H}(s, a)$$

where β is a decaying weight factor that controls the influence of the human input.

Preference Feedback. In preference-based reinforcement learning (PbRL), human teachers compare pairs of action trajectories, providing feedback in the form of relative preferences rather than numerical rewards [58]. Formally, given two trajectories τ_i, τ_j, the teacher indicates which they prefer, denoted $\tau_i \succ \tau_j$. The goal is to infer a latent reward function R^{ψ} that explains these preferences, assuming that the preferred trajectory has a higher cumulative reward :

$$P(\tau_i \succ \tau_j) = \frac{\exp\Big(R^{\psi}(\tau_i)\Big)}{\exp\Big(R^{\psi}(\tau_i)\Big) + \exp\Big(R^{\psi}(\tau_j)\Big)}.$$

Preference-based methods avoid the need for hand-crafted numeric rewards and are well-suited for non-expert users. Once the reward model is learned, standard RL (e.g. policy optimization) can be applied using the inferred rewards. However, PbRL relies heavily on the quality of relative comparisons, and may require many trajectory pairs to converge unless active or meta-learning strategies are used [289, 26].

2.5 Summary

This chapter traced how theories of development, research in social robotics, and empirical work on learning-by-teaching converge toward the idea that robots can function not merely as instructional tools but as credible peers and learners. Prior studies have demonstrated the value of scaffolding, fallibility, and peer-like roles, yet most computational models remain narrow: they either rely on scripted behaviours, Wizard-of-Oz control, or domain-specific adaptations. These approaches fall short of capturing the incremental, contingent, and socially meaningful dynamics that make peer-to-peer teaching effective.

What remains underexplored is how robots can embody genuine learning trajectories—making errors, showing progress, and adapting in response to children's feedback—in a way that reflects the mechanisms established in developmental theory. Addressing this gap requires models that do not simply simulate teaching interactions but that can learn interactively from children in ways that are transparent, responsive, and pedagogically aligned.

This thesis takes up that challenge. It develops and evaluates an interactive reinforcement learning model that enables social robots to act as teachable peers in primary school contexts. By grounding the design in developmental principles, implementing computational models that allow robots to adapt through child feedback, and testing these systems in classroom studies, the work aims to demonstrate how teachable robots can support both children's cognitive gains and their social-emotional engagement. The next chapter introduces this model and the methodological framework used to investigate it in real educational settings.

Chapter 3

A Cognitive Model for Peer-Like Learning

[Epigraph removed for copyright reasons. Original source: Yoda, Star Wars: The Last Jedi*, dir. Rian Johnson (Lucasfilm, 2017)]*

3.1 Introduction

The previous chapter highlighted the importance of peer-like interaction, transparency, and responsiveness in supporting learning-by-teaching (LbT) scenarios with social robots. Building on this foundation, this chapter presents a cognitive model that brings these principles into practice through an interactive learning architecture guided by user feedback. The goal is to enable the robot to behave as a teachable peer: starting as a novice, adapting in real time to human input, and making its learning trajectory visible and contingent.

To inform the design of this model, the chapter first reports an online study comparing how different robot learning approaches, varying in contingency and interactivity, are perceived by users. Based on these findings, an Interactive Reinforcement Learning (RL) framework is introduced, allowing the robot to incrementally refine its policy through binary evaluative feedback. This framework supports responsive and transparent learning behaviour and provides the foundation for its evaluation in classroom studies presented in the following chapters.

3.2 Online Study on Perceived Contingency and Responsiveness

3.2.1 Study Motivation and Research Questions

For a robot to function effectively as a teachable peer, it should not only learn competently but also be perceived as a responsive and contingent learner. As discussed in Chapter 2, prior work in HRI has shown the pedagogical potential of novice-like robots [124, 24], but much of this research has relied on scripted or Wizard-of-Oz behaviours. Some learning-based peer robots can adapt to children's teaching [158], but these models are typically engineered for specific tasks and environments, making them difficult to transfer to novel settings without substantial redesign.

Teachable robots are promising candidates for peer-like learning because they allow children to guide a robot's development. However, not all human-guided strategies are equally suited to classroom settings. Demonstration assumes that users can provide optimal trajectories, which is unrealistic for children and introduces a high interactional burden [50]. Instruction via natural language is more intuitive but remains constrained by speech recognition challenges, particularly in noisy classrooms and with child speakers [64, 198, 255, 89].

Feedback-based modalities, by contrast, are simpler and more accessible. Evaluative feedback, involving binary judgments, and preference-based feedback, which compares alternatives, have both shown technical feasibility [142, 171, 58]. Yet their effectiveness in child–robot interaction depends not only on algorithmic performance but also on how users perceive responsiveness, clarity, and ease of teaching. These aspects remain underexplored. Although this concern partly overlaps with explainable AI (XAI), which seeks to make model decisions transparent [4], the focus here is different. Rather than exposing internal computations, the goal is to ensure *interactional transparency*: that users clearly perceive the causal link between their feedback and the robot's behaviour. This quality is crucial for designing teachable robots that are both adaptive and socially credible.

To examine these issues, an online study was conducted to compare how different robot learning models, varying in their contingency and interactivity, are perceived by users. The study further investigates how these perceptions are shaped by user characteristics and task type. These insights aim to inform the selection of learning models that best support engagement, trust, and instructional agency in child–robot LbT interactions.

Three research questions guide this investigation:

- **RQ1.1.** How do different robot teaching methods (evaluative feedback, preference-based feedback, and non-interactive learning) shape users' perceptions of robot responsiveness, controllability, and teaching success?
- **RQ1.2.** To what extent do individual user characteristics (e.g., personality traits, prior experience with robots) influence preferences for specific teaching methods?
- **RQ1.3** How does the nature of the learning task (e.g., navigation, manipulation, classification) affect the perceived agency and anthropomorphism of the robot?

3.2.2 Experimental Design

An online between-subjects study was conducted to explore how users perceive different robot learning behaviours across various teaching modalities and task types. Rather than directly interacting with the robot, participants watched short video clips depicting the robot being taught by a human in different learning scenarios. The study followed a 3×3 factorial design, crossing three robot teaching methods with three pedagogical tasks. Each participant was randomly assigned to one of the nine experimental conditions and viewed corresponding videos before completing the questionnaire. The specific video clips can be viewed on the project website [1].

[1]https://sites.google.com/view/teachable-robots

Participants An a priori power analysis was conducted using G*Power (version 3.1) for a MANOVA (α = .05, power = .95, number of groups = 9), assuming a small effect size ($f^2(V) = 0.0625$), which indicated a recommended sample size of $N = 144$. A total of 138 participants ($M_{age} = 22.87$, $SD = 1.78$; 80 men, 56 women, 1 non-binary) were recruited via the online platform Prolific [2]. All were between 18 and 26 years old and received monetary compensation for their participation (£1.80). Although the broader aim of this research is to support child–robot interaction, this preliminary study used young adults as a proxy population due to practical constraints, with the goal of informing the design of robot behaviours for future use in classrooms. The study received ethical approval from the Sheffield Hallam University Research Ethics Committee (Application ID: ER56422859).

Teaching Conditions. Participants were randomly assigned to one of three robot learning conditions, as depicted in Figure 3.1:

- **Evaluative Feedback:** Participants observed the robot improving through trial-and-error based on binary feedback provided by a human teacher using a joystick.
- **Preference-Based Feedback:** Participants saw the experimenter select preferred robot behaviours via a web interface, with preferences then transferred to guide the robot's learning.
- **Download Condition:** In this control condition, participants watched the robot executing a pre-trained policy without user intervention.

Task Types. To examine how perceptions might vary by context, participants were also assigned to one of three task types, as illustrated in Figure 3.2. These tasks varied in complexity and social framing, enabling assessment of how the nature of the task influences perceptions of robot behaviour.

- **Navigation Task:** The robot had to navigate through a maze to reach a goal while avoiding obstacles.

[2] https://www.prolific.com/

Figure 3.1: Overview of the teaching condition. From left to right: Teaching with Evaluative Feedback, teaching with Preferences and Download condition.

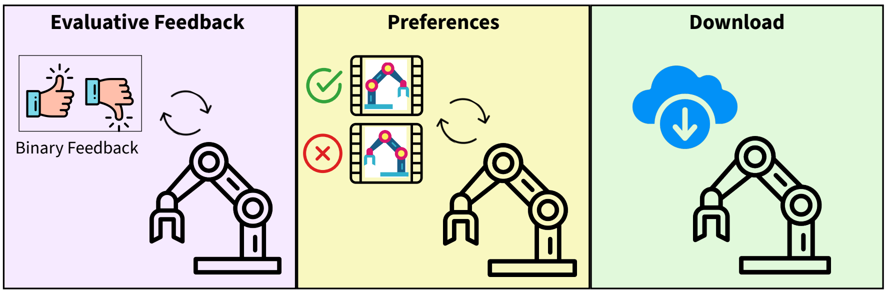
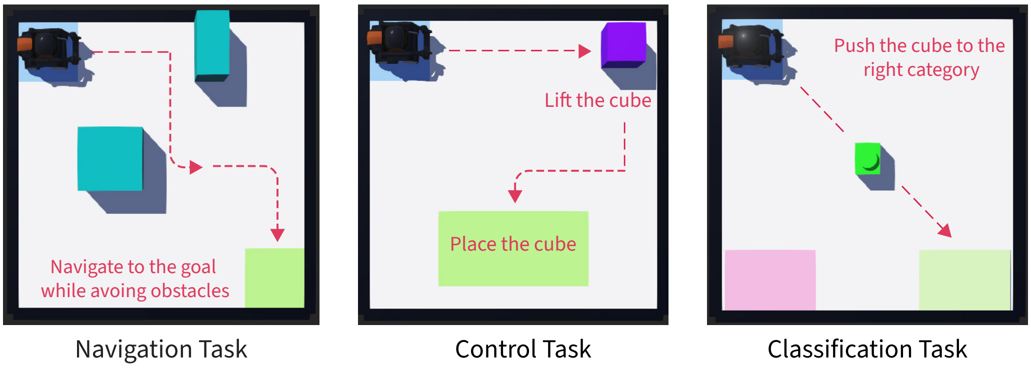


Figure 3.2: Experimental tasks: Participants observe the teaching process of a robot within one of three types of tasks: navigation, control and classification.



- **Control Task:** This task required the robot to approach, lift, and place a cube in a designated location, involving both navigation and manipulation.
- **Classification Task:** The robot was tasked with categorising objects into two predefined classes. It could choose to assign an object to category A, category B, or skip classification.

3.2.3 Measures

To assess participants' perceptions of the robot and the teaching method, a range of standardised and adapted instruments were used. All items were presented as part of an online questionnaire administered after the testing phase. Unless otherwise specified, responses were collected using 5-point Likert scales (1 = Strongly Disagree, 5 = Strongly Agree). Internal consistency for all multi-item scales was assessed using Cronbach's alpha, with detailed reliability statistics reported in Appendix A.

Robot Perception Metrics

- **Perceived Responsiveness:** Assessed using two adapted items from the Godspeed Animacy scale [18], measuring the extent to which the robot appeared to react contingently to the teaching process.
- **Perceived Control:** Measured with a 3-item scale adapted from [71], capturing participants' sense of influence over the robot's behaviour under the given teaching method.
- **Anthropomorphism, Intelligence, and Likability:** Evaluated using adapted subscales from the Godspeed questionnaire [18], each consisting of five items.
- **Perceived Success:** Captured using a binary question asking participants whether they believed the robot had learnt successfully.

Teaching Method Evaluation

- **Ease of Use and Acceptance:** Participants rated the teaching method's ease of use, usefulness, and their intention to use it in the future. These were measured using eight items adapted from the Technology Acceptance Model (TAM) [23].

Individual Differences

- **Personality Traits:** The Mini-IPIP inventory was used to measure the Big Five personality traits: Neuroticism, Extraversion, Openness (Intellect), Agreeableness, and Conscientiousness [74]. The scale includes 20 items (four per trait).
- **Robotics Experience:** A 5-item scale adapted from [185] assessed participants' prior experience and familiarity with robots.

3.2.4 Procedure

The study was conducted online using the Qualtrics platform [3] and followed a structured sequence comprising four main phases:

1. **Pre-survey:** Participants first completed a pre-survey that collected demographic information (age, gender, education level), prior experience with robots, and personality traits. This phase ensured baseline comparability across groups and enabled subsequent analysis of individual differences.

2. **Training Phase:** Participants were then randomly assigned to one of the nine experimental conditions (3 teaching methods × 3 task types). Each participant watched a short training video illustrating the robot's learning process within their assigned conditions. To control for outcome-based bias, all training videos were designed to conclude with equivalent robot performance, regardless of the learning condition.

3. **Testing Phase:** After training, participants viewed a set of four video clips showing the robot performing the task in new test scenarios. Each set included: two successful trials, one ambiguous trial (suboptimal but goal-reaching), and one failure. The 75% success rate was selected to reflect realistic performance dynamics, in line with prior HRI work [296]. The test videos were identical across teaching conditions within each task type, to ensure comparability.

4. **Post-survey:** Finally, participants completed a questionnaire assessing their perceptions of the robot and the teaching method, using the measures described in Section 3.2.3. An attention check was included between the testing phase and the questionnaire to ensure data reliability.

All video materials were pre-recorded using a WoZ approach to maintain consistent robot behaviour across participants and conditions. The EF condition involved real robot interaction, while the PB condition used preference-trained trajectories in sim-

[3] https://www.qualtrics.com/en-gb/

Table 3.1: Descriptive statistics (Mean and SD) for each variable across teaching conditions.

Variable	Download Condition	Evaluative Feedback	Preferences
Anthropomorphism	0.40 (0.15)	0.45 (0.18)	0.47 (0.14)
Control	0.71 (0.21)	0.79 (0.11)	0.77 (0.16)
Responsiveness	**0.73 (0.16)**	**0.82 (0.13)**	0.77 (0.17)
Intelligence	0.58 (0.17)	0.64 (0.17)	0.65 (0.15)
Likability	0.69 (0.11)	0.72 (0.19)	0.74 (0.15)
Ease of use	**0.74 (0.13)**	**0.80 (0.17)**	0.80 (0.16)
Usefulness	0.71 (0.11)	0.74 (0.17)	0.74 (0.17)
Intent of use	0.75 (0.14)	0.80 (0.14)	0.78 (0.15)

ulation. The Download condition followed a fully scripted execution. Experimental videos can be found in the project website [4].

3.2.5 Results

This section presents the findings from the quantitative analyses of the online study. Multivariate and non-parametric statistical tests were used depending on normality and variance assumptions.

Effects of Teaching Method on Perception

A multivariate analysis of variance (MANOVA) revealed a significant overall effect of teaching method on participants' perception of the robot, specifically on perceived *responsiveness*, *control*, and *ease of use* (Pillai's Trace = 0.11, $F(8, 266) = 1.99$, $p =$ .048). Table 3.1 presents the descriptive statistics (means and standard deviations) for each variable across conditions, and Figure 3.3a depicts the significant differences between the teaching conditions.

Responsiveness. A Kruskal-Wallis test showed a significant difference across teaching conditions ($H(2) = 6.46$, $p = .040$). Post-hoc Wilcoxon rank-sum tests (Bonferroni-corrected) indicated that robots taught through *evaluative feedback* were perceived as significantly more responsive ($M = 8.14$, $SD = 1.33$) than those

[4] https://sites.google.com/view/teachable-robots

in the *Download* condition ($M = 7.33$, $SD = 1.58$; $p = .014$, $d = 0.56$). No significant differences were found between the Evaluative Feedback and Preference-Based conditions.

Ease of Use. A significant main effect was also found for perceived ease of use ($H(2) = 6.37$, $p = .041$). Both interactive methods (Evaluative Feedback and Preference-Based) were rated as more user-friendly than the Download condition, although post-hoc comparisons did not reach significance after Bonferroni correction.

Other Measures. The remaining metrics showed no statistically significant differences across teaching conditions. However, descriptive trends suggested that:

- **Intent to use** and **usefulness** were consistently high across all conditions.
- **Perceived control**, **intelligence**, and **likability** of the robot were rated notably above average.
- **Anthropomorphism** was rated as relatively average.

These findings suggest that while users could differentiate between robot learning modalities in terms of responsiveness and usability, general attitudes toward the robot's usefulness, likability, and overall competence were stable across conditions.

Effects of User Characteristics

Individual differences in personality traits and prior experience with robots influenced how participants evaluated the robot and the teaching method.

Extraversion. In the Evaluative Feedback condition, extraversion was positively correlated with perceived responsiveness ($r = .42$, $p = .003$), intelligence ($r = .46$, $p = .001$), and likability ($r = .36$, $p = .015$). No significant correlations were observed in the other teaching conditions.

Intellect (Openness). Intellect was positively associated with ease of use in the Preference-Based condition ($r = .42$, $p = .002$), and with both perceived responsiveness ($r = .36$, $p = .013$) and likability ($r = .35$, $p = .015$) in the Evaluative Feedback condition. No significant correlations were found in the Download condition.

Robotics Experience. In the Download condition, prior robotics experience correlated positively with perceived usefulness ($r = .37$) and anthropomorphism ($r = .38$), both at $p = .013$. No significant correlations were observed in the other conditions.

No significant associations were found for gender, education level, or other personality traits.

Effects of Task Type

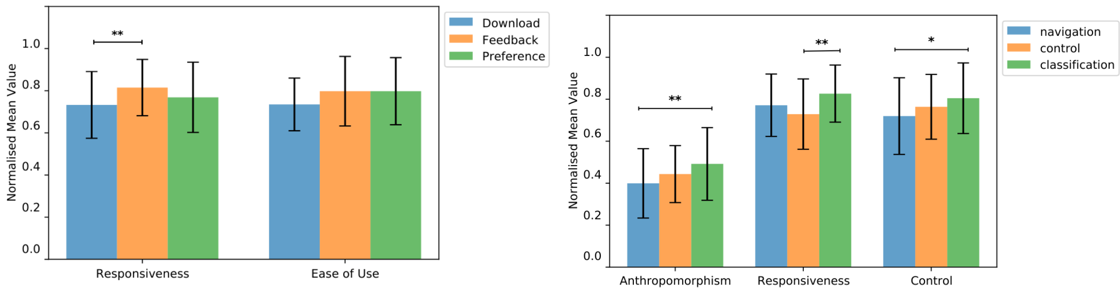
A separate MANOVA was conducted to examine whether the type of task influenced participants' perceptions, independent of teaching condition. A significant multivariate effect was found (Pillai's Trace = 0.14, $F(6, 268) = 3.41$, $p = .003$). Figure 3.3b illustrates the significant differences between the tasks.

Anthropomorphism. A Kruskal-Wallis test revealed a significant difference in perceived anthropomorphism ($H(2) = 7.39$, $p = .025$). Wilcoxon comparisons showed significantly higher ratings in the *classification* task ($M = 12.29$, $SD = 4.32$) than in *navigation* ($M = 9.98$, $SD = 4.13$; $p = .009$, $d = 0.55$). No difference was found with the control task.

Control The nature of the task also affected perceived control over the robot ($H(2) = 6.15$, $p = .046$). Control ratings were higher for the *classification* task ($M = 12.07$, $SD = 2.52$) compared to *navigation* ($M = 10.8$, $SD = 2.74$; $p = .02$, $d = 0.48$).

Responsiveness Participants rated the robot as more responsive in the *classification* task ($M = 8.27$, $SD = 1.36$) than in the *control* task ($M = 7.29$, $SD = 1.68$;

Figure 3.3: (a) Comparison of the responsiveness and ease of use between the teaching conditions. (b) Comparison of the anthropomorphism, control and responsiveness between the tasks.



(a) **(b)**

$p = .005$, $d = 0.63$). No significant difference was observed between classification and navigation.

No significant differences were found across tasks for perceived intelligence or likability.

3.2.6 Discussion

The findings of this study offer different insights into how different teaching modalities affect user perceptions of teachable robots, and how individual characteristics and task type further shape those perceptions.

In line with RQ1.1, the teaching method significantly influenced perceived responsiveness and ease of use. Robots that learned through evaluative feedback were seen as more responsive than those following a pre-trained policy in the Download condition. This supports the idea that visible, contingent robot learning plays a critical role in establishing the robot as a socially credible learner. Interestingly, preference-based learning, while technically interactive, did not significantly enhance perceived responsiveness. This may be due to the more indirect and less interpretable nature of preference-based learning, where the cause-effect relationship between user input and robot behaviour is less immediate and transparent. Despite these differences in responsiveness, both interactive methods were rated as more user-friendly than the non-interactive condition. This suggests that even minimal human involvement in the learning process may increase users' engagement and acceptance of the teaching role. However, it is noteworthy that no significant differences were observed across conditions for perceived anthropomorphism, intelligence, likability, usefulness, or intent to use. These attributes may require longer or more immersive interactions to shift, and thus were less influenced by the brief exposure provided in this study.

With respect to RQ1.2, personality traits, particularly extraversion and intellect (openness), emerged as important predictors of user experience. Extraverted participants responded more positively to the robot in the Evaluative Feedback condition, suggesting that socially outgoing individuals may be more comfortable in dynamic, trial-and-error teaching interactions. Similarly, participants who scored high on the Intellect/Openness dimension perceived both interactive conditions as more usable and engaging, possibly due to a greater openness to novelty and tolerance for ambiguity. These findings underscore the importance of considering user traits when designing robot teaching interfaces, especially in educational contexts where individual learners may vary in confidence, motivation, and communication style.

In relation to RQ1.3, task type significantly influenced how the robot was perceived. Across conditions, the classification task elicited higher ratings of anthropomorphism, control, and responsiveness than more mechanical tasks like navigation. This effect may stem from the classification task's closer resemblance to an educational game, which likely made the interaction feel more social, goal-oriented, and pedagogically meaningful, thereby encouraging users to attribute more human-like qualities to the robot and feel a greater sense of agency.

Taken together, these findings support the conclusion that evaluative feedback is a promising teaching modality for educational HRI. It supports contingency, is perceived as responsive, and was experienced as intuitive and engaging by participants. Although preference-based learning can guide robot policy successfully, participants perceived it as less transparent or responsive when compared to more direct forms of interaction.

While not conclusive, the results suggest that evaluative feedback may offer a more interpretable and engaging interaction style, making it a suitable choice for the interactive architecture described in the next section. The findings also point to the importance of adapting robot behaviour to individual user traits and task demands.

3.3 Cognitive Model for Peer-Learning Robots

The focus of the thesis is to build a cognitive model for robots that would allow them to act as peers. The design of this cognitive model is grounded in three principles established in Chapter 2. First, peer-to-peer learning requires a partner who can collaborate interactively rather than act as a pre-programmed tutor. Second, learning-by-teaching scenarios are most effective when the robot starts as a novice and demonstrates visible improvement across the interaction, thereby sustaining the child's motivation to teach [124, 90]. Third, positioning children as teachers requires methods of guidance that are appropriate for their cognitive and interactional capacities: they must be easy to use, transparent in how they influence the robot, and not impose excessive cognitive load [50, 64].

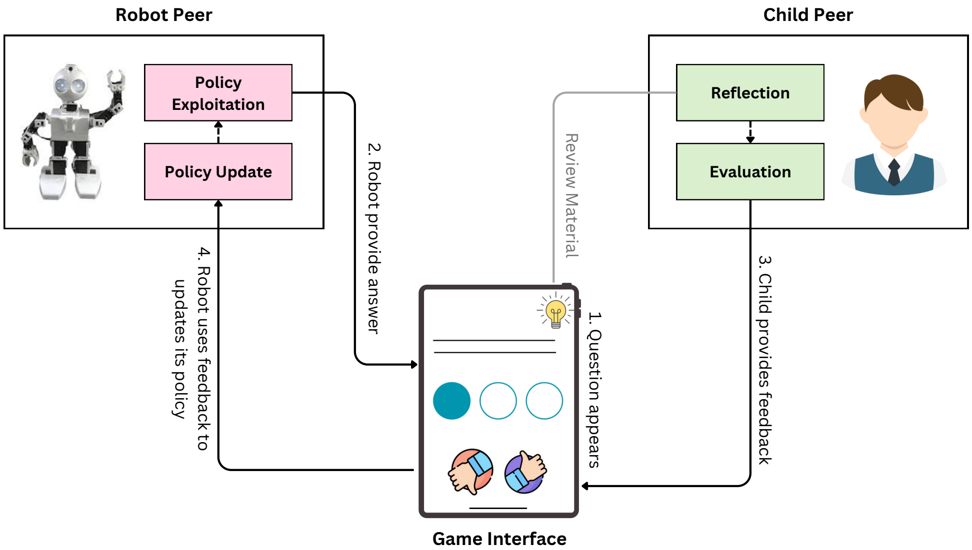
Among the different teaching modalities, evaluative feedback emerges as a particularly suitable candidate for enabling teachable robots in classroom settings. It requires only simple binary judgments (correct/incorrect), closely mirrors how children naturally correct one another [52], and makes the robot's learning progress transparent [9]. Compared to demonstrations, which demand the ability to provide optimal trajectories, or natural language instructions, which are limited by speech recognition in noisy classrooms, evaluative feedback places minimal burden on young users while still supporting effective peer-like interaction [50, 171]

The online study presented earlier provided additional support for this choice: participants perceived robots trained with evaluative feedback as more responsive and intuitive than those following a pre-trained policy. While not conclusive, these findings validate the broader rationale for selecting evaluative feedback as the foundation of the model.

Building on these considerations, the proposed cognitive model enables the robot to begin with no prior knowledge and incrementally refine its behaviour through repeated interaction with a child tutor. By grounding learning in *Interactive*

Reinforcement Learning (RL), the model ensures that each feedback signal directly shapes the robot's policy, making the learning trajectory both contingent and visible to the child.

Figure 3.4: Overview of the cognitive model for Learning-by-Teaching with Interactive Reinforcement Learning. (1) A question is displayed with multiple choice options. (2) Policy Exploitation: The robot peer uses its current policy to select an answer. (3) The child peer reflects on the robot's answer (Reflection), optionally reviews the material, and decides whether the robot's response is correct or incorrect (Evaluation). (4) The robot updates its policy (Policy Update) based on the child's feedback. This loop continues as the robot progressively refines its knowledge, while the child simultaneously reinforces their own understanding through active teaching.



3.3.1 Interactive Reinforcement Learning Peer Model

Interactive RL enables agents to adapt their behaviour directly from human feedback, reducing reliance on predefined reward functions and making learning accessible to non-expert users. This property is particularly relevant in educational contexts, where children act as teachers and must be able to guide a robot's behaviour without technical expertise. A number of frameworks have been proposed in this space, demonstrating the feasibility of shaping agent behaviour through human-delivered signals. For example, Knox and Stone [142] introduced the TAMER framework, in which binary evaluative feedback is interpreted as a reward signal to guide policy updates. Loftin et al. [171] extended this approach by modelling human feedback strategies to accelerate convergence, while MacGlashan et al. [176] showed that human feedback is policy-dependent and introduced COACH (Convergent Actor–Critic by Humans), an algorithm that interprets human critique as an advantage signal within an actor–critic model. More recent work has extended Interactive RL into deeper and more expressive architectures. For instance, the Deep TAMER framework applies neural network-based function approximation to TAMER, enabling real-time human feedback to be effective in complex, high-dimensional environments such as Atari games [282, 10]. Emerging studies also explore emotion-based feedback as a signal for learning. Researchers have demonstrated that human facial expressions, captured during training sessions, can be interpreted as evaluative feedback to guide a robot's learning effectively [163], and that emotional cues may serve as intuitive reinforcement signals in interactive drone training scenarios [214].

In the context of education, however, certain constraints narrow the set of viable options. A teachable robot must adapt quickly within short interactions, provide visible improvements that reinforce the child's role as a teacher, and minimise the cognitive demands placed on the child. From this perspective, TAMER offers several advantages. It is straightforward to implement, requires no prior specification of task-specific reward functions, and supports rapid convergence within discrete environments. Crucially, the direct mapping between a child's binary judgment (correct/incorrect) and the robot's policy update is intended to make the learning process transparent and contingent, qualities that are important for sustaining engagement in LbT scenarios.

For these reasons, the TAMER framework was adopted as the basis of the peer-learning model presented here. Although not a new algorithm, its properties align closely with the requirements of child–robot educational interaction: ease of use, responsiveness, and pedagogical credibility. Specifically, after each action, the robot receives an evaluation signal from the user (+1 for a correct answer, -1 for an incorrect one) and updates its Q-values using the following rule:

$$Q'(s,a) = Q(s,a) + \alpha \cdot (h - Q(s,a)), \tag{3.1}$$

where $Q(s, a)$ is the current value estimate for state–action pair (s, a), h is the feedback signal, and α is the learning rate. This allows the robot to incrementally refine its policy $\pi(a \mid s)$ over time.

While TAMER itself is not a new algorithm, its application here represents a novel adaptation to the educational domain. Most prior implementations of TAMER and related frameworks have focused on technical benchmarks or constrained laboratory tasks, without addressing the pedagogical requirements of child–robot interaction. The contribution of this work lies in reframing Interactive RL as a cognitive model for peer-like learning: making the robot's errors and improvements visible, ensuring that feedback can be provided through child-appropriate modalities, and embedding the model within a flexible game interface that can be easily customised by teachers across domains. In this sense, the originality of the model is not in proposing a new algorithmic mechanism, but in aligning an existing Interactive RL framework with developmental and educational principles to support learning-by-teaching in real classroom contexts.

3.3.2 Game Interface and Policy Formulation

The proposed cognitive model requires a formulation that allows the robot to update its behaviour incrementally in response to feedback. To this end, the learning task is modelled as a Markov Decision Process (MDP), where each quiz question corresponds to a state $S = \{question_1, \ldots, question_n\}$ and the set of possible answers defines the action space $A = \{option_1, \ldots, option_m\}$. The robot's policy $\pi(a|s)$ maps questions to answers and is continuously refined through feedback, ensuring that each child's input directly shapes the robot's behaviour. This formalisation allows the TAMER framework introduced in the previous section to be applied in an educational setting: the binary evaluative feedback provided by the child is interpreted as a reward signal, driving incremental policy improvement, until the optimal one is reached whereby the robot learns to map each question to the correct answer.

To situate this process in a context familiar and engaging for children, a game-based interface was designed. The robot assumes the role of a novice learner and the child acts as a tutor, mirroring peer-teaching dynamics while keeping the interaction lightweight. Each round presents a multiple-choice question on a touchscreen tablet, the robot selects an answer, and the child indicates whether the choice was correct. This framing highlights the child's agency as a teacher while making the robot's progress transparent across trials. The choice of a quiz format makes the interface broadly adaptable: any learning activity that can be structured as a multiple-choice question set, whether in language, mathematics, history, or science, can be integrated into the game. This contrasts with much prior work in learning-by-teaching, where tasks were narrowly engineered for specific domains such as handwriting [158] or scripted vocabulary games [82]. By contrast, the present interface generalises across subjects and can be customised by teachers without technical expertise.

The interface was implemented using Flask [5] and deployed on touchscreen tablets. It presents quiz items, collects feedback, and communicates with the robot's learning module. It was deliberately designed to be intuitive, child-friendly, and adaptable across different learning domains. Its modular structure also supports future extensions to other feedback modalities, thereby providing a flexible platform for deploying teachable robots in varied classroom scenarios.

3.4 Summary

This chapter presented the design rationale and technical formulation of a teachable robot model grounded in Interactive RL. The first part reported the findings of an online user study evaluating different robot learning behaviours. Results showed that evaluative feedback was perceived as more responsive and easier to use compared to non-contingent alternatives, and that individual traits such as extraversion and task framing significantly shaped user perceptions. These insights informed the development of a cognitive model in which the robot acts as a peer learner, improving its policy through real-time evaluative input.

The proposed model formalises the child–robot interaction as an MDP and uses the TAMER framework to update the robot's behaviour based on binary

[5]https://flask.palletsprojects.com/en/stable/

feedback. This design enables fast convergence, interpretability, and contingency; key qualities for effective LbT interactions. By aligning empirical evidence with system implementation, this chapter establishes the foundations for deploying the peer learner in real-world educational settings.

The next chapter evaluates this model in classroom experiments with primary school children, examining its impact on learning outcomes, engagement, and perceptions of the robot across different conditions.

Chapter 4

Peer-like Robots: Evaluating Interactive RL in Primary School Classrooms

"It is the time you have wasted for your rose that makes your rose so important."

— Antoine de Saint-Exupéry, *The Little Prince*

4.1 Introduction

Building on the cognitive model introduced in Chapter 3, this chapter examines how the Interactive RL-based model performs when deployed in real classroom settings. While most prior work on LbT with robots has relied on scripted or WoZ behaviours, this study investigates the feasibility and educational impact of an autonomous, feedback-driven robot that learns directly from children's evaluative input.

The chapter extends Interactive RL to real classrooms, testing its effectiveness beyond controlled lab settings. Two learning tasks were used to compare a robot-assisted learning-by-teaching condition with a tablet-based self-practice control. The goal is to understand not only whether children can effectively teach such a robot, but also how this interaction influences their own learning, engagement, and behaviour.

This chapter contributes new insights into the real-world use of adaptive peer-like robots in education and the conditions under which they may support personalised and reflective learning.

4.2 Study Motivation and Research Questions

This study investigates the practical impact of the proposed cognitive model for teachable robots when embedded in real classroom interactions. Instead of evaluating only the model's technical performance, the focus is on how teaching an autonomous robot influences children's learning, engagement, and self-regulation.

Specifically, the study explores whether children can effectively guide the robot's learning, whether this peer-teaching experience enhances retention compared to independent practice, and how these effects vary by learner profile and task type. A between-subjects design compared the robot-based LbT condition with a matched tablet-based control, using the Interactive RL framework introduced earlier to adapt the robot's behaviour based on children's feedback.

The study is guided by the following research questions:

- **RQ2.1: Can primary school children successfully guide a robot's learning using Interactive Reinforcement Learning in a classroom setting?**
  This question examines the feasibility of deploying Interactive RL in real educational environments, focusing on children's ability to provide accurate, timely feedback that supports the robot's learning.

- **RQ2.2: Does engaging in a teaching interaction with a robot enhance children's knowledge retention compared to independent practice?**
  Here, we explore whether the act of teaching a robot contributes to deeper learning and longer-term memory consolidation, in comparison to more traditional self-guided learning.

- **RQ2.3: Are children with lower initial knowledge more likely to benefit from the learning-by-teaching condition than their higher-**

achieving peers?
This question investigates the potential for teachable robots to act as a form of adaptive support, offering greater benefit to students who begin with less prior knowledge.

- **RQ2.4: How does the type of learning task influence the effectiveness of the robot-based learning-by-teaching interaction?**
This final question addresses whether the cognitive demands of different task types (e.g., factual recall vs. conceptual reasoning) affect the outcomes of teaching a robot, with implications for future task design.

4.3 Methodology

4.3.1 Participants

A total of 58 children aged 8–9 years (Year 4, UK) participated in the study. All attended the same primary school in the United Kingdom, were distributed across two classes, and studied French as an additional language. Parental consent was obtained in advance, and verbal assent was collected from the children on the day of the study. To protect participant privacy, data were pseudonymised using a scheme that combined each child's class number with the first letter of their teacher's name. This approach enabled matching retention scores with earlier data while maintaining confidentiality. One child was excluded from analysis for scoring zero on the pre-test, as the LbT approach targets knowledge consolidation rather than initial acquisition. Additional data were removed due to technical issues, logging errors, or mismatched pseudonym IDs. The final dataset included 53 children: 33 boys, 16 girls, and 4 who preferred not to specify their gender.

4.3.2 Learning Tasks

The learning tasks were presented as quiz-style games, implemented as a web-based interface using Flask and deployed on touchscreen tablets. Each question corresponded to a state in the underlying MDP, and each answer option mapped to a possible action, following the task formulation described in Section 3.3.2. The interface supported real-time interaction, functioning both as the input/output channel for the robot and as the medium through which children delivered feedback.

Two learning activities were selected in collaboration with primary school teachers to ensure pedagogical alignment. The chosen tasks focused on content not yet covered in class, with the aim of minimising ceiling effects while presenting an appropriate cognitive challenge. Prior work suggests that overly simple tasks risk disengagement [263], whereas more demanding activities can promote learning and reflection [264, 48]. Based on these criteria, two domains were targeted: vocabulary memorisation and grammar rule inference.

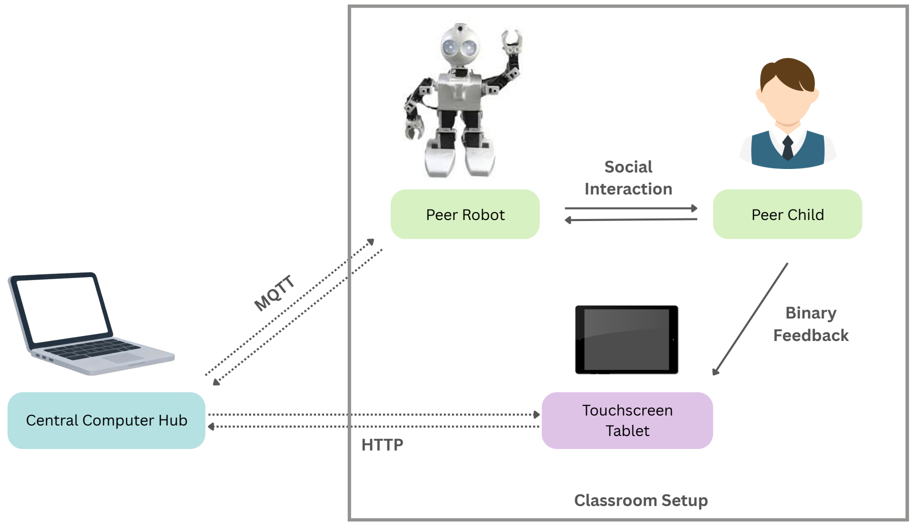
Body Parts Game This vocabulary task required the robot to match English body part names (*hand, head, foot, belly,* and *eye*) to their correct French translations. In each round, an English word was displayed with three translation options, and the robot selected one. The task drew on previous research showing the benefits of social robots in language learning, particularly when iconic gestures are used to ground word meanings [69]. To reinforce its answers, the robot pointed to the corresponding body part after each response. This task was modelled as an MDP with five states (target words) and three actions (possible translations).

Grammar Game The grammar task involved classifying French words as *feminine, masculine,* or *plural* based on their determiner (*La, Le, Un, Une, Les,* or *Des*). The robot selected a category, and the child evaluated its decision. This activity was designed to support rule inference and metacognitive engagement, which have been shown to benefit from learning-by-teaching scenarios [25]. The task was modelled as an MDP with six states (determiners) and three actions (grammatical categories).

4.3.3 Study Design

The learning scenario was implemented using a hybrid system composed of three interconnected components: (1) a JD humanoid robot, (2) a touchscreen tablet

Figure 4.1: System architecture of the learning scenario. The setup comprises three interconnected components: a central computer hub running the pre-programmed learning pipeline, a JD humanoid peer robot, and a touchscreen tablet serving as the game interface. The hub communicates with the robot via MQTT and with the tablet via HTTP. The child interacts directly with the tablet to provide binary evaluative feedback, which is relayed to the hub to update the robot's policy in real time. Social interaction between the peer robot and the child unfolds within the classroom setup.



running the learning game, and (3) a central computer hub managing synchronisation and control. These components communicated via MQTT protocols, enabling real-time feedback loops between the child's input and the robot's behaviour. This configuration, illustrated in Figure 4.1, supports continuous updates to the robot's learning policy during the interaction.

The robot used in the study was JD[1], a small humanoid platform equipped with articulated limbs and LED-animated eyes. Although designed primarily for teaching programming concepts, JD was adapted for use as a peer-like learner in this study. To enhance child engagement, the robot's default voice was replaced with a child-like synthetic voice generated using ElevenLabs[2]. Additional behaviours, such as head movements aligned with speech and context-specific gestures, were developed to promote naturalistic and socially credible interaction. The robot was programmed using the EZ-Robot ARC platform, within which all behaviours, including head movements aligned with speech and context-specific gestures, were pre-programmed

[1]https://www.ez-robot.com
[2]https://elevenlabs.io/

and triggered upon receipt of the corresponding MQTT signals from the central computer hub.

The touchscreen tablet served as the main interface for the learning activity. It displayed quiz questions and multiple-choice options, along with a progress tracker and a help button. After the robot selected an answer, the screen presented binary feedback buttons, allowing the child to indicate whether the robot's response was correct or incorrect. The tablet interface across the different conditions is illustrated in Appendix B (Figure B.2).

Each game round followed a structured sequence: the robot first introduced itself and explained the task, then attempted to answer each question. It waited for child feedback, prompting the child after ten seconds of inactivity. If no response was given within an additional fifteen seconds, the robot encouraged the child to use the help button. Feedback was acknowledged both verbally and visually, with eye colour changes (green for correct, red for incorrect). When an error was made, the correct answer was reviewed collaboratively. In the vocabulary task, the robot also reinforced learning by pointing to the relevant body part during the review. At the end of each session, the robot thanked the child and simulated falling asleep to signal the conclusion of the activity. A video demonstration of a full game round is available online[3].

To maintain ecological validity, all sessions were conducted within the children's regular classroom environment. Multiple one-on-one child–robot interactions were carried out in parallel within the same room, preserving the naturalistic peer-group setting and minimising experimenter intervention. This design choice aimed to reduce potential social desirability bias and suggestibility effects, which are known to influence child responses in controlled lab settings [298]. Maintaining the classroom context was especially important given the study's focus on peer-like interaction and collaborative learning dynamics.

[3] https://youtu.be/7APP2ZUaizQ

4.3.4 Experimental Conditions

The study was structured as two independent between-subjects experiments; one for each learning task (Body Parts and Grammar). This design allowed learning effects to be attributed to the interaction condition without introducing carryover or familiarity effects that might confound outcomes in a within-subject design. Conducting two separate experiments also minimised the risk of ceiling effects in post-tests by avoiding repeated exposure to the same task.

Each experiment included two conditions:

- **Learning-by-Teaching (LbT condition)**: Children taught a social robot by providing binary evaluative feedback on its answers. The robot adapted in real time using the Interactive RL model, simulating a peer-like learning trajectory.
- **Self-Practice (Tablet condition)**: Children completed the same learning task independently on a touchscreen tablet, with no robot present. They selected the option they believed was correct; the interface then indicated whether the choice was correct and, if not, displayed the correct answer.

In the Self-Practice condition, children could not submit their answer immediately; instead, a 10-second delay was imposed before they could respond. This ensured equivalent reflection time across conditions, prevented children from guessing too quickly, and matched the pacing of the LbT condition, where feedback could only be given once the robot had produced its answer. By controlling timing in this way, we reduced the risk of random responding and avoided fatigue differences between groups.

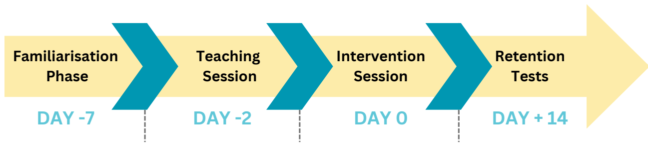
Moreover, to preserve classroom cohesion and prevent social exclusion, all children participated in both learning conditions, but across different tasks. For example, children who interacted with the robot in the Body Parts game completed the Grammar game on the tablet, and vice versa. This structure ensured ecological validity while enabling controlled comparisons between the learning conditions. Figure 4.2 shows the spatial organisation of the sessions within the school.

Figure 4.2: Floor plan of the study setup for both conditions. (a) Learning by Teaching condition: Children interacted with the robot in a classroom setting, with robots and tablets distributed around the room. (b) Self-Practice condition: Children worked independently on tablets in a communal area.

4.3.5 Procedure

The study was conducted over a three-week period and followed a four-phase protocol: (1) a familiarisation session, (2) a teacher-led lesson, (3) the main intervention session, and (4) a delayed retention test. The overall timeline is illustrated in Figure 4.3.

Figure 4.3: Study timeline. Familiarisation Session (Day −7) to introduce children to the robot; Teaching Session (Day −2) to deliver baseline instruction; Intervention Session (Day 0) involving the robot or tablet; and Retention Test (Day +14) to assess long-term knowledge retention.



Familiarisation Session. To reduce novelty effects and increase children's comfort with the robot, a familiarisation session was conducted one week before the intervention. The JD robot was introduced in the classroom, where it greeted the children, demonstrated simple movements, and performed a short dance. Children then engaged in a group activity on the classroom smartboard, where they took turns using the feedback buttons to teach the robot the names of fruits in French. This activity introduced the mechanics of the upcoming study in a playful and low-pressure context. The session ended with a Q&A where children could ask questions about the robot's abilities and behaviour.

Teaching Session. Two days before the intervention, teachers delivered structured lessons covering French body parts and grammatical determiners using standardised materials prepared by the research team. This ensured all children were introduced to the relevant content prior to the study and positioned the robot as a learning partner rather than a source of instruction. The two-day gap between teaching and intervention was intended to mitigate ceiling effects and maintain ecological validity.

Intervention Session. Each class was divided into three rotating groups: one completed the task with the robot (LbT condition), another used the same game on a tablet without the robot (Self-Practice condition), and the third engaged in an unrelated computer science activity. Children experienced both robot and tablet conditions across different tasks: those who taught the robot in the Body Parts game used the tablet for the Grammar game, and vice versa.

Each session lasted approximately 20 minutes and followed a consistent structure: a pre-test to assess baseline knowledge, 15 iterations of the learning game, an immediate post-test, and a brief questionnaire. Robot interactions took place in the classroom, while tablet sessions were conducted in a nearby communal area. This setup preserved the classroom context while enabling parallel sessions and minimising direct experimenter involvement.The learning game was implemented in practice through the TAMER-based cognitive model presented in Chapter 3 described in Algorithm 1. The pseudo-code illustrates how child feedback is integrated into the robot's policy update and action selection in real time.

Retention Test. Two weeks after the intervention, children completed the same knowledge assessments on tablets to measure retention. These were conducted during normal class hours and mirrored the original testing procedure. All children completed retention tests for both tasks, as every participant had experienced both tasks in either the LbT or Self-Practice condition. To conclude the study and thank participants, each child received a certificate, and the JD robot performed a final dance routine in the classroom.

4.3.6 Measures

To evaluate the impact of the learning-by-teaching interaction, three types of measures were used: (1) knowledge assessments, (2) in-game behavioural metrics, and (3) post-intervention questionnaires. These instruments were designed to capture learning outcomes, user engagement, and subjective perceptions of the experience.

Knowledge Assessments. Children completed three touchscreen-based tests for each task: a pre-test (administered before the learning activity), a post-test (immediately after), and a retention test (two weeks later). All tests were delivered

Algorithm 1 Interactive RL for Learning-by-Teaching a Peer Robot

Require: State set S (questions), action set A (answer options), learning rate α, iterations T, feedback window τ_{fb}, hint window τ_{hint}

Init: $Q(s, a) \leftarrow 0$ for all (s, a) ▷ reset Q at the start of each child session
for $t = 1$ to T **do**
  $s_t \leftarrow \textsc{DrawQuestion}(S)$
  $a_t \leftarrow \textsc{Max}(Q[s_t, \cdot])$
  $\textsc{RobotSelectsAndSpeaks}(a_t)$; $\textsc{ShowOnTablet}(a_t)$
  $h \leftarrow \textsc{WaitFeedback}(\tau_{\text{fb}})$ ▷ $h \in \{+1, -1\}$
  if $h = \textsc{None}$ **then**
    $\textsc{PromptChild}()$
    $h \leftarrow \textsc{WaitFeedback}(\tau_{\text{hint}})$
  end if
  $Q[s_t, a_t] \leftarrow Q[s_t, a_t] + \alpha \cdot (h - Q[s_t, a_t])$ ▷ TAMER update
  $\textsc{Log}(t, s_t, a_t, h, Q[s_t, a_t], \text{time}, \text{help\_ms}, \text{fb\_acc})$
end for

on tablets using the same interface as the learning game to ensure consistency. Each test comprised three rounds to reduce the effect of random guessing.

For the *Body Parts* task, the tests included: (1) matching English body part names to their French equivalents, (2) matching French words to English, and (3) matching French terms to corresponding images, each aimed at assessing vocabulary memorisation.

For the *Grammar* task, children classified French nouns as masculine, feminine, or plural based on their determiners. To evaluate rule generalisation rather than rote memorisation, the test included novel words that had not been presented during the intervention session.

Knowledge gain and retention gain were computed by subtracting pre-test scores from post-test and retention test scores, respectively.

In-Game Behavioural Metrics. Interaction logs from the learning activity were used to capture real-time engagement and learning behaviour. Specifically, the following metrics were recorded:

- **Feedback Accuracy:** Whether the child correctly identified the robot's answers as correct or incorrect.

- **Time per Iteration:** The time elapsed between the robot's answer and the child's feedback.

- **Hint Usage:** Whether, and how often, the child used the hint button before providing feedback.

These measures were used to infer cognitive engagement, reflection, and help-seeking behaviour across conditions.

Perception Questionnaire. Following the intervention, children completed a short self-report questionnaire to evaluate their experience. The instrument measured perceived competence, task enjoyment, and engagement. Items were adapted from the Intrinsic Motivation Inventory (IMI) [236] to align with the classroom context and participant age.

To ensure usability and accessibility, the questionnaire used a child-friendly 5-point Likert scale represented as a star-rating system. Questions were displayed individually in random order, and children responded by tapping the appropriate number of stars. This design was informed by previous work on self-report measures for children and was pre-tested in classroom settings to ensure clarity [45]. Internal consistency scores (Cronbach's alpha) for each subscale, and its corresponding phrasing are provided in Appendix A.

4.4 Results

4.4.1 Effectiveness of Teaching Through Feedback

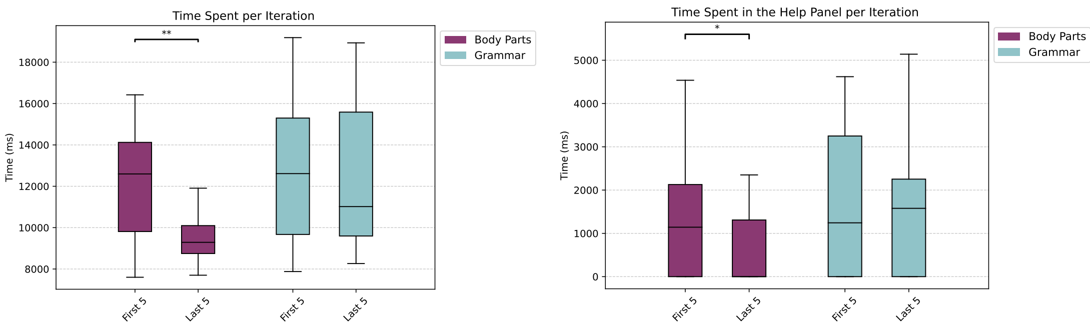
To address RQ2.1, we examined whether children could guide the robot's learning through Interactive RL by providing accurate feedback and adapting their tutoring behaviour over time.

Children were generally able to provide accurate feedback to the robot, with a mean accuracy of $M = 0.89$ ($SD = 0.13$) in the Body Parts Game and $M = 0.74$ ($SD = 0.29$) in the Grammar Game. These results indicate that children understood the goal of the interaction and assumed the role of a teacher, actively supporting the robot's learning process by identifying and correcting its mistakes.

In addition, interaction data suggest that children adapted to their role as a tutor over time. As shown in Figure 4.4a, their response time per iteration significantly decreased during the Body Parts Game, from a median of 12596.52 ms in the first five rounds to 9289.20 ms in the last five, $U = 607.00$, $z = 2.90$, $p < .01$, $r = .38$. This indicates that children became quicker in evaluating the robot's answers as the session progressed.

A similar pattern was observed in their use of the help panel (Figure 4.4b). Children initially spent a median of 1142.40 ms using the help feature, which decreased to 0.00 ms by the final rounds, $U = 563.50$, $z = -2.61$, $p < .05$, $r = -0.34$. This reduction suggests growing confidence in their ability to assess the robot's responses without additional support. These trends were not observed in the Grammar Game.

Figure 4.4: Comparison of time spent during the game between the first and last five iterations. (a) Time spent per iteration. (b) Time spent on the help panel. Significant adaptation was observed in the Body Parts Game.



(a) **(b)**

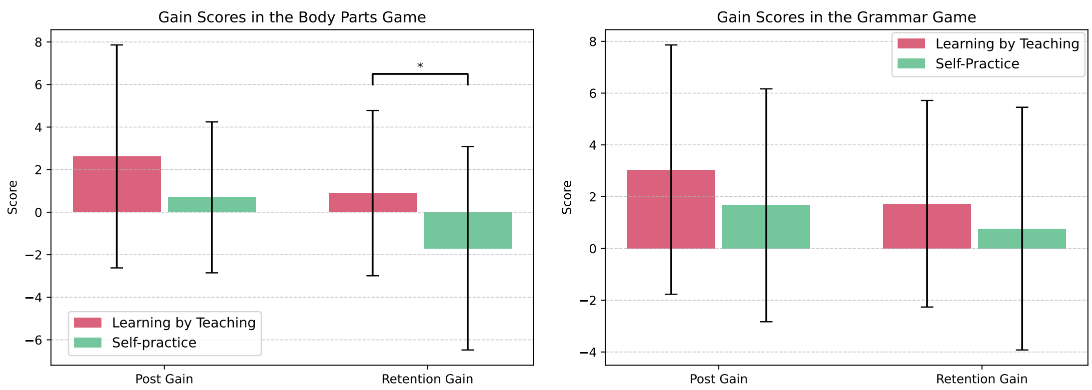
Taken together, these patterns suggest that children became increasingly familiar with the task and began to engage more strategically with the tutee robot. This supports the feasibility of using Interactive RL as a learning model for robot tutees in classroom settings, with children effectively guiding the robot's learning while adjusting their own tutoring behaviour over time.

4.4.2 Learning-by-Teaching Versus Self-Practice

To address RQ2.2, we compared the learning outcomes of learning-by-teaching a social robot with the self-practice condition, while also examining whether the perception of the task differed between the two approaches.

Figure 4.5 shows that, on average, children in the LbT condition achieved higher post-test and retention gains compared to those in the Self-Practice condition in the Body Parts Game, where this difference reached statistical significance. Children in the LbT condition showed a mean retention gain of $M = 0.90$ ($SE = 3.89$), compared to a negative gain of $M = -1.70$ ($SE = 4.78$) in the Self-Practice group, $t(50) = 2.16$, $p < .05$, $r = .29$. This suggests that LbT supported better long-term learning in the vocabulary task, and that children in the control condition may have forgotten or performed worse than at baseline. No significant differences were found between conditions in the Grammar Game.

Figure 4.5: Retention and post-test gains by condition. (a) Body Parts Game: higher retention in LbT. (b) Grammar Game: no significant difference.

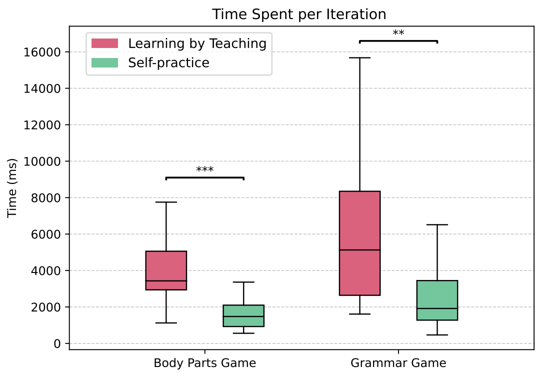


(a) **(b)**

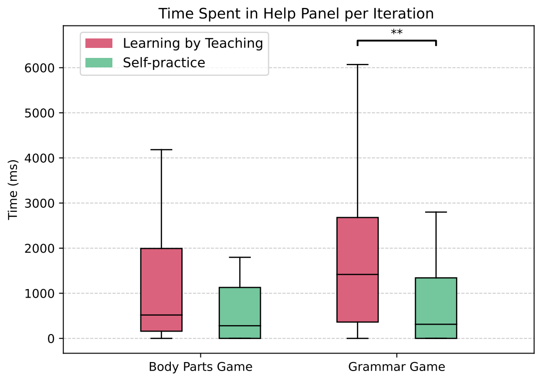
Children in the LbT condition spent significantly more time per iteration than those in the Self-Practice condition, as shown in Figure 4.6a. In the Body Parts Game, their median iteration time was 3431.66 ms, more than double that of the Self-Practice group (1476.40 ms), $U = 571.0$, $z = 0.88$, $p < .0001$, $r = .12$. A similar pattern was observed in the Grammar Game, where children in the LbT condition took longer to respond ($Mdn = 5127.50$ ms) compared to the control group ($Mdn = 1916.03$ ms), $U = 535.0$, $z = 0.76$, $p < .001$, $r = .11$. These findings suggest that children remained engaged with the task for longer periods in the LbT condition.

Children in the LbT condition also spent more time using the help panel on average, though this difference was only statistically significant in the Grammar Game. As shown in Figure 4.6b, children in the LbT group spent a median of 1417.13 ms reviewing the material, compared to 312.80 ms in the Self-Practice group, $U = 468.0$, $z = 2.56$, $p < .01$, $r = .35$. Additionally, in the LbT condition, time spent on the help panel was positively correlated with retention gain ($r = 0.30$, $p < .05$), whereas no such relationship was observed in the Self-Practice condition ($r = -0.01$, $p > .05$). This suggests that children in the LbT condition engaged more actively in self-reflection and material review, which may have contributed to their higher retention gain.

Figure 4.6: Time spent per iteration (a) and on the help panel (b). Significant differences emerged across both tasks, particularly in LbT.



(a)



(b)

Table 4.1 summarises children's self-reported ratings of engagement, perceived competence, enjoyment, and task difficulty across the two learning conditions. No significant differences were found between the LbT and Self-Practice groups, except for engagement in the Grammar Game, where children in the LbT condition reported significantly higher engagement ($p < .05$). An exploratory analysis further showed that, in the LbT condition, time spent per iteration was positively correlated with perceived engagement ($r = 0.365$, $p < .05$), a pattern not observed in the Self-Practice group ($r = -0.01$, $p > .05$). Perceived difficulty was rated as moderate across both conditions, suggesting that the LbT interaction was not experienced as more demanding than independent practice.

Table 4.1: Perceived experience ratings across tasks and conditions. Significant differences in bold.

Variable	Body Parts Game			Grammar Game		
	LbT	**Self-Practice**	**p**	**LbT**	**Self-Practice**	**p**
Task Enjoyment	4.3 (0.83)	4.2 (0.81)	0.92	4.2 (0.97)	4.2 (1.07)	0.99
Perceived Competence	4.0 (0.81)	4.3 (0.81)	0.41	4.3 (1.07)	4.4 (1.07)	0.95
Engagement	3.7 (1.07)	4.1 (0.99)	0.11	**4.2 (0.99)**	3.3 (1.46)	**0.03**
Difficulty	2.3 (0.94)	2.3 (1.06)	0.96	2.6 (1.26)	2.6 (1.46)	0.73

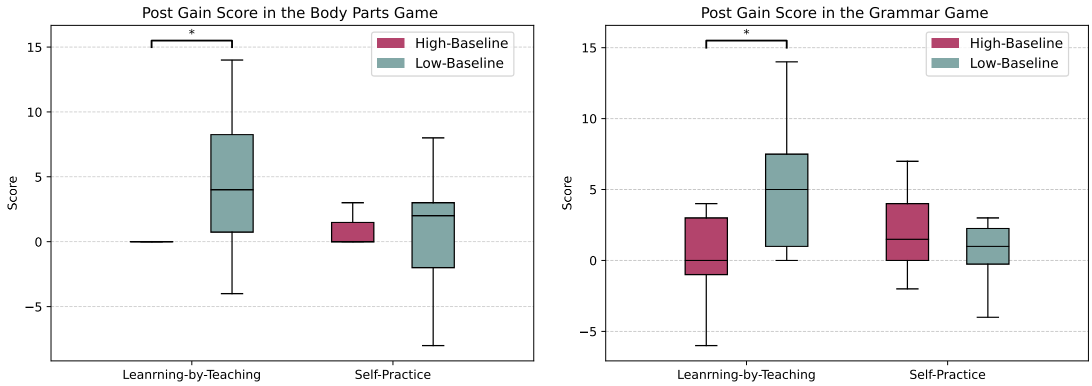
4.4.3 Influence of Prior Knowledge

To address RQ2.3, we explored whether prior knowledge moderated learning gains. Participants were divided into two groups for each task and condition based on the median pre-test score: a Low-Baseline group (scores at or below the median) and a High-Baseline group (scores above the median). The median cut-off scores, along with demographic information such as gender distribution and pre-test performance for each group, are provided in the Appendix B.

Figure 4.7 shows that, within the LbT condition, students with lower prior knowledge (Low-Baseline) achieved significantly greater post-test gains than their higher-performing peers (High-Baseline). This effect was not observed in the Self-Practice condition, where gains were minimal for both learning methods. Specifically, in the Body Parts Game, Low-Baseline students showed a median gain of 4.00, compared to 0.00 for High-Baseline students, $U = 52.00$, $z = -2.86$, $p < .05$, $r = -0.51$. A similar pattern emerged in the Grammar Game, with Low-Baseline learners achieving a median gain of 5.00, again significantly higher than the High-Baseline group ($Mdn = 0.00$), $U = 29.00$, $z = -2.07$, $p < .05$, $r = -0.44$.

Figure 4.8 shows that this effect also extended to retention gains, with Low-Baseline students in the LbT condition retaining more knowledge than their High-Baseline peers. This difference was significant in the Body Parts Game, where Low-Baseline students had a median retention gain of 2.00, compared to -1.00

Figure 4.7: Post-test learning gains in LbT vs. Self-Practice, split by prior knowledge. Stronger effects observed for Low-Baseline learners in LbT.

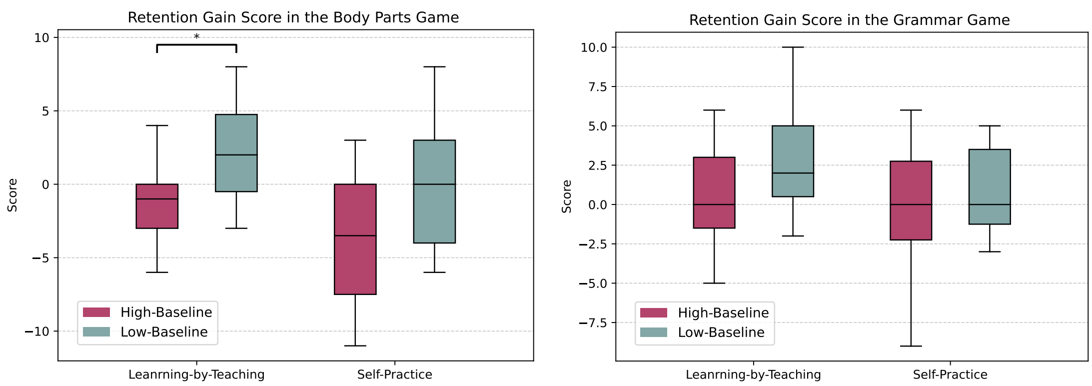


(a) **(b)**

for High-Baseline students, $U = 51.00$, $z = -3.37$, $r = -0.59$. A similar trend was observed in the Grammar Game, where Low-Baseline students ($Mdn = 2.00$) outperformed the High-Baseline group ($Mdn = 0.00$), though the difference did not reach statistical significance ($U = 40.00$, $p > .05$), with a medium effect size ($r = -0.44$). In contrast, retention gains in the Self-Practice condition were minimal for both groups, with Low-Baseline students showing a median gain of 0.

These results suggest that the LbT condition was particularly beneficial for students with lower prior knowledge, enabling them to achieve greater learning gains than their higher-performing peers, an effect not observed in the Self-Practice condition.

Figure 4.8: Retention gain scores by baseline knowledge. LbT particularly benefited learners with low prior knowledge.



(a) **(b)**

4.4.4 Influence of Task Type

Finally, to address RQ2.4, we investigated how task type affected interaction and perception within the LbT condition.

Although ratings for enjoyment and competence were similarly high across tasks, engagement was significantly higher in the Grammar Game ($Mdn = 4.67$) than the Body Parts Game ($Mdn = 4.00$), $U = 209.50$, $z = -2.08$, $r = -0.29$.

Correlational analysis revealed that, in the Grammar Game, iteration time was positively associated with enjoyment ($r = .56$, $p < .01$), competence ($r = .41$, $p = .05$), and engagement ($r = .46$, $p < .05$). Time spent on the help panel was also positively correlated with retention gain ($r = .61$, $p < .01$). These relationships were absent in the Body Parts Game, suggesting deeper cognitive engagement in rule-based tasks.

Taken together, these findings highlight the potential of inference-based tasks to promote deeper engagement, self-reflection, and stronger retention when paired with an interactive peer robot. The positive associations between interaction time, help usage, and learning outcomes suggest that rule-inference tasks may better support reflective learning processes, an effect not observed in the more memorisation-focused Body Parts Game.

4.5 Discussion

4.5.1 Interactive RL as a Cognitive Model for Peer-Learning Robots

The results show that children effectively adopted the role of tutor by consistently providing accurate feedback to the robot, in line with prior work showing that children can successfully guide a robot's learning using evaluative feedback, demonstrating their attentiveness and ability to self-evalaute the robot's learning [158, 295]. In the Body Parts Game, children became more efficient over time, as shown by the decreasing response time across iterations. This mirrors previous findings in LbT research, where repeated tutoring interactions help learners refine their understanding through practice and self-monitoring [79, 232].

The observed reduction in help panel usage further suggests that children grew more confident in their role and increasingly relied on their own knowledge. This behavioural adaptation indicates that the Interactive RL setup supported sustained tutoring engagement, particularly in tasks that allowed for progressive skill development. By contrast, this adaptation was not observed in the Grammar Game, where help panel usage remained relatively stable. One possible explanation is that the task may have been more challenging for this age group, as it introduced multiple unfamiliar words and grammatical forms. Although this was not reflected in the self-reported difficulty ratings, it may have led children to rely more consistently on the help panel for support [25, 232].

Overall, these findings indicate that children were able to take on the tutor role effectively, responding to the robot's learning needs and adjusted their behaviour accordingly, with clearer signs of adaptation emerging in tasks that focused on straightforward recall.

4.5.2 Learning-by-Teaching Enhances Engagement and Retention

This study explored the impact of teaching a social robot versus independent practice, with a focus on learning outcomes and engagement. An important finding of this study is that the presence of a social robot did not distract children or hinder their learning. In contrast to previous work suggesting that robots may reduce task focus or interfere with performance in educational settings [135], children in the LbT condition maintained or even exceeded the learning gains observed in the Self-Practice condition. This suggests that when appropriately framed as a peer learner, a robot can serve as a productive learning partner rather than a source of distraction.

While overall learning gains were higher in the LbT condition, the difference was not statistically significant across both tasks. One possible explanation is that children had already begun consolidating knowledge prior to the intervention. During both the familiarisation and teaching sessions, children were told they would be teaching the robot, which may have prompted them to prepare mentally for that role. Prior research has shown that even the act of preparing to teach, without actually teaching, can enhance learning outcomes [25, 147]. This preparatory effect may have benefited both groups and reduced the observable difference during the main intervention.

Despite this, children in the LbT condition achieved significantly higher retention in the Body Parts Game, indicating a positive long-term effect of teaching the robot. The social nature of the interaction likely contributed to this outcome. Prior work has shown that physically embodied agents can enhance learning through increased engagement and attention [162, 135, 164], particularly when they combine gesture and verbal feedback [279]. The robot's behaviour may have activated a sense of responsibility in the child, consistent with the Protégé Effect, which suggests that learners put more effort into mastering material when they are teaching someone else [46].

This interpretation is further supported by behavioural data. Children in the LbT condition spent more time per iteration and used the help panel more frequently, especially in the Grammar Game, where help usage was positively correlated with retention. These patterns suggest that teaching the robot promoted active review and reflection during the task, consistent with prior findings that tutoring a social agent can encourage metacognitive engagement [232, 25].

In sum, children teaching the robot demonstrated higher engagement and stronger retention, and importantly, the robot did not distract from learning. On the contrary, it appeared to support sustained interaction and reflective behaviour, particularly in tasks that required greater attention or review.

4.5.3 Children with Lower Prior Knowledge Benefit More from Learning-by-Teaching

The results also indicate that prior knowledge played an important role in shaping learning outcomes, particularly in the LbT condition. In both the Body Parts and Grammar tasks, students with lower pre-test scores (Low-Baseline) achieved greater learning gains than their higher-performing peers. This difference was statistically significant in the Body Parts Game for both post-test and retention scores, while a similar, but non-significant, trend was observed in the Grammar Game.

These findings are consistent with earlier research suggesting that learners with lower prior knowledge benefit more from LbT, as the process encourages them to engage in self-explanation and error correction [25, 232]. Teaching the robot may have given these students structured opportunities to reflect on their understanding, helping them identify gaps and reinforce core concepts through interaction.

Interestingly, this pattern was not observed in the Self-Practice condition, where no clear differences emerged between Low- and High-Baseline students. This suggests that students with weaker initial knowledge may not have engaged as deeply with the material when working independently. In contrast, the interactive and social nature of LbT may have acted as a scaffold that supported deeper engagement for this group.

It is also important to consider that High-Baseline students had less room for improvement due to their already strong prior knowledge, which may have contributed to their smaller gains, a common ceiling effect in learning research.

Together, these findings show that LbT was particularly beneficial for children with lower prior knowledge, enabling them to achieve greater gains than would be expected through self-guided practice alone.

4.5.4 Rule-Inference Tasks Promote Deeper Engagement in Learning-by-Teaching

The findings also suggest that the type of learning task shaped how children engaged with the LbT interaction. In particular, children who completed the Grammar Game reported significantly higher engagement than those who played the Body Parts Game. This may reflect the greater complexity of the grammar task, which introduced unfamiliar structures and required children to infer rules. Although cognitive demand was not directly measured, this result aligns with research showing that effortful learning can increase engagement and satisfaction when learners perceive that they are making progress [27].

Despite this, no significant differences were found between the two tasks in terms of self-reported enjoyment or perceived competence, likely due to ceiling effects in the questionnaire responses, as previously discussed.

While overall interaction time and help usage did not differ significantly between tasks, correlational analyses revealed clear differences in how children engaged. In the Grammar Game, time spent per iteration was positively associated with reported enjoyment, competence, and engagement. Moreover, time spent on the help panel was strongly correlated with retention gain. These associations were not observed in the Body Parts Game, suggesting a more reflective learning process in the grammar task.

These patterns point to deeper cognitive engagement in the Grammar Game, possibly due to the need for rule-based reasoning rather than simple recall. This interpretation is supported by earlier research indicating that LbT is especially effective for tasks that involve higher-order thinking, such as conceptual understanding and problem-solving [25, 232]. The lower and decreasing use of the help panel in the Body Parts Game further suggests that it relied more on rote memorisation, which may not require the same level of review or reflection during the activity.

Overall, the Grammar Game seemed to promote higher engagement and deeper learning processes, as reflected in help panel usage and its association with retention.

However, due to ceiling effects in self-reported measures, we cannot draw firm conclusions about whether children perceived the task as more effective or enjoyable.

4.6 Limitations and Future Work

While this study offers valuable insights into the impact of LbT with a social robot in a real classroom setting, several limitations should be acknowledged.

First, the study was conducted in a single UK primary school with a specific age group and curricular context. This may limit the generalisability of the findings to other educational settings, age ranges, or student populations. Future research should seek to replicate the experiment across multiple schools, including those with diverse socio-economic backgrounds and varying levels of technological familiarity, in order to assess the robustness and broader applicability of the results.

Second, the study relied on self-report measures of engagement, enjoyment, and perceived competence. These measures showed signs of ceiling effects, which are common in child-focused research and may have reduced the sensitivity to detect differences between conditions [76]. Although the use of child-friendly Likert scales was necessary to ensure accessibility, future studies could benefit from incorporating more objective or multimodal indicators of engagement, such as gaze tracking, facial expression analysis, or speech features, to gain a more nuanced picture of children's experiences [2].

In addition, although the tasks were co-designed with teachers to align with curriculum goals, it is possible that the Grammar Game was more cognitively demanding or less developmentally appropriate for some children. While no differences were reported in perceived difficulty, the need for sustained help usage and slower interaction patterns may reflect task complexity that warrants closer examination in future work.

Finally, while this study evaluated the effectiveness of Interactive RL as a mechanism for peer-like robot learning, it did not compare this model to alternative approaches such as scripted behaviours, Wizard-of-Oz control, or different learning frameworks. Future research could investigate how different levels of robot autonomy and learning transparency affect children's teaching behaviour, learning outcomes, and perceptions of the robot.

Together, addressing these limitations can help refine the design and deployment of teachable robots in educational contexts, and contribute to the development of more adaptive, inclusive, and scalable Learning-by-Teaching systems.

4.7 Summary

This chapter presented a classroom-based study evaluating the use of Interactive RL as a cognitive model for peer-like social robots in primary education. By enabling the robot to learn from children's evaluative feedback in real time, the study explored how LbT could be implemented beyond controlled laboratory settings.

The findings demonstrated that children were able to tutor the robot effectively, providing accurate and consistent feedback while adjusting their behaviour over time. This supports the feasibility of Interactive RL as a mechanism for sustaining bi-directional teaching interactions in naturalistic classroom environments.

Children in the LbT condition achieved significantly higher retention gains in the vocabulary task and showed greater behavioural engagement across both learning activities. The study also revealed that children with lower prior knowledge benefited most from teaching the robot, suggesting that this approach may serve as a form of personalised scaffolding that supports educational equity.

Task type was found to influence the quality of the interaction: the rule-inference Grammar Game elicited higher engagement, more reflective behaviour, and stronger associations with retention, compared to the memorisation-based Body Parts Game. These results underscore the importance of cognitive task design when implementing robot-mediated learning.

Beyond learning outcomes, the study also demonstrated the practical viability of deploying multiple autonomous robots in real classrooms with minimal experimenter intervention. This contributes to the growing body of evidence supporting the integration of adaptive, socially interactive robots into everyday educational practice.

Together, these findings highlight the potential of Interactive RL to move beyond scripted behaviours, enabling robots to act as more responsive learning partners that may support deeper engagement, personalised learning, and classroom integration in real-world educational settings.

Chapter 5

Impact of Involvement on Trust Toward Robots

"Why, sometimes I've believed as many as six impossible things before breakfast."

- Lewis Carroll, *Through the Looking-Glass*

5.1 Introduction

The previous chapter presented an interactive RL framework implemented in a real-world classroom setting, where children taught a social robot and demonstrated improved retention and engagement. While those findings focused on cognitive and learning outcomes, they nonetheless raise important considerations for the future deployment of such systems, particularly regarding how children might perceive and evaluate a learning robot as a social and epistemic partner. This chapter extends that line of inquiry by examining how different levels of user involvement in teaching the robot influence perceptions of anthropomorphism and trust (affective, cognitive, and epistemic). In doing so, it investigates the psychological mechanisms that underlie users' relational stance toward the robot, and how individual differences mediate these effects. The chapter presents the experimental results and offers a detailed analysis of the social-cognitive dimensions of trust in human-robot interaction.

5.2 Study Motivation & Research Questions

Social bonds and trust are fundamental to children's development and learning. From early childhood, children begin forming affective and cognitive trust with their peers, which support cooperation, emotional sharing, and effective communication [110]. Affective trust refers to an emotional bond characterised by feelings of closeness, safety, and interpersonal care. It is associated with attitudes about an agent's perceived warmth and benevolence. In contrast, cognitive trust rests on more knowledge-driven, reasoned judgments regarding an agent's reliability, competence, and trustworthiness of intentions [223, 179, 187, 159].

Developmental studies show that by age three, children begin forming affective trust by starting to show preferences for certain peers, gradually deepening friendships through reciprocal sharing and confiding in one another [110]. Cognitive trust develops in parallel as children's theory of mind and social reasoning improve, allowing them to evaluate whether peers act with integrity and consistency [35, 105]. This trust is often stronger for familiar friends compared to unknown or disliked peers [39].

In the context of social robotics, presenting robots as peers can elicit similar trust-related responses. Children have been observed to form emotional bonds with robots [123]. In a study conducted by Turkle et al, [269], children expressed a desire to nurture or be nurtured by the robot, leading to attachments that extended beyond simple interaction. While these social bonds can enhance learning engagement, they also raise ethical questions about the nature and consequences of such attachments.

Beyond affective and cognitive trust lies epistemic trust; the trust placed in others as reliable sources of knowledge [145, 40, 257]. This form of trust is particularly critical in learning scenarios, as it influences whether children accept information as truthful and worthy of integration into their understanding [146]. Research indicates that preschoolers begin to demonstrate selective epistemic trust from around age three, with a more consistent preference for accurate and knowledgeable informants emerging by age four [146]. This extends to social robots as well; studies have shown that children are capable of discerning the reliability of robots as information sources. For instance, research indicates that children aged five prefer to learn from a competent robot over an incompetent human, highlighting their ability to evaluate epistemic characteristics independently of social cues [260].

However, the dynamics of epistemic trust towards peer-like robots remain underexplored. While children may assess a robot's competence, the nature of their involvement in the learning process is likely to influence how trust develops. Involvement has been theorised as a critical determinant of psychological ownership, since the more a user invests effort and personal input into shaping a system, the stronger their sense of "this is mine" becomes [157]. Empirical work in HRI supports this view, showing that active contribution to a robot's behaviour can enhance users' feelings of control and self-investment [156, 154, 155].

In educational contexts, active teaching has been shown to enhance trust in a robot's knowledge and capabilities [56]. One mechanism that may underlie this effect is psychological ownership, which has been linked to trust: when users feel that a system reflects their own input and aligns with their intentions, they are more likely to regard it as reliable and trustworthy [212, 157] Conversely, the risk of misplaced or excessive trust raises important ethical considerations, especially when learners attribute epistemic authority to a system that remains fallible [248].

The growing prominence of generative AI technologies, such as conversational agents, tutoring systems, and educational tools, that increasingly interact with children underscores the urgent need to study epistemic trust [167, 31]. Children are exposed to AI systems capable of generating information, advice, and creative content, making it essential to understand how they evaluate the reliability of these sources and how their involvement with such systems influences their trust and learning outcomes. Importantly, a moderate level of trust in the robot is beneficial, as it enables children to engage effectively in learning with and from their peer-like robotic companions [278]. However, trust must also be calibrated appropriately. Excessive trust may lead individuals, including children, to uncritically accept misinformation or over-rely on the robot, while insufficient trust may prevent them from considering the robot as a valuable learning partner at all.

Therefore, this chapter explores how varying degrees of involvement in teaching a peer-like robot influence the development of epistemic trust, psychological ownership, and broader trust dynamics. While the present study was conducted with adults, the findings are intended to inform the design of future child–robot educational contexts, where these constructs are most relevant.

To address these gaps, the present study investigates how different levels of involvement in teaching a peer-like robot shape users' perceptions of trust and ownership. Specifically, the following research questions are posed:

- **RQ3.1: Does participant involvement in a robot's RL process influence determinants of psychological ownership (PO) and overall psychological ownership toward the robot?**
  This question investigates whether actively teaching the robot increases users' sense of ownership and connection to it. Prior work shows that when people are involved in shaping a system, they are more likely to feel it is "theirs" [156, 71]. Psychological ownership theory explains this by identifying three key determinants: control, percieved knowledge, and self-investment [212, 129]. When these are present, people not only feel more responsible for the system but also more motivated and emotionally attached to it [68, 204, 251, 21]. This study therefore examines how involvement in teaching influences these ownership dynamics.

- **RQ3.2: Does participant involvement in the robot's RL process affect anthropomorphism of the robot?**
  This question investigates whether increased involvement in teaching the robot enhances perceptions of its human-like qualities. Anthropomorphism in HRI is often described through dimensions such as autonomy (the robot's ability to act independently) [83], sociability (its capacity for social interaction) [191], agency (attributions of intentionality and goal-directedness) [103], and animacy (the sense of being alive or lifelike) [87]. Greater engagement may amplify these motivations, fostering stronger anthropomorphic perceptions that, in turn, can support the development of trust [107, 108, 226, 28]. This study further

examines whether increased involvement in teaching reduces perceived disturbance, the extent to which the robot is experienced as distracting, intrusive, or disruptive to the task.

- **RQ3.3: Does participant involvement influence trust toward the robot?** This question examines whether greater involvement in teaching the robot leads to increased trust, encompassing affective trust , cognitive trust, and epistemic trust. Prior research suggests that active engagement shapes trust indirectly, by fostering self-investment, perceived control, and social connection, which are known to support these dimensions of trust [157]. This study hypothesises that greater involvement in the robot's RL process will lead to higher trust toward the robot.

5.3 Methodology

5.3.1 Participants

An a priori power analysis for t-tests with Bonferroni correction was conducted using G*Power (effect size $d = 0.50$, power = 0.80, $\alpha = 0.0167$), indicating a required sample size of 144 participants to detect medium effects across multiple comparisons. Participants were recruited at Bielefeld University (Germany). Initially, data were collected from 145 participants. The final sample comprised 143 adult participants (91 females, 36 males, 4 others, 3 preferred not to say; $M_{age} = 25.22$, $SD_{age} = 7.41$) who met the inclusion criteria: at least 18 years old, native or fluent German speakers, normal color vision, no outlier responses (>3 SD) in more than 5% of variables of interest, and complete questionnaire responses.This study was therefore conducted with an adult population to allow for a controlled investigation of involvement and trust dynamics. While this approach provides internal validity, the findings may not directly generalise to children, whose developmental differences could alter these processes. Future research is needed to examine how these dynamics unfold in child–robot educational contexts.

5.3.2 Experimental Task

Participants were invited to engage in a reasoning game involving the completion of a card sequence, supported by a social robot, as illustrated in Figure 5.1. Each card featured a geometric shape (circle, square, or triangle) in one of four possible colours (red, blue, green, or yellow). Only the first card in the sequence (card 0) was visible; all subsequent cards were concealed.

The task was governed by two abstract rules: one dictating the order of colours and the other the order of shapes. These rules were displayed for 30 seconds prior to the beginning of each trial. This time constraint was introduced to ensure participants engaged with the task without relying on continuous visual support, while also limiting the likelihood of ceiling effect [61]. Once the task began, the rules could not be reviewed anymore.

The characteristics of the task were determined through a pre-study involving 90 participants. In this preliminary investigation, variations in sequence length and the presence or absence of a memorisation component were systematically examined. The configuration that was ultimately selected was one which induced a neutral perception of the task, while maintaining a balanced cognitive load; sufficiently challenging to avoid ceiling effects, but not so demanding as to produce disengagement from the task. This balance was considered necessary to reliably measure epistemic trust.

Each sequence began with the single visible card, followed by a grey placeholder representing skipped cards (four cards), and then several black cards which participants were required to complete, as showed in Figure 5.1. From the first black placeholder onward, participants were instructed to infer both the shape and colour of each card based on their understanding of the two rules previously shown.

To enable the evaluation of epistemic trust, the robot was programmed to follow a colour rule that differed from the one provided to participants, as shown in Figure 5.2. This rule mismatch was not disclosed. Consequently, the robot's suggestions occasionally coincided with participants' answers, but diverged in some trials. This manipulation allowed the assessment of whether participants would

Figure 5.1: Reasoning game used to assess epistemic trust, where participants inferred missing cards based on colour and shape rules.

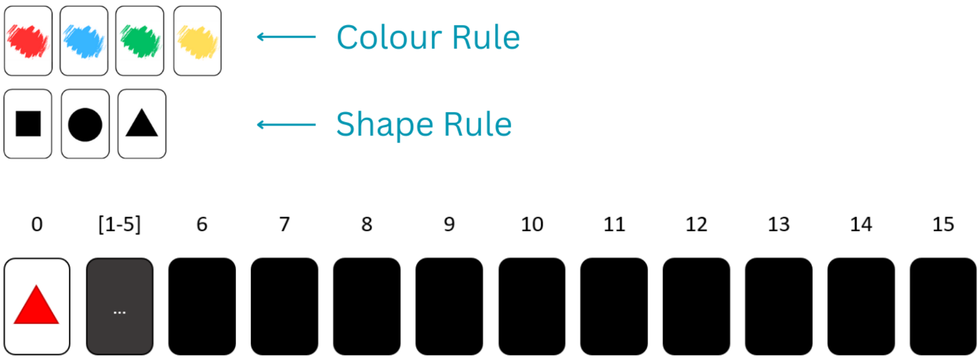
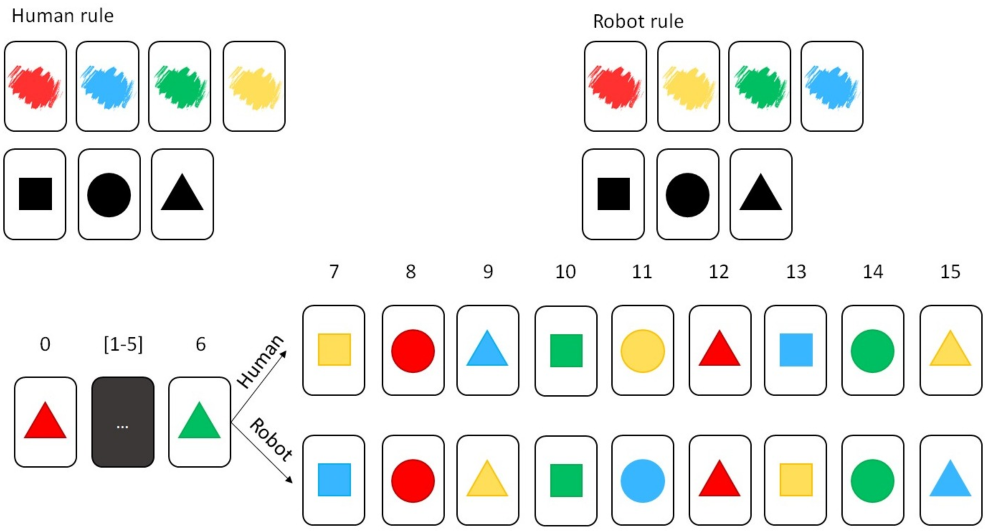


Figure 5.2: Illustration of the rule mismatch in the card reasoning game, showing the participant's rule, the robot's rule, and the correct answer.



accept or reject the robot's responses, even when those conflicted with the logic they had remembered, thus serving as a behavioural measure of epistemic trust.

Following each participant response, the robot provided its own answer. Participants were then given the opportunity to revise their answer in light of the robot's suggestion.

This interaction design enabled the examination of participants' trust-related behaviours, particularly the degree to which their decisions were influenced by the robot's differing input.

5.3.3 Experimental Conditions

A between-subjects design was employed to investigate how different levels of user involvement in teaching a robot to discriminate shapes affected participants' psychological ownership, anthropomorphism, and trust. The four experimental conditions systematically varied the degree to which participants contributed to the robot's reinforcement learning of basic shapes (circle, square, triangle), ranging from no involvement to direct hands-on teaching. This manipulation was grounded in the hypothesis that greater involvement in shaping the robot's behaviour would strengthen participants' beliefs about its cognitive competence, and thereby increase their willingness to treat its input as epistemically valid during the subsequent reasoning task involving the same shape categories (see Section 5.3.2).

The four conditions were defined as follows:

- **Control condition:** Participants in this condition received no information about the robot's learning process or its perceptual abilities. They engaged in an unrelated filler task while the robot was purportedly being trained out of sight.
- **Low Involvement condition:** Participants were told that the robot had undergone a training process to learn about shapes, but were not shown how it occurred. This condition provided minimal informational access to the robot's learning.
- **Medium Involvement condition:** Participants observed the robot being trained by the experimenter to learn about shape. They were able to see the learning procedure and how the robot responded to feedback, without directly participating in the training.
- **High Involvement condition:** Participants actively trained the robot themselves by providing corrective feedback through a reinforcement interface. They guided the robot in learning to discriminate shapes, thereby playing a direct role in shaping its behaviour.

Participants were randomly assigned to one of the four conditions. The final sample included 37 participants in the control condition, 38 in the low involvement condition, 33 in the medium involvement condition, and 35 in the high involvement condition.

5.3.4 System Implementation

For the experimental conditions involving training,a desktop interface connected to a NAO robot was implemented [101], with all components developed in Python. The shape learning task was modelled as an MDP, and the robot was trained using the Interactive RL framework introduced in Chapter 3. In this setup, the robot selected one of two shapes based on a target label, pointed toward it, and vocalised its choice. Human feedback, provided through the interface, was encoded as binary evaluative signals and used to update the robot's policy in real time, until convergence to the the optimal policy ($\simeq 15 iterations$).

In the subsequent reasoning game, the robot's behaviour was fully scripted and identical across all conditions, ensuring that any differences in trust-related behaviour could be attributed to participants' prior involvement in the training phase rather than to variation in the robot's performance. To assess epistemic trust, the robot's responses included both correct and systematically biased suggestions (see Figure 5.2), allowing to examine whether participants would accept or reject its input when it conflicted with the learned rules. In addition, participants explicitly rated how likely they believed the robot's answers were correct for both colour and shape trials.

5.3.5 Procedure

The study was conducted individually in a quiet room at Bielefeld University. Upon arrival, participants were welcomed, provided with an information sheet, and asked to give informed consent. They were then seated in front of a computer and introduced to the NAO robot.

All participants began with a short familiarisation phase in which the robot introduced itself, performed simple movements, and invited participants to imitate its gestures. This was intended to reduce novelty effects and establish a minimal level of comfort.

Participants then entered the training phase, where the level of involvement with the robot's shape-learning task varied according to the assigned condition (see Section 5.3.3). To control for time-on-task, participants in the low involvement and control conditions completed a filler memory game of equal duration, using card stimuli drawn from the later reasoning task.

Finally, all participants completed the reasoning game described in Section 5.3.2, followed by questionnaires and a debriefing.

5.3.6 Measures

A combination of self-report questionnaires and behavioural data was used to assess participants' perception of the robot, psychological responses to the interaction, and epistemic trust during the reasoning task. Unless otherwise specified, all Likert items were rated on a 7-point scale.

Manipulation Check

To assess whether the experimental manipulation had the intended effect, participants completed a questionnaire immediately after the training or filler task.

- **Perceived user involvement:** measured using five self-generated items to evaluate the participants' sense of contribution and influence regarding how the robot functioned.

- **Perceived robot autonomy:** assessed using four self-generated items based on the definition of autonomy by Beer et al. [22].

Dependent measure (pre-test)

Following the manipulation check, participants completed a set of self-report measures designed to assess the psychological effects of the experimental manipulation prior to engaging in the reasoning task. These measures were included to capture participants' subjective responses to the interaction and their relationship with the robot at that point in the study.

- **Determinants of Psychological Ownership:** Two items for perceived control, two items for self-investment, and three items for perceived knowledge were used, all adapted from Delgosha and Hajiheydari [71].
- **Psychological Ownership:** Three items based of sense of psychological ownership from [71], adapted from [272] to HRI context, and three items adapted from [139] taping self-extension.
- **Perceived closeness:** This was assessed using the Inclusion of Other in the Self (IOS) scale [12] It consists of a single-item pictorial measure of interpersonal closeness. Participants were asked to select the image that best represented the relationship between themselves and the robot, with overlapping circles indicating higher perceived closeness.
- **Antropomorphism:** 16 items of Human-Robot Interaction Evaluation Scale (HRIES) [256].
- **Cognitive and affective trust toward robots:** 20 items from the Multi-Dimensional Measure of Trust (MDMT) scale [177].

Behavioural Measure (during the game)

During the reasoning task, participants responded to several in-game prompts designed to capture both behavioural and self-reported indicators of epistemic trust and task-related confidence, for each itteration.

- **Conformation to the robot's suggestion:** measured on trials where the robot's answer was intentionally incorrect. On these trials, participants' final answer was compared to their initial response to determine whether they changed their answer to match the robot's suggestion. Responses were categorised as: full conformation (both shape and colour changed to match the robot), conformation for shape only, conformation for colour only, or no conformation. This measure served as the primary behavioural indicator of epistemic trust.

- **Perceived confidence in the robot's answer:** After each trial, participants rated how confident they were in the correctness of the robot's suggestion, with one rating for shape and one for colour, on a 7-point Likert scale.

Dependent measures (post-task)

Following the reasoning task, a set of control measures was administered to account for potential individual differences in task performance and technology perception.

- **Self-confidence in task performance:** assessed using a single self-generated item, asking participants how confident they were in the correctness of their answers throughout the reasoning task.
- **Prior experience with robots:** measured using three items adapted from [222], capturing participants' familiarity and past interactions with robotic systems.
- **Affinity for technology interaction:** measured using the 9-item ATI scale [93], which assesses individuals' general tendency to interact with and enjoy using technology.
- **Task workload:** assessed using a shortened version of the NASA Task Load Index (NASA-TLX) [109], excluding the physical effort item. The resulting five-item version captured perceived mental demand, temporal demand, performance, effort, and frustration.
- **Tendency to anthropomorphise:** measured using the 15-item Individual Differences in Anthropomorphism Questionnaire (IDAQ) [285].
- **Need for cognition:** assessed using the 6-item NCS-6 short form [170], which captures individuals' tendency to engage in and enjoy effortful cognitive activities.

- **Need for closure:** measured using 15 items from the scale developed by [228], targeting participants' desire for certainty and aversion to ambiguity.
- **Demographic variables:** included age, gender, education level, employment status, and German language proficiency.

In addition, participants in the high involvement condition answered two self-generated items asking whether they believed their feedback had been accurate during the training phase, and if not, whether this had been intentional. Finally, two open-text questions were included to assess attention and potential suspicion regarding the true purpose of the study.

5.4 Results

5.4.1 Effects of User Involvement on Perceived Ownership, Control, Knowledge, and Closeness

This section investigates whether varying levels of user involvement in the robot's learning process influenced participants' perceptions of self-investment, control, knowledge, ownership, and closeness. Descriptive statistics for all variables across the four experimental conditions are summarised in Table 5.1. Inferential statistics are presented below.

Perceived Self-investment. A one-way ANOVA revealed a significant effect of condition on self-investment, $F(3, 139) = 12.20$, $p < .001$, $\eta^2 = .21$. Planned contrasts indicated that participants in the high-involvement condition reported significantly greater self-investment than those in the combined medium, low, and control conditions, $t(139) = 5.48$, $p < .001$, $r = .42$. The comparison between the medium-involvement condition and the combined low and control conditions was only marginally significant, $t(139) = 1.75$, $p = .086$, $r = .17$, as seen in Figure 5.3a.

An ANCOVA was conducted with prior robot experience and individual tendency to anthropomorphise (IDAQ) entered as covariates. The analysis revealed a significant

Table 5.1: Descriptive statistics for Psychological Ownership and its proposed determinants across conditions. Values are reported as mean (std). Statistically significant differences are highlighted in bold.

Variable	High-Involvement	Medium-Involvement	Low-Involvement	Control
Self-investment	**3.10 (1.31)**	**2.06 (1.51)**	**1.53 (0.87)**	**1.57 (1.22)**
Perceived Knowledge	**2.30 (1.13)**	**2.46 (1.19)**	**1.90 (0.91)**	**1.93 (1.13)**
Perceived Control	2.56 (1.31)	2.33 (1.16)	2.33 (1.37)	1.86 (1.18)
Closeness	1.51 (0.77)	1.45 (1.02)	1.50 (0.88)	1.38 (0.63)
Psychological Ownership	1.55 (0.55)	1.51 (0.55)	1.50 (0.44)	1.50 (0.65)

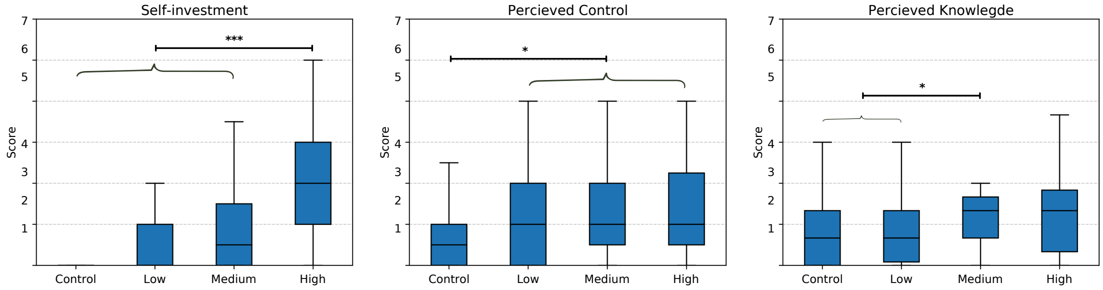
effect of condition on self-investment, $F(3, 137) = 10.57$, $p < .001$, $\eta_p^2 = .19$. Among the covariates, robot experience was a significant positive predictor of self-investment, $F(1, 137) = 4.37$, $p = .038$, $\eta_p^2 = .03$, and IDAQ also significantly predicted self-investment, $F(1, 137) = 8.85$, $p = .003$, $\eta_p^2 = .06$.

Across all conditions, self-investment scores remained relatively low, though the high-involvement group reported the highest mean ($M = 3.10$, $SD = 1.31$).

Perceived Control. Across all conditions, perceived control ratings were generally low (see Table 5.1), suggesting a potential floor effect [61]. A one-way ANOVA revealed no significant main effect of condition on perceived control, $F(3, 139) = 1.88$, $p = .136$, $\eta^2 = .04$. However, a planned contrast comparing the control condition to all involvement conditions showed a significant difference, $t(139) = -2.31$, $p = .024$, $r = .19$, indicating that participants in the control group experienced significantly lower perceived control compared to those exposed to any level of involvement, as depicted in Figure 5.3b. No significant differences emerged between the low and medium/high involvement groups, $t(139) = -0.44$, $p = .662$, $r = .04$, nor between the medium and high conditions, $t(139) = -0.74$, $p = .464$, $r = .09$.

To account for individual tendency to anthropomorphise, an ANCOVA was conducted with IDAQ scores as a covariate. The effect of conditions remained non-significant, $F(3, 138) = 1.73$, $p = .164$, $\eta_p^2 = .04$, while IDAQ emerged as a marginal predictor of perceived control, $F(1, 138) = 3.65$, $p = .058$, $\eta_p^2 = .03$, with higher anthropomorphism tendencies slightly associated with increased perceived control ($= 0.24$).

Figure 5.3: Boxplots illustrating the effects of involvement condition (Control, Low, Medium, High) on (a) Self-Investment, (b) Perceived Control, and (c) Perceived Knowledge. Scores range from 1 to 7. Significant contrasts are marked with brackets (*$p<.05$, ***$p<.001$). Higher involvement levels were associated with greater self-investment and perceived knowledge, while perceived control was significantly lower in the control group compared to the others.



(a) **(b)** **(c)**

Perceived Knowledge. Perceived knowledge scores were generally low across all conditions (see Table 5.1). A one-way ANOVA revealed a marginal effect of condition on perceived knowledge, $F(3, 139) = 2.20$, $p = .091$, $\eta^2 = .05$. Planned contrasts showed that participants in the medium involvement condition reported significantly higher perceived knowledge than those in the low and control conditions, $t(139) = 2.26$, $p = .028$, $r = .21$. No significant difference was observed between the high involvement and other conditions, $t(139) = 0.95$, $p = .344$, $r = .08$, nor between the low and control groups, $t(139) = -0.10$, $p = .919$, $r = .01$, as shown in Figure 5.3c.

An ANCOVA was then conducted with robot experience and task workload entered as covariates. The effect of condition was no longer significant, $F(3, 137) = 1.99$, $p = .119$, $\eta_p^2 = .04$. Both covariates were significant predictors of perceived knowledge: robot experience, $F(1, 137) = 5.31$, $p = .023$, $\eta_p^2 = .04$, and task workload, $F(1, 137) = 6.91$, $p = .010$, $\eta_p^2 = .05$, where higher robot experience and greater task workload were associated with higher perceived knowledge.

Psychological Ownership. Psychological ownership scores were low across all conditions (see Table 5.1). A one-way ANOVA revealed no significant effect of condition on PO, $F(3, 139) = 0.08$, $p = .973$, $\eta^2 < .01$.

To ensure the robustness of the measure, an exploratory factor analysis was conducted. The analysis indicated that the three self-extension items demonstrated weak or inconsistent loadings (one loading below the conventional .40 threshold and another showing cross-loadings), whereas the three core PO items [71] loaded strongly (all $> .80$) on a single factor. A follow-up analysis using only the core PO items also revealed no significant effect of condition, $F(3, 139) = 0.60$, $p = .614$, $\eta^2 = .01$.

Given that prior robot experience was significantly correlated with the core PO measure, an ANCOVA was conducted with robot experience entered as a covariate. The effect of condition remained non-significant, $F(3, 138) = 0.64$, $p = .588$, but robot experience emerged as a significant predictor of psychological ownership, $F(1, 138) = 4.46$, $p = .037$, $\eta_p^2 = .03$, indicating that participants with more prior exposure to robots reported slightly higher levels of ownership.

Perceived Closeness. Perceived closeness to the robot was low across all conditions (see Table 5.1). A one-way ANOVA revealed no significant effect of condition on IOS scores, $F(3, 139) = 0.19$, $p = .902$, $\eta^2 < .01$. An ANCOVA was then conducted with robot experience, affinity for technology interaction, and IDAQ scores entered as covariates. The effect of condition remained non-significant, $F(3, 136) = 0.20$, $p = .895$, $\eta_p^2 = .004$. Affinity for technology emerged as a significant positive predictor of perceived closeness, $F(1, 136) = 8.46$, $p = .004$, $\eta_p^2 = .06$, and tendency to anthropomorphism was also significant, $F(1, 136) = 3.99$, $p = .048$, $\eta_p^2 = .03$, indicating that participants with higher affinity for technology and stronger tendency to anthropomorphise reported greater perceived closeness to the robot.

5.4.2 Effects of User Involvement on Anthropomorphisation

This analysis examined whether the level of user involvement in training the robot influenced the degree to which participants attributed human-like qualities to the robot, including autonomy, sociability, agency, animacy, and disturbance. Descriptive statistics for all variables across the four experimental conditions are summarised in Table 5.2. Inferential statistics are presented below.

Table 5.2: Descriptive statistics for Anthropomorphism dimensions across experimental conditions. Values are reported as mean (std).

Variable	High-Involvement	Medium-Involvement	Low-Involvement	Control
Perceived Autonomy	3.89 (1.20)	4.10 (1.25)	3.88 (1.38)	4.06 (1.14)
Perceived Sociability	4.83 (1.15)	4.84 (1.25)	4.66 (1.21)	5.07 (1.15)
Perceived Animacy	3.06 (1.01)	3.31 (1.17)	2.89 (0.97)	3.30 (1.00)
Perceived Disturbance	2.39 (1.09)	2.44 (1.35)	2.80 (1.41)	2.36 (1.33)

Perceived Autonomy. A one-way ANOVA revealed no significant effect of condition on perceived robot autonomy, $F(3, 139) = 0.30$, $p = .826$, $\eta^2 = .01$. Across all groups, perceived autonomy ratings remained relatively stable and centred around the scale's midpoint (Table 5.2), suggesting that the robot was generally viewed as moderately autonomous.

Perceived Sociability. No significant effect of condition was found on perceived sociability, $F(3, 139) = 0.73$, $p = .538$, $\eta^2 = .02$. Sociability ratings were consistently above average across conditions as seen in Table 5.2, indicating that the robot was broadly perceived as sociable, irrespective of involvement level. An ANCOVA was then conducted with affinity for technology interaction and need for closure entered as covariates. The effect of condition remained non-significant, $F(3, 137) = 1.03$, $p = .383$, $\eta_p^2 = .02$. Both covariates were significant predictors of perceived sociability: affinity for technology, $F(1, 137) = 16.82$, $p < .001$, $\eta_p^2 = .11$, and need for closure, $F(1, 137) = 20.03$, $p < .001$, $\eta_p^2 = .13$. Participants with higher affinity for technology and greater need for closure perceived the robot as more sociable.

Perceived Agency. There was no significant effect of condition on perceived agency, $F(3, 139) = 0.24$, $p = .871$, $\eta^2 = .01$. Mean ratings were consistently above the scale midpoint across all conditions (Table 5.2), indicating that the robot was generally viewed as moderately agentic. An ANCOVA was then conducted with tendency to antropomorphise and need for cognition entered as covariates. The effect of condition remained non-significant, $F(3, 137) = 0.85$, $p = .468$, $\eta_p^2 = .02$. IDAQ showed a marginal effect on percieved agency, $F(1, 137) = 3.39$, $p = .068$, $\eta_p^2 = .02$, suggesting that participants with stronger tendency to anthropomorphise tended to attribute slightly higher agency to the robot. Need for cognition was not a significant predictor, $F(1, 137) = 0.46$, $p = .499$, $\eta_p^2 = .003$.

Perceived Animacy. A one-way ANOVA showed no significant effect of condition on animacy attribution, $F(3, 139) = 1.35$, $p = .260$, $\eta^2 = .03$. Ratings were generally low to moderate across conditions (Table 5.2), suggesting that the robot was not strongly perceived as animate. An ANCOVA including IDAQ as a covariate revealed a significant main effect of tendency to anthropomorphise on perceived animacy, $F(1, 138) = 7.62$, $p = .007$, $\eta_p^2 = .05$, with higher IDAQ scores associated with increased animacy ratings ($\beta = 0.29$). The effect of condition remained non-significant when controlling for this covariate, $F(3, 138) = 1.62$, $p = .187$.

Perceived Disturbance. Conditions had no significant effect on perceived disturbance, $F(3, 139) = 0.89$, $p = .446$, $\eta^2 = .02$. Ratings were consistently below the midpoint across conditions (see Table 5.2), indicating that the robot was generally not perceived as disturbing. An ANCOVA was then conducted with robot experience and affinity for technology interaction entered as covariates. The effect of condition remained non-significant, $F(3, 137) = 0.87$, $p = .458$, $\eta_p^2 = .02$. Neither covariate was a strong predictor of perceived disturbance: robot experience was not significant, $F(1, 137) = 1.64$, $p = .202$, $\eta_p^2 = .01$, and affinity to technology showed only a marginal effect on percieved disturbance, $F(1, 137) = 3.46$, $p = .065$, $\eta_p^2 = .03$, with higher scores associated with slightly lower perceived disturbance.

5.4.3 Effect of Involvement on Trust toward Robots

This section investigates whether the level of user involvement in teaching the robot influenced different forms of trust, including cognitive, affective, and epistemic trust. Descriptive statistics for all variables across the four experimental conditions are summarised in Table 5.3. Inferential statistics are presented below.

Cognitive Trust. A one-way ANOVA revealed no significant effect of conditions on cognitive trust, $F(3, 139) = 0.71$, $p = .551$, $\eta^2 = .01$. Across all conditions,

Table 5.3: Descriptive statistics for Trust dimensions across experimental conditions. Values are reported as mean (std).

Variable	High-Involvement	Medium-Involvement	Low-Involvement	Control
Cognitive Trust	4.06 (0.91)	3.95 (0.78)	4.02 (0.88)	4.24 (0.87)
Affective Trust	4.44 (1.10)	4.39 (1.10)	4.11 (1.00)	4.57 (1.17)
Epistemic Trust (shape)	5.61 (1.46)	5.57 (1.26)	5.33 (1.53)	5.91 (1.29)
Epistemic Trust (colour)	3.93 (1.45)	4.18 (1.39)	4.32 (1.11)	4.42 (1.24)

cognitive trust ratings were generally above the midpoint of the scale (see Table 5.3), suggesting that participants perceived the robot as moderately competent regardless of their level of involvement. An ANCOVA was then conducted with tendency to anthropomorphise and need for closure entered as covariates. The effect of conditions remained non-significant, $F(3, 137) = 1.00$, $p = .397$, $\eta_p^2 = .02$. Both covariates were significant predictors of cognitive trust: IDAQ, $F(1, 137) = 6.30$, $p = .013$, $\eta_p^2 = .04$, and need for closure, $F(1, 137) = 11.75$, $p < .001$, $\eta_p^2 = .08$. Participants with stronger anthropomorphism tendencies and higher need for closure reported higher levels of cognitive trust in the robot.

Affective Trust. A one-way ANOVA showed no significant effect of condition on affective trust, $F(3, 139) = 1.17$, $p = .322$, $\eta^2 = .02$. Ratings were generally above the midpoint across conditions (see Table 5.3), indicating that participants tended to perceive the robot as emotionally engaging, regardless of their level of involvement. An ANCOVA was then conducted with affinity for technology interaction, IDAQ, and need for closure entered as covariates. The effect of conditions remained non-significant, $F(3, 136) = 1.56$, $p = .201$, $\eta_p^2 = .03$. However, all three covariates significantly predicted affective trust: affinity for technology, $F(1, 136) = 6.49$, $p = .012$, $\eta_p^2 = .05$; IDAQ, $F(1, 136) = 6.69$, $p = .011$, $\eta_p^2 = .05$; and need for closure, $F(1, 136) = 7.55$, $p = .007$, $\eta_p^2 = .05$. Participants with higher affinity for technology, stronger tendency to anthropomorphise, and greater need for closure reported higher levels of affective trust in the robot.

Epistemic Trust (Robot's Shape Suggestion). A one-way ANOVA found no significant effect of condition on participants' confidence in the robot's shape suggestion, $F(3, 139) = 1.09$, $p = .356$, $\eta^2 = .02$. Across all conditions, epistemic trust ratings were generally high (see Table 5.3), indicating that participants tended to express confidence in the robot's shape suggestions regardless of their level of involvement. An ANCOVA was then conducted with task performance and mental workload entered as covariates. The effect of condition remained non-significant, $F(3, 137) = 1.21$, $p = .309$, $\eta_p^2 = .03$. Performance was a strong positive predictor of epistemic trust towards shapes, $F(1, 137) = 15.17$, $p < .001$, $\eta_p^2 = .10$, such that participants with higher task performance reported greater confidence in the robot's shape suggestions. Mental workload showed a marginal negative effect, $F(1, 137) = 2.85$, $p = .094$, $\eta_p^2 = .02$, indicating that participants who experienced higher workload tended to express slightly lower epistemic trust.

Epistemic Trust (Robot's Colour Suggestion). A one-way ANOVA showed no significant effect of condition on participants' confidence in the robot's colour suggestion, $F(3, 139) = 0.91$, $p = .437$, $\eta^2 = .02$.

5.4.4 Relationships Between User Involvement, Ownership, and Trust

To examine how user involvement impacts psychological ownership and how ownership and trust relate to epistemic trust, a series of mediation and regression analyses were conducted.

Mediation of User Involvement on Psychological Ownership. Three mediation models tested whether perceived self-investment, control, and knowledge mediated the relationship between perceived user involvement and psychological ownership.

User involvement significantly predicted self-investment ($\beta = 0.55$, $p < .001$, 95% CI [0.42, 0.68]), and self-investment in turn significantly predicted psychological ownership ($\beta = 0.10$, $p = .002$, 95% CI [0.04, 0.17]). The total effect of user involvement on psychological ownership was also significant ($\beta = 0.10$, $p = .002$, 95% CI [0.04, 0.16]). When examining the direct pathway, that is, the effect of involvement on ownership after controlling for self-investment, the effect was not significant ($\beta = 0.06$, $p = .111$, 95% CI [–0.01, 0.13]). The indirect pathway, representing the effect of involvement on ownership through increased self-investment, was also not significant ($\beta = 0.04$, $p = .196$, 95% CI [–0.02, 0.11]).

User involvement significantly predicted perceived control ($\beta = 0.29$, $p < .001$, 95% CI [0.15, 0.43]). However, perceived control did not significantly predict psychological ownership when controlling for involvement ($\beta = 0.07$, $p = .052$, 95% CI [–0.00, 0.14]). The total effect of involvement on psychological ownership was significant ($\beta = 0.10$, $p = .002$, 95% CI [0.04, 0.16]), and the direct effect also remained significant after accounting for perceived control ($\beta = 0.09$, $p = .009$, 95% CI [0.02, 0.15]). The indirect pathway through perceived control was not significant ($\beta = 0.01$, $p = .336$, 95% CI [–0.01, 0.04]), indicating that perceived control did not mediate the relationship between involvement and ownership.

In contrast, for perceived knowledge, mediation was supported. User involvement significantly predicted perceived knowledge ($\beta = 0.22$, $p < .001$, 95% CI [0.10, 0.35]), and perceived knowledge in turn significantly predicted psychological ownership ($\beta = 0.18$, $p < .001$, 95% CI [0.10, 0.26]). The total effect of involvement on psychological ownership was significant ($\beta = 0.10$, $p = .002$, 95% CI [0.04, 0.16]). The direct effect of involvement remained significant when controlling for perceived knowledge ($\beta = 0.06$, $p = .045$, 95% CI [0.00, 0.12]), but was reduced in magnitude. Importantly, the indirect effect of involvement on ownership through perceived knowledge was significant ($\beta = 0.04$, $p = .004$, 95% CI [0.01, 0.08]), indicating that perceived knowledge partially mediated the relationship between involvement and psychological ownership.

Psychological Ownership as Predictor of Trust. Psychological ownership significantly predicted both cognitive and affective trust. For cognitive trust, the regression model was significant, $F(1, 141) = 16.85$, $p < .001$, with psychological ownership positively predicting trust, $\beta = 0.52$, 95% CI [0.27, 0.76], $R^2 = .11$. For affective trust, psychological ownership also had a significant effect, $F(1, 141) = 8.99$, $p = .003$, $\beta = 0.49$, 95% CI [0.17, 0.81], $R^2 = .06$.

Anthropomorphism as Mediator Between Autonomy and Trust. Mediation analyses tested whether anthropomorphism mediated the relationship between perceived autonomy and trust.

Perceived autonomy significantly predicted anthropomorphism ($\beta = 0.23$, $p < .001$, 95% CI [0.15, 0.31]), and anthropomorphism in turn strongly predicted cognitive trust ($\beta = 0.95$, $p < .001$, 95% CI [0.80, 1.10]). The total effect of autonomy on cognitive trust was significant ($\beta = 0.29$, $p < .001$, 95% CI [0.19, 0.40]). When anthropomorphism was included in the model, the direct effect of autonomy on cognitive trust was reduced but remained significant ($\beta = 0.09$, $p = .043$, 95% CI [0.00, 0.18]). The indirect effect through anthropomorphism was also significant ($\beta = 0.20$, $p < .001$, 95% CI [0.12, 0.30]), indicating partial mediation: autonomy increased perceptions of anthropomorphism, which in turn enhanced cognitive trust.

Concerning affective trust, perceived autonomy significantly predicted anthropomorphism ($\beta = 0.23$, $p < .001$, 95% CI [0.15, 0.31]), and anthropomorphism strongly predicted affective trust ($\beta = 1.10$, $p < .001$, 95% CI [0.90, 1.31]). The total effect of autonomy on affective trust was significant ($\beta = 0.33$, $p < .001$, 95% CI [0.19, 0.47]). However, when anthropomorphism was included in the model, the direct effect of autonomy on affective trust was no longer significant ($\beta = 0.09$, $p = .124$, 95% CI [–0.03, 0.22]). The indirect effect through anthropomorphism was significant ($\beta = 0.23$, $p < .001$, 95% CI [0.14, 0.35]), indicating full mediation: the effect of autonomy on affective trust was carried entirely by anthropomorphism.

Cognitive and Affective Trust as Predictors of Epistemic Trust. A multiple regression was conducted to examine whether cognitive and affective trust predicted epistemic trust, measured through participants' confidence in the robot's shape suggestions. The overall model was significant, $F(2, 140) = 3.38$, $p = .037$. Affective trust was a significant positive predictor ($\beta = 0.39$, $p = .019$), while cognitive trust was not ($\beta = -0.24$, $p = .256$).

A second regression examined whether cognitive and affective trust predicted confidence in the robot's colour suggestions. The model was not statistically significant, F(2, 140) = 1.98, p = .142. Neither cognitive trust ($\beta = 0.25$, p = .214) nor affective trust ($\beta = 0.00$, p = .983) were significant predictors.

A third model tested whether cognitive and affective trust predicted behavioural conformation to the robot's shape suggestions. The model was significant, F(2, 140) = 3.70, p = .027, explaining 5.0% of the variance ($R^2 = .050$). Cognitive trust significantly predicted conformation (β= 0.033, p = .018), while affective trust was not a significant predictor (= –0.011, p = .301).

A final regression assessed whether cognitive and affective trust predicted behavioural conformation to the robot's colour suggestions. The model was not statistically significant, F(2, 140) = 0.94, p = .393, explaining 1.3% of the variance ($R^2 = .013$). Neither cognitive trust (β= 0.025, p = .357) nor affective trust $\beta =$ –0.001, p = .946) significantly predicted conformation in this context.

5.5 Discussion

5.5.1 User Involvement Promotes Self-Investment but Fails to Elicit Psychological Ownership

This study hypothesised that greater involvement in teaching the robot would enhance users' sense of PO and its proposed determinants: self-investment, perceived control, perceived knowledge, and closeness. The findings offered only limited support for this hypothesis. Specifically, involvement significantly increased self-investment, but no consistent effects were observed for perceived control, perceived knowledge, or closeness.

Self-investment emerged as the most sensitive indicator to the manipulation of involvement. Participants in the high-involvement condition reported significantly greater self-investment than those in all other conditions. This finding aligns with psychological ownership theory, which posits that dedicating more time and effort to a target fosters a stronger sense of investment in it [211].

However, despite this statistical effect, absolute self-investment scores remained relatively low across all conditions, indicating that participants only experienced a limited sense of investment in the robot. Individual differences, particularly anthropomorphism tendencies and prior experience with robots, also influenced self-investment. This suggests that self-investment is shaped by both situational factors (such as the level of involvement) and dispositional factors (such as individual tendencies), rather than reflecting a general state of engagement.

In contrast, perceived control was not significantly impacted by condition overall. However, a planned contrast revealed that participants in the control condition reported significantly lower control than those in any of the involvement conditions. This suggests that even minimal involvement, or simply being informed about the robot's learning process, can bolster users' sense of control. Still, this effect did not scale linearly with involvement level, and control ratings were generally low, potentially indicating a floor effect.

Perceived knowledge showed a marginal effect of condition, with high ratings in both the high- and medium-involvement conditions. Although only the contrast comparing medium involvement to the low and control groups reached statistical significance, the high-involvement group also reported relatively high levels. These findings suggest that user involvement, regardless of intensity, may foster a form of metacognitive confidence. Indeed, perceived knowledge can be understood as a metacognitive judgment, reflecting how individuals monitor and evaluate what they or others know [91, 81, 148]. In the present context, being actively involved in teaching the robot may have encouraged participants to monitor and reflect more closely on the robot's learning process, thereby enhancing their perceived knowledge. Additionally, perceived knowledge was positively predicted by both mental workload and prior robot experience, indicating that greater cognitive effort and familiarity with robots jointly shaped how much knowledge participants attributed to the robot.

Mediation analyses further highlighted the role of perceived knowledge: while involvement significantly predicted self-investment, control, and knowledge, only perceived knowledge emerged as a significant mediator of psychological ownership, with a significant indirect effect ($\beta = 0.04$, $p = .004$) alongside a direct effect. This pattern contrasts with prior literature on psychological ownership, where perceived control and self-investment are typically the most consistent predictors, while intimate knowledge plays a more variable role [186, 205]. In the present study, however, the task explicitly involved teaching, which may have heightened participants' focus on knowledge as a salient dimension, thereby making perceived knowledge a stronger route to ownership than is usually reported.

Despite initial expectations, perceived closeness to the robot was not significantly influenced by involvement condition. Ratings were low across all groups, suggesting limited social bonding. However, closeness was significantly predicted by participants' affinity for technology, and marginally by anthropomorphism tendencies, indicating that trait-level factors may play a larger role in shaping perceived closeness than short-term task manipulations. One possible explanation is that individuals who identify as technology enthusiasts may more readily form self–target associations with robots, which facilitates feelings of closeness. Similarly, people who are more prone to anthropomorphise are likely to attribute self-centric knowledge and qualities to the robot, thereby fostering a stronger sense of relational closeness [83].

Similarly, psychological ownership was unaffected by the involvement manipulation, with uniformly low scores across conditions. Exploratory factor analysis indicated that the three "self-extension" items, often used to measure PO, showed weak performance in this context, with one item loading below the .40 threshold and another exhibiting cross-loadings. In contrast, the three core PO items loaded strongly onto a single factor and demonstrated greater construct validity. However, a follow-up analysis using only these core items also revealed no significant effect of condition.

These findings suggest that while self-investment and perceived knowledge were somewhat sensitive to involvement, the broader construct of psychological ownership was not. Prior work highlights perceived control as a central antecedent of psychological ownership [297, 186], and several studies on customisation similarly show that involvement enhances ownership primarily by increasing users' sense of control [157]. From this perspective, the present results run counter to theoretical expectations: despite higher involvement, participants did not report greater perceived control. A likely explanation is that the manipulation of involvement did not provide participants with genuine decision power. Even in the high-involvement condition, participants mainly followed instructions to teach the robot the “correct” way, rather than making autonomous choices about the robot’s learning. Without meaningful choice, the sense of control required for ownership to emerge may have been limited.

Another likely factor is personal relevance, defined as the extent to which an artefact resonates with a user’s goals, values, or sense of self [217]. Higher personal relevance can increase the likelihood that users experience the artefact as “theirs,” thereby strengthening psychological ownership. Although underexplored in HRI, initial findings suggest that personal relevance may influence ownership [205]. In this study, however, the robot was not presented in a way that connected to participants’ personal goals, which may have limited the emergence of PO regardless of involvement.

In sum, while user involvement influenced some proximal constructs, particularly self-investment and perceived knowledge, these effects did not robustly translate into stronger psychological ownership. This suggests that involvement alone may be insufficient: to elicit meaningful PO in HRI, it may be necessary to provide users with choice, foster personal relevance, or extend the duration and depth of the interaction.

5.5.2 User Involvement Does Not Affect Anthropomorphism, but Individual Traits Do

This study hypothesised that greater involvement in training the robot would lead to increased anthropomorphism, based on the assumption that active engagement fosters social presence and human-like attribution [83, 286]. Specifically, it was expected that participants who taught or observed the robot’s learning process would be more likely to perceive it as autonomous, sociable, agentic, and animate, and less likely to view it as disturbing.

Despite the manipulation, anthropomorphic attributions remained stable across conditions, with the robot generally perceived as moderately sociable and agentic, but low in animacy and disturbance. It is possible that the short duration or the fact that participants only taught the robot shapes, rather than engaging in a socially or personally meaningful activity, limited the emotional engagement needed to alter anthropomorphic perceptions.

However, individual differences emerged as stronger and more consistent predictors. The tendency to anthropomorphise significantly predicted agency and animacy, and showed marginal effects for sociability, suggesting that individuals who are more prone to attributing human-like traits to non-human agents also perceived the robot as more human-like. This aligns with theoretical accounts indicating that anthropomorphism reflects a stable disposition that varies across individuals and persists across contexts [285, 151, 242].

The lack of relationship between tendency to anthropomorphise and disturbance is particularly noteworthy. One might expect that a tendency to anthropomorphise an entity would reduce feelings of unease, as familiarity and human-likeness are generally associated with greater comfort. However, disturbance is thought to reflect the uncanny valley [17]; a negative affective response that occurs when an agent is neither fully mechanical nor convincingly human. In this case, NAO's appearance may have been too far from the human-likeness threshold to trigger such discomfort. This could explain why disturbance ratings were low overall and unaffected by anthropomorphism tendencies.

In addition to the tendency to anthropomorphise, other dispositional traits predicted anthropomorphic perception. Affinity for technology was linked to higher sociability and lower disturbance, while need for closure was positively associated with sociability. Taken together, the results point to user traits as the primary drivers of anthropomorphism, outweighing the effects of short-term involvement with the system.

In addition to the tendency to anthropomorphise, other dispositional traits predicted anthropomorphic perception. Affinity for technology was linked to higher sociability and lower disturbance, while need for closure was positively associated with sociability. This is consistent with the idea that individuals with a higher need for closure,who prefer intuitive judgments and quick resolution, may be more likely to anthropomorphise, using it as a heuristic or cognitive shortcut to make sense of the robot as an entity [193]. Taken together, these findings point to user traits as the primary drivers of anthropomorphism, outweighing the effects of short-term involvement with the system.

Sociability and agency ratings were consistently above average, indicating that the robot was generally perceived as a capable and socially present partner, even though these aspects were not directly manipulated in the study. In contrast, animacy ratings remained low, underscoring the difficulty of eliciting life-like perception through task interaction alone. Enhancing perceptions of animacy may require more expressive motion, naturalistic behaviour, or emotionally congruent responses.

Overall, these findings indicate that anthropomorphic attributions are not easily modulated through short-term involvement alone. Instead, they appear to be anchored in individual predispositions, influenced by technology attitudes, cognitive styles, and prior experience.

5.5.3 User Involvement Does Not Influence Trust, but Dispositional and Perceptual Factors Do

This study hypothesised that greater involvement in teaching the robot would enhance participants' trust across cognitive, affective, and epistemic dimensions. However, the results provide no evidence that the level of involvement significantly influenced any form of trust. Across all conditions, trust ratings remained relatively stable. Cognitive and affective trust scores were consistently above the midpoint, while epistemic trust was not significantly affected by involvement level. These findings suggest that, in the context of a short, one-shot interaction, trust in the robot was not sensitive to differences in how participants engaged with its learning process.

Instead, trust was more strongly associated with individual traits and perceptions of the robot. For both cognitive and affective trust, anthropomorphism tendencies and need for closure emerged as significant predictors. Higher NFC scores predicted greater trust, consistent with the idea that individuals who prefer certainty may be more inclined to adopt trust as a heuristic to reduce ambiguity and achieve closure [153, 229]. This may reflect the motivational nature of NFC in reducing ambiguity: trust, as conceptualised by Möllering [184], involves a "leap of faith" that bridges the gap between known cues and unknown outcomes. Individuals high in NFC may be more willing to make that leap in order to resolve uncertainty and establish a stable interpretation of the robot's trustworthiness. Similarly, participants who more readily anthropomorphised the robot were more likely to ascribe competence and emotional warmth to it. This positive relationship between anthropomorphism and trust is consistent with a broader body of research highlighting the role of human-likeness in enhancing perceived reliability and rapport in HRI [28, 78, 226, 294].

Affinity for technology also contributed to affective trust, though its role was less pronounced. However, affinity for technology did not significantly predict cognitive trust in this study. Individuals with higher affinity for technology are more likely to approach new systems with enthusiasm and openness, which can foster more positive attitudes toward the robot and greater ease in interacting with it [227, 13]. This dynamic tends to enhance affective trust, without necessarily influencing cognitive trust. In contrast, cognitive trust tends to be shaped more by task performance, explicit expertise cues, or perceptions of system functionality. This distinction aligns with emerging findings in HRI research. For example, Cantucci et al. [41] found that robot personality traits affected affective trust but did not significantly impact cognitive trust or trust-based behaviour.

Additionally, a significant relationship was observed between psychological ownership and trust, with participants who felt a stronger sense of ownership over the robot reporting higher levels of both cognitive and affective trust. This finding aligns with prior research suggesting that a sense of personal investment or identification with a system can foster trust [71, 66, 155]. Moreover, perceived autonomy influenced trust indirectly through anthropomorphism, which mediated the relationship for both cognitive and affective dimensions. While autonomy alone did not strongly predict trust, evidence suggests its effect may operate through perceived human-likeness or social presence. For example, autonomy has been shown to moderate the translation of competence into trust in robot contexts [41]. Robotic mind attribution / human-likeness has also been linked to increased trust [65], and social presence is known to foster trust in human-machine interactions by reducing uncertainty [194, 239].

In the case of epistemic trust, participants' confidence in the robot's shape suggestions was not significantly affected by involvement condition. However, two psychological factors: self-confidence and mental workload strongly predicted confidence in the robot's shape suggestions.

Participants with greater confidence in their own reasoning were also more confident in the robot's responses. This pattern may reflect social projection, the tendency to assume others think or perform similarly to oneself [225, 234]. Projection effects are typically stronger toward ingroup members, and in this study the robot was explicitly framed as a peer, which may have encouraged participants to treat it as part of their ingroup. However, because the tendency to anthropomorphise did not predict this effect, and anthropomorphism is often a prerequisite for perceiving a robot as a peer or ingroup member, this interpretation remains tentative and requires further confirmation.

Moreover, higher perceived workload was associated with lower epistemic trust. This finding echoes prior work showing an inverse relationship between cognitive load and trust in robots and automation. For instance, Ahmad et al. [1] observed that participants engaged in a collaborative game with robots reported declining trust as their cognitive load increased, indicating that mental demands limit the attentional resources available to sustain confidence in a robot's input. This interpretation is consistent with cognitive engineering accounts, which conceptualise workload, trust, and situation awareness as interdependent constructs. Parasuraman et al. [200] argue that high workload reduces the attentional capacity required to monitor and evaluate automated agents, thereby complicating the calibration of trust.

Importantly, cognitive and affective trust predicted epistemic trust differently across contexts. For shape suggestions, affective trust emerged as the only significant predictor of reported confidence in the robot's answers, while cognitive trust predicted behavioural conformation, suggesting that while participants may feel confident due to emotional connection, actual behavioural deference is driven more by competence-related beliefs. In contrast, neither form of trust predicted epistemic trust for colour suggestions, likely reflecting participants' detection of the robot's consistent inaccuracy in that domain. This divergence between reported confidence and behavioural adaptation highlights the multi-layered nature of epistemic trust and the importance of assessing both self-report and behavioural data.

Finally, it is important to acknowledge a potential limitation related to the way epistemic trust was measured in this study. Because the aim was to examine how participants' involvement in teaching a robot influenced their trust, the task used to assess epistemic trust needed to directly reflect the concepts that participants had taught the robot. This constraint ensured a meaningful link between the learning interaction and the subsequent evaluation. However, it also meant that the epistemic trust task was based on a simple classification activity, which may not fully capture the complexity of epistemic trust as defined in the cognitive sciences. Future work could draw on more established paradigms, such as Wason selection task [283], which probe how individuals balance their own reasoning with external input, to provide a richer assessment of epistemic trust in HRI.

Overall, these findings suggest that user involvement in robot learning may not be sufficient to affect trust in the short term, especially when baseline trust levels are already moderate to high. Instead, trust was more strongly shaped by stable individual differences and inferred characteristics of the robot, such as competence, familiarity, and autonomy. Future work should explore whether longer-term interactions, collaborative learning, or higher-stakes decision contexts might yield stronger effects of involvement on trust calibration.

5.6 Summary

This chapter examined how user involvement in a robot's learning process influences perceptions of anthropomorphism, PO, and different types of trust. Contrary to expectations, increased involvement did not significantly affect anthropomorphic attributions or trust levels. Instead, both anthropomorphism and trust were better explained by individual traits such as the tendency to anthropomorphise, need for closure, and affinity for technology.

Epistemic trust was similarly unaffected by involvement but was shaped by participants' cognitive state and disposition, with mental workload and self-confidence playing key roles. Furthermore, the relationship between trust types and behavioural outcomes was context-dependent: affective trust predicted confidence in the robot's suggestions, while cognitive trust predicted behavioural conformation, but only when the robot's performance was reliable.

While these results suggest that short-term involvement may not be sufficient to modulate trust or anthropomorphism, they underscore the importance of tailoring educational interactions with robots to individual learner profiles. Designing robot systems that can detect and adapt to user traits may help foster more effective and trustworthy learning environments. Future research should investigate whether more prolonged, emotionally engaging, or high-stakes educational contexts lead to stronger effects of involvement on trust and learning outcomes.

Chapter 6

Teaching Robot through Natural Language

[Epigraph removed for copyright reasons. Original source: A.A. Milne, Winnie-the-Pooh *(London: Methuen, 1926)]*

6.1 Introduction

While earlier chapters explored how children can engage in teaching interactions with social robots using evaluative feedback within an Interactive RL framework, these interactions were primarily built around a single mode of communication. In contrast, human teaching, particularly in child-led contexts, is inherently multimodal, involving not just evaluation, but also correction and forward-looking guidance. To bridge this gap, this chapter presents a novel framework that enables robots to interpret and learn from multiple types of natural language advice using large language models. The framework is evaluated in an HRI study comparing it to a baseline system limited to evaluative input. The chapter explores whether this multimodal approach improves learning efficiency, reduces teaching effort, and supports more natural and personalised interactions, advancing toward the broader goal of developing peer-like educational robots that adapt to diverse teaching styles.

6.2 Study Motivation and Research Questions

The previous chapters showed that children can successfully teach robots through Interactive RL using evaluative feedback. However, this architecture was limited to processing a single input type, restricting the range of natural teaching behaviours that children typically employ.

In practice, human teaching is rarely limited to one channel. Children and adults alike tend to mix different kinds of feedback, shifting fluidly between evaluation, correction, and guidance depending on the situation. These strategies emerge early and adapt to the learner's needs and context. Foundational work by Wood et al. [293] demonstrated that adult tutors scaffold learning by flexibly adjusting their instructional strategies to suit the child's developmental stage, employing more guidance with younger children and corrective or evaluative strategies with older ones. Further studies have shown that even preschoolers are capable of providing targeted feedback: by age four, children begin selecting informative evidence to guide a learner's understanding, including examples that help revise incorrect hypotheses [19], and by age five, they begin to offer specific corrective input [231]. In peer tutoring, older children adapt their teaching depending on the learner's engagement, offering clarification and guidance when needed [118]. Such behaviours have also been observed in human-robot contexts. For example, it has been shown that children use different strategies when teaching a robot versus a peer, offering more coaching and responding to the robot's questions and feedback [201].

Robots have likewise been shown to benefit from diverse forms of teaching signals. In Interactive RL, evaluative feedback, typically in the form of scalar or binary assessments of past actions, has been used to shape learning in the absence of handcrafted reward functions [142]. Corrective feedback, where the appropriate action is specified by the human, has been found to reduce exploration space and accelerate convergence [44]. Guidance, in the form of anticipatory advice, has been shown to be effective in steering learning, particularly in the early stages of exploration [266]. Despite these individual successes, current Interactive RL systems are typically limited to processing a single type of feedback, lacking the flexibility to accommodate the multimodal instructional styles naturally used by human teachers. For example, it has been shown that both evaluative and corrective feedback improve learning, yet these modalities were implemented through distinct mechanisms and not integrated within a single framework [189]. Furthermore, a recent survey has highlighted that most reinforcement learning from human feedback frameworks remain narrowly scoped, focusing on scalar rewards, preferences, or demonstrations, without generalising to more naturalistic and flexible forms of input [133]. This fragmentation limits the capacity of current systems to support interaction protocols that reflect the dynamics of real-world teaching [50].

Recently, efforts have been made to enhance RL by leveraging Large Language Models (LLMs). Prior work has explored using LLMs to generate reward signals through fine-tuning on user data [197, 14], guiding intermediate learning tasks with linguistic instructions [102, 42], or formulating goal-conditioned rewards based on semantic similarity between generated captions and desired goals [77]. While these approaches demonstrate the potential of LLMs to enrich reinforcement learning, they focus predominantly on offline or synthetic settings. To date, no studies have investigated the use of LLMs for real-time interpretation of human advice in Interactive RL contexts.

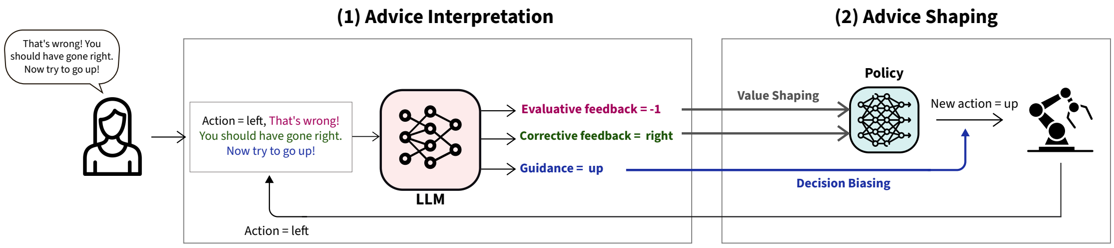
To address this gap, the ECLAIR framework (Evaluative Corrective Guidance Language as Reinforcement) has been developed. ECLAIR leverages LLMs to interpret free-form natural language feedback and classify it into structured categories: evaluative, corrective, or guidance. These signals are then used to shape the robot's behaviour through tailored integration mechanisms. By supporting a unified model for processing diverse feedback types, ECLAIR enables more natural, flexible, and cognitively aligned teaching interactions.

The study is guided by the following research questions:

- **RQ4.1:** Does ECLAIR improve learning from human advice compared to baseline approaches?

- **RQ4.2:** Is ECLAIR more feedback-efficient, requiring fewer teaching signals to achieve comparable performance?

Figure 6.1: Overview of the ECLAIR framework. Human natural language feedback is processed using a large language model, which classifies the feedback into one or more categories: evaluative, corrective, or guidance. Each type is then integrated into the learning process using a dedicated shaping strategy.



- **RQ4.3:** How sensitive is ECLAIR to variations in prompt structure, and how do such variations affect advice interpretation?

Based on these questions, it was expected that (i) ECLAIR would lead to faster convergence and higher task performance than single-modality Interactive RL, (ii) fewer teaching signals would be needed to reach equivalent performance levels, and (iii) performance would remain robust across prompt variations, with minimal impact on classification accuracy or downstream learning.

6.3 The ECLAIR Framework

This section details the implementation of the ECLAIR framework, which enables reinforcement learning from diverse forms of natural language feedback using LLMs and tailored shaping strategies. Specifically, the framework operates in two stages: (1) the interpretation of human advice using a language model, and (2) the integration of this advice into the agent's policy via shaping techniques. An overview of the framework is provided in Figure 6.1.

6.3.1 Advice Interpretation

In the first phase of the framework, human-provided natural language feedback is interpreted into a structured format that the robot can use. A pre-trained LLM is prompted to serve as a feedback classifier. The model receives a structured prompt that includes task information, the robot's previous action, and the verbal feedback provided by the user. It is asked to output the interpreted advice in terms of three possible feedback types:

- **Evaluative feedback** assesses whether the robot's last action was correct or incorrect.
- **Corrective feedback** specifies what the robot should have done instead.
- **Guidance** suggests what the robot should do in the next step.

The LLM is guided using a detailed system prompt that defines the task's structure (e.g., goal, action space) and the meaning of each feedback type. Few-shot examples are included to improve interpretation accuracy. These examples show how naturalistic feedback maps onto structured outputs, ensuring consistent and semantically accurate classification. If the provided advice does not include one or more of the feedback categories, the model is instructed to return `None` for that category. The prompt used is provided in Appendix C.

Once the LLM generates a structured output, a parser extracts the interpreted information and translates it into numerical values for downstream use. For instance, evaluative feedback is converted into a binary reward signal, corrective feedback into an alternative action label, and guidance into an action suggestion. If no feedback is provided, the system proceeds without update.

This interpretation process is designed to operate in real time after the robot takes an action and receives a response from the user, supporting continuous interaction without rigid input constraints.

6.3.2 Advice Shaping

Following interpretation, the structured feedback is integrated into the robot's learning algorithm through different shaping strategies, depending on the type of advice received.

Evaluative Feedback. Previous studies have shown that evaluative feedback is most effectively used as immediate information about the value of an action, positioning it as an ideal candidate for directly modifying the action-value function within the learning algorithm [116]. Similar to Knox et al.[143], the Q-augmentation method is adopted to integrate the evaluative feedback, denoted as h, into the learning process. For each observed state-action pair $< s, a >$, the Q-value is updated as follows:

$$Q'(s, a) = Q(s, a) + \alpha(h - Q(s, a)) \tag{6.1}$$

where h is the human-provided evaluation and α is the learning rate, set empirically.

Corrective Feedback. When the robot selects an incorrect action, the corrective feedback specifies the optimal action, denoted as a_c, that would have been more appropriate. For its integration into the learning process, a methodology paralleling Q-augmentation is used; instead of altering the action-value of the observed state-action pair $< s, a >$, the Q-value for the state and the optimal action $< s, a_c >$ is updated. This method of adjustment is essentially equivalent to providing positive feedback for the corrected action a_c, encouraging its selection in the future. The update is formulated as follows:

$$Q'(s, a_c) = Q(s, a_c) + \alpha(1 - Q(s, a_c)) \tag{6.2}$$

Guidance Feedback. Guidance is provided to steer the exploration strategy toward the desired action of the user. Unlike evaluative and corrective feedback, which influence the learning algorithm's value estimations, guidance feedback directly modifies the policy's output at decision time. To incorporate this feedback into the learning process, the decision-biasing strategy [189] is considered, where the robot prioritises the action suggested by the user, denoted as a_g, at decision time. Specifically, the policy is adjusted such that for the next state s_{t+1}, the chosen action is explicitly set to a_g:

$$\pi(s_{t+1}) = a_g \tag{6.3}$$

This override is applied independently of the robot's exploration mechanism, ensuring that human guidance can take priority when available.

By combining these shaping strategies under a single framework, ECLAIR supports a more human-like and flexible teaching process. Rather than constraining users to a single type of feedback or predefined interface, the framework allows teachers to provide feedback naturally, using the modality that best suits the interaction, whether it be evaluation, correction, or direction.

6.4 Methodology

6.4.1 Participants

Twelve adult participants were recruited for the study ($M_{age} = 30$; 10 men, 2 women). The sample included a mix of university students and staff members with varying levels of prior experience with robots. Participants were not required to have any specific technical background, and no exclusion criteria were applied beyond age (18+) and fluency in English.

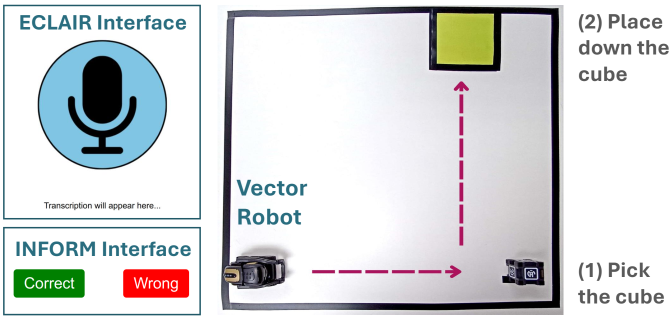
6.4.2 Study Design

To evaluate the performance of the ECLAIR framework, a human-robot interaction study was conducted in which participants were asked to teach a robot a simple control task. The experimental setup is illustrated in Figure 6.2.

Robotics Platform. The study employed *Vector*,[1] a compact and commercially available social robot. Owing to its small size, Vector is well-suited for tabletop interactions. The robot is equipped with basic navigation and manipulation capabilities, including omnidirectional wheels for movement and a lift mechanism for physical engagement with object, such as picking up or placing its interactive cube.

[1]https://ddlbots.com/products/vector-robot

Figure 6.2: Experimental setup: Vector robot, *pick and place* task and screenshots from the interfaces.



System Configuration. The experimental system consisted of three main components: the robot, an Android tablet, and a Linux-based machine running Ubuntu. The Android tablet served as the user interface through which participants could provide natural language feedback. The Ubuntu machine hosted the reinforcement learning process and coordinated communication between the robot and tablet via the Robot Operating System (ROS).

LLMs and Speech Recognition. To interpret user feedback, `gpt-3.5-turbo` model was employed [36], selected for its effectiveness in understanding and generating natural language. Since the robot accepted spoken input, Whisper model [218] was used for speech-to-text transcription. This ensured that verbal advice from users could be reliably processed into textual input for interpretation.

Reinforcement Learning Algorithm. Q-learning was used as the underlying RL algorithm [262]. As a model-free algorithm suitable for discrete action spaces, Q-learning was chosen for its convergence properties and its alignment with the shaping strategies used in ECLAIR. This setup enabled a focused evaluation of ECLAIR's effectiveness as an advice interpretation framework within a controlled task environment.

6.4.3 Task Domain

The task used in this study involves a simplified *pick-and-place* scenario, in which the robot is required to start from a predefined position, navigate toward a cube, pick it up, and place it in a designated target location.

The environment is modelled as a 5×5 grid, yielding 25 discrete cells. The robot's position within the grid is defined by its (x, y) coordinates and encoded as a unique cell location L_{cell}. The full task state is described by the tuple $S = (L_{cell}, \text{is\_picked})$, where is_picked is a binary indicator representing whether the cube has been successfully picked up. This results in a state space comprising 50 distinct states.

The robot is equipped with six primitive actions: moving up, down, left, or right, and executing the `pick` and `place` actions. Each episode begins with the robot positioned at a predefined start location (marked with an X) and terminates either when the cube has been placed or after a maximum of 15 time steps, whichever comes first. The robot's objective is to complete the task using the minimum number of time steps.

6.4.4 Experimental Conditions

A within-subject design was employed to evaluate the contribution of the ECLAIR framework in interpreting and integrating diverse forms of human feedback. Each participant interacted with the system under two experimental conditions:

- **ECLAIR Condition.** In this condition, participants were asked to teach the robot using spoken natural language. Their verbal feedback was transcribed using Whisper and interpreted by an LLM to extract evaluative, corrective, and guidance signals. These were then integrated into the learning process through the appropriate shaping strategies defined by the ECLAIR framework.
- **Evaluative-Only Baseline (Tablet Interface).** In this condition, participants interacted with a tablet interface that allowed them to provide binary feedback by selecting either a "correct" or "incorrect" button after each robot

action. No natural language was used, and only evaluative feedback was collected. This condition mirrors the architecture previously proposed previous works, serving as a comparative baseline for assessing the added value of ECLAIR's multimodal feedback integration.

6.4.5 Procedure

Participants completed both experimental conditions sequentially in a within-subject design, with the order counterbalanced across individuals. Each condition consisted of eight training episodes, during which participants were asked to guide the robot to complete the pick-and-place task. The number of episodes was determined through preliminary simulations, which indicated that eight episodes were sufficient to support learning convergence under a cooperative teaching strategy, while also maintaining a manageable study duration and accommodating battery constraints.

Before beginning the main study, participants completed one practice episode per condition to become familiar with the task mechanics and the respective interaction modality. A short break was provided between conditions to minimise carryover effects. On average, the total session lasted approximately 1 hour and 30 minutes per participant, resulting in over 18 hours of data across the full participant pool.

In addition to the main interactive experiment, a series of controlled evaluations were conducted to assess the robustness of ECLAIR's LLM-based feedback interpretation. These evaluations did not involve participant interaction but used a labelled dataset of 20 natural language advice instances collected from pilot sessions. Three tests were performed: (1) comparing few-shot and zero-shot prompting strategies by removing examples from the LLM prompt, (2) testing sensitivity to prompt rewording using semantically equivalent but syntactically varied instructions, and (3) translating the dataset into four languages: Arabic, French, Spanish, and Italian, to evaluate multilingual performance. For each test, label accuracy was computed across three random seeds to ensure consistency.

6.4.6 Measures

To evaluate the effectiveness of the ECLAIR framework and compare it against the evaluative-only baseline, three categories of data were collected and analysed:

- **Task Performance.** For each episode, the robot's performance was assessed using a score equal to the negative number of time steps required to complete the task. Penalties were applied for incorrect or incomplete executions: a -15 penalty if the cube was not picked up by the end of the episode, and a -10 penalty if it was placed in an incorrect location. This score was used solely for evaluation purposes and was not used to shape the learning process.

- **Feedback Frequency and Type.** All feedback provided by participants was logged and categorised. In the ECLAIR condition, the number of instances of evaluative, corrective, and guidance feedback was recorded based on the LLM's classification. In the evaluative-only condition, only binary positive or negative labels were logged. This allowed for comparison of interaction patterns and feedback density across conditions.

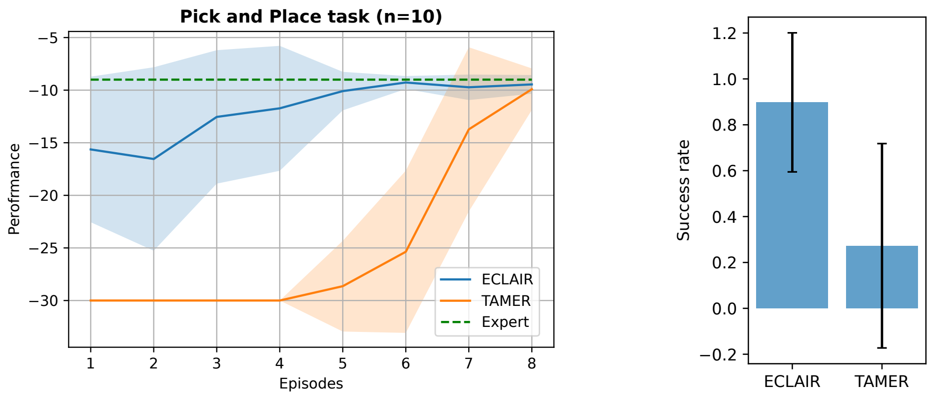
- **Robot Behaviour Logs.** For each episode, complete logs of the robot's state transitions and executed actions were recorded. These traces enabled detailed inspection of the learning process and supported the reconstruction of learning curves, policy changes, and decision consistency over time.

6.5 Results

6.5.1 Task Performance and Learning Progression

Figure 6.3 presents a comparative evaluation of ECLAIR and the evaluative-only baseline (TAMER) on the pick-and-place task. As shown in Figure 6.3a, ECLAIR exhibited faster learning, with performance converging toward expert-level behaviour within the first few episodes. By contrast, TAMER demonstrated a slower learning curve, with improvements observed only in later episodes. This difference was statistically significant, with ECLAIR achieving over 70% greater task success across the trials ($p < 0.0001$), as illustrated in Figure 7.5b.

Figure 6.3: Evaluation of ECLAIR and TAMER on the *Pick-and-Place* task. (a) Performance curve of the models. The expert performance (green dashed line) sets the targeted behaviours. ECLAIR quickly converges towards expert performance, while TAMER does so at a slower pace. (b) Success rates of the training session. ECLAIR is significantly more successful than TAMER.



(a) **(b)**

The differences in exploration patterns were further examined through a state visitation heatmap (Figure 6.4). ECLAIR showed a concentrated visitation pattern along the optimal trajectory, whereas TAMER's visits were more dispersed, indicating broader, less focused exploration.

Figure 6.4: Heatmap of state visitations during training. ECLAIR displays a concentrated activity along the optimal trajectory, while TAMER exhibits a broader exploration.

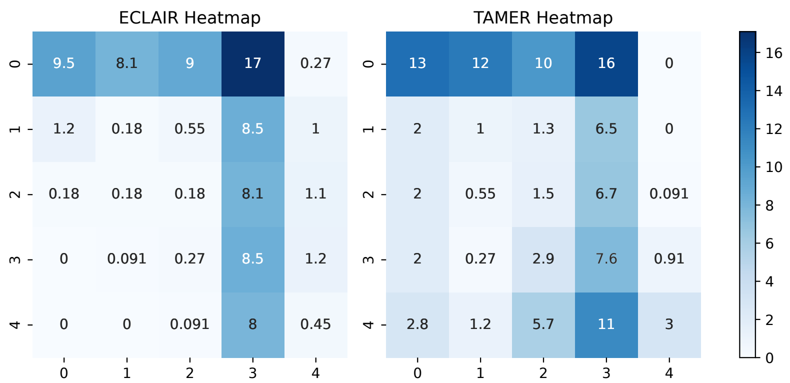
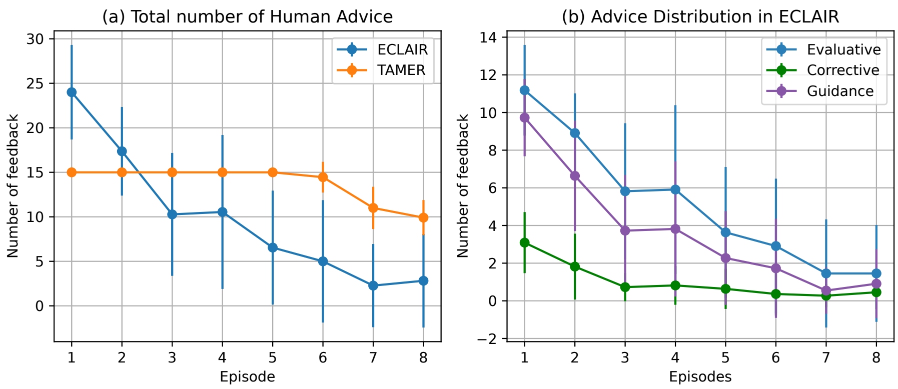


Figure 6.5: Distribution of human feedback across training episodes. (a) Compares the total advice provided in the ECLAIR and TAMER models, highlighting ECLAIR's initial higher engagement that decreases over time, opposite to TAMER's consistent feedback pattern. (b) Breaks down the types of feedback used within ECLAIR, showing a preference for evaluative and guidance feedback over corrective feedback.



6.5.2 Feedback Frequency and Types

Figure 6.5a shows the quantity of feedback provided by participants across episodes in both conditions. In the early stages of training, ECLAIR received significantly more feedback, which tapered off as the robot improved. In contrast, TAMER elicited a more consistent volume of feedback throughout the session.

The distribution of feedback types in the ECLAIR condition is presented in Figure 6.5b. Participants primarily used evaluative and guidance feedback, with corrective feedback being less frequently observed. This suggests that as the robot became more accurate through guidance, fewer corrective interventions were needed.

6.5.3 Robustness of the LLM-Based Advice Interpreter

To assess the generalisation capacity of the ECLAIR interpretation module, label accuracy was computed across several controlled variations in the prompt and language settings.

Prompting Strategy. Figure 6.6a shows that interpretation accuracy remained stable across both few-shot and zero-shot conditions. The absence of performance degradation when removing examples from the prompt suggests that the LLM's pretraining was sufficient to handle the task-specific classification without requiring additional in-context examples.

Prompt Rewording. As shown in Figure 6.6b, changing the wording of the system prompt while keeping the semantic intent intact had no significant effect on label accuracy. Accuracy scores remained consistent across multiple seeds, indicating that ECLAIR is robust to minor variations in prompt phrasing.

Language Flexibility. To evaluate multilingual performance, the advice dataset was translated into Arabic, French, Spanish, and Italian. Figure 6.6c presents the results, showing comparable classification accuracy across all four languages. These findings suggest that the LLM's multilingual training enables ECLAIR to interpret advice effectively across different languages, extending its applicability beyond English-speaking users.

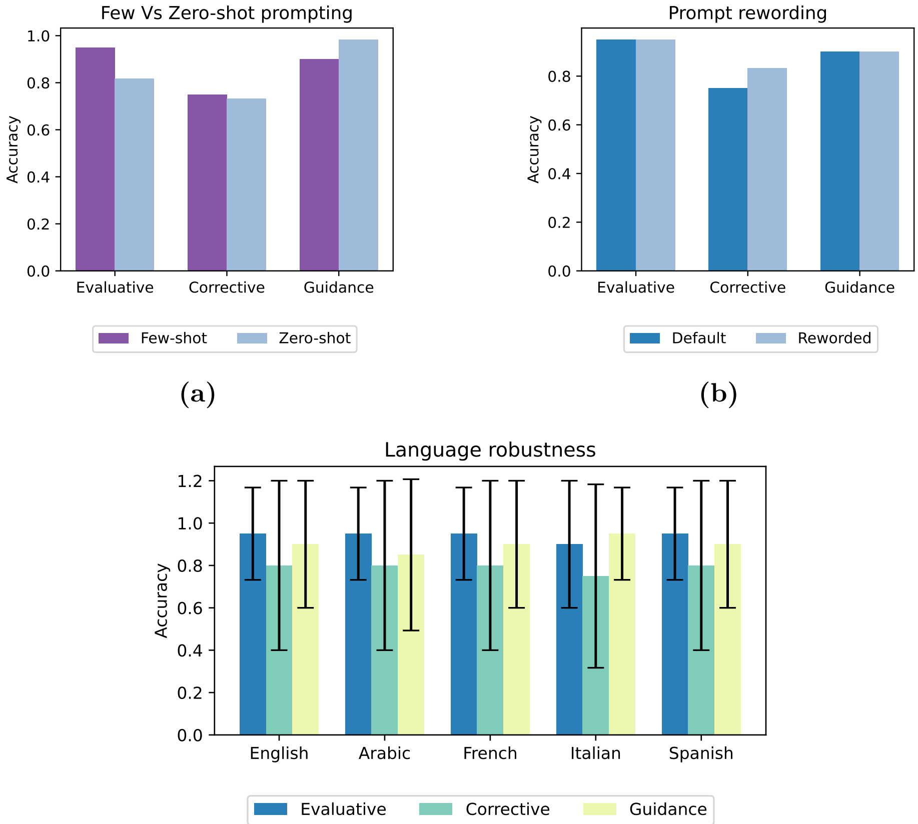
6.6 Discussion

This section interprets the study findings in relation to the research questions that guided the evaluation of the ECLAIR framework, and discusses their broader implications for educational applications, particularly in child–robot interaction.

6.6.1 ECLAIR Enables Faster and More Targeted Robot Learning

The results demonstrate that ECLAIR significantly enhanced the robot's learning efficiency and policy convergence compared to the evaluative-only baseline (TAMER). Robots trained with ECLAIR achieved higher success rates, displayed more focused exploration, and reached optimal behaviour more rapidly, as evidenced by both the performance curves and state visitation heatmaps. These findings are consistent with prior work in Interactive RL, where the inclusion of corrective and anticipatory signals has been shown to guide agents more effectively than scalar rewards alone

Figure 6.6: Evaluating ECLAIR's interpretation accuracy across prompting methods and languages.



(a) (b) (c)

[266, 104]. Recent research also supports the value of allowing multiple feedback modalities in producing more robust and accelerated policy learning [57].

By permitting the use of evaluative, corrective, and guidance feedback, ECLAIR supports more flexible and proactive teaching interactions. This aligns with educational theories of scaffolding and peer instruction, where effective teachers adapt their instructional strategies based on the learner's needs and responses [293, 274]. In the context of ECLAIR, forward-looking guidance played a critical role in reducing the robot's trial-and-error behaviour, enabling participants to directly steer the learning process rather than merely react to errors.

From an educational standpoint, the multimodal flexibility of ECLAIR has potential value in child–robot learning settings. Children are likely to differ in their preferred teaching approaches: some may emphasise praise, while others may rely more on hints or corrections. A framework that can accommodate such variability could allow children to adopt teaching styles that feel intuitive and personal, supporting more peer-like interactions in which the robot behaves as a responsive learning companion rather than a passive tool. While this remains to be empirically tested with child participants, prior HRI studies suggest that opportunities for personalisation and agency are linked to greater motivation, engagement, and deeper learning outcomes [160].

6.6.2 Multimodal Feedback Reduces Teaching Effort

ECLAIR not only improved learning outcomes but also reduced the amount of human input required over time. Participants interacting with ECLAIR exhibited a decreasing trend in the frequency of feedback as the robot's performance improved, while those in the TAMER condition maintained a more consistent level of input throughout. This efficiency aligns with findings by Chisari et al. [57], who showed that multimodal feedback strategies reduce teaching burden while improving task performance.

The observed interaction pattern, characterised by high initial guidance followed by evaluative reinforcement, mirrors classic scaffolding strategies in human tutoring, where support is gradually faded as learner competence increases [54]. This strategy was also observed in educational HRI studies, such as Senft et al. [245], and Pareto et al. [201], where children dynamically adjusted their teaching based on the learner's responses.

In classroom contexts, efficiency is certainly important: robots that can learn quickly from fewer interventions are more practical in time-constrained educational environments. At the same time, findings from Chapter 4 showed that longer interaction time was associated with greater retention, suggesting that efficiency should not come at the cost of sustained engagement. The challenge, therefore, is to engineer the right balance; providing enough responsiveness to keep interactions natural and manageable, while still allowing children sufficient opportunity to elaborate, reflect, and consolidate their own learning through teaching. Future work should examine how frameworks like ECLAIR can be integrated into educational designs that balance efficiency with the pedagogical value of extended interaction.

6.6.3 ECLAIR Is Robust Across Prompts and Languages

The interpretation module of ECLAIR, powered by a large language model, exhibited strong robustness to prompt variation and multilingual usage. Classification accuracy remained consistent across few-shot and zero-shot prompting, semantically reworded prompts, and translations into four different languages. These findings are aligned with results from instruction-tuned language models, which have demonstrated stable performance across prompt formulations in controlled tasks [141, 172].

Moreover, the system's consistent performance across Arabic, French, Spanish, and Italian confirms that ECLAIR's advice interpretation capabilities extend to diverse linguistic settings. This is particularly relevant in multicultural classrooms, where language diversity is common. Allowing children to interact with the robot in their native or preferred language not only supports inclusivity but also reduces barriers to participation and engagement. Studies on multilingual large language models have similarly demonstrated that semantic understanding can generalise well across languages [301].

Importantly, this robustness also alleviates the need for children to adhere to rigid verbal templates. Instead, they can express themselves naturally, using varied phrasing and styles. This facilitates more authentic, fluid learning interactions, and reflects real-world communication better than constrained interfaces. As such, ECLAIR supports a form of interaction that is both technically effective and socially aligned with children's natural communicative behaviours.

6.7 Summary

This chapter introduced ECLAIR, a novel framework for enabling reinforcement learning from diverse forms of natural language feedback using LLMs. Motivated by developmental and educational research highlighting the multimodal nature of human teaching, ECLAIR was designed to interpret evaluative, corrective, and guidance signals and to integrate them into the robot's learning process using dedicated shaping strategies.

The framework was implemented and evaluated in a within-subject HRI study. Twelve adult participants taught a social robot to perform a pick-and-place task under two conditions: ECLAIR, which supported multimodal verbal advice, and a baseline condition limited to binary evaluative feedback. Results showed that ECLAIR led to significantly faster learning, higher success rates, and more efficient task completion, while also reducing the overall amount of feedback required. Analysis of feedback patterns revealed that participants naturally used a mixture of feedback types, favouring evaluative and guidance input over corrective feedback.

Further experiments demonstrated that ECLAIR's LLM-based interpretation module is robust to prompt variations and performs consistently across multiple languages, supporting its adaptability for use in multilingual and naturalistic learning environments.

Together, these findings demonstrate the technical feasibility of ECLAIR and its potential educational relevance. By aligning robot learning mechanisms with human pedagogical strategies, the framework shows how natural language can be leveraged to interpret diverse forms of feedback in real time. Although the present study was conducted with adults, the approach has clear promise for future child–robot teaching interactions, where variability in teaching styles and spontaneous feedback are central. ECLAIR therefore represents an important step toward scalable, language-enabled, peer-like robots for educational settings, pending validation in classroom contexts.

Chapter 7

Sleep-Inspired Consolidation in Teachable Robots

[Epigraph removed for copyright reasons. Original source: J.K. Rowling, Harry Potter and the Deathly Hallows *(London: Bloomsbury, 2007)]*

7.1 Introduction

This chapter explores how a robot can consolidate learning from evaluative human feedback to acquire more generalisable, goal-directed behaviours. While earlier chapters investigated how children can engage in teaching interactions with robots through Interactive RL using evaluative feedback, and explored how this approach can be extended with natural language to enable robots to interpret richer multimodal advice, both approaches relied on feedback processed in a largely myopic fashion, where learning occurred only in response to immediate signals. This chapter introduces INFORM, a sleep-inspired consolidation framework that allows robots to revisit and reinterpret past trajectories offline, moving beyond purely reactive learning toward more durable and generalisable knowledge.

The remainder of the chapter outlines the architecture and learning objectives of INFORM, describes the experimental protocol used in both discrete and continuous tasks, and presents results evaluating policy performance, robustness, and reward recovery. The discussion considers implications for cognitive modelling and educational HRI, positioning INFORM as a learning mechanism that is more closely aligned with human cognitive processes.

7.2 Study Motivation and Research Questions

In Interactive RL, human feedback plays a central role, yet differs fundamentally from traditional environment-based rewards. While standard RL optimises dense reward signals over long time horizons, human feedback is typically communicative and local, serving to indicate whether an action was correct rather than to drive long-term behaviour optimisation [117]. As a result, it often conflicts with the assumptions of reward-maximising agents.

Ho et al. [114] showed that such feedback is state and action-contingent, guided by pedagogical intent rather than long-term reinforcement structure. Positive feedback, in particular, is commonly used to affirm momentary correctness rather than to shape policy trajectories. Consequently, RL agents can misinterpret these signals, leading to positive cycles; where agents repeatedly visit familiar states to collect positive feedback, regardless of whether this aligns with the task goal.

This phenomenon was further validated by Knox and Stone [143], who observed that in episodic tasks, human trainers overwhelmingly provided positive feedback, causing agents with high discount factors to fall into these positive cycles. To prevent this, they advocated for strong discounting (low γ), which shifts the learning focus to immediate feedback. Specifically, myopic agents ($\gamma \approx 0$) avoid positive loops and align more closely with the teacher's immediate intent, thereby improving task performance. Prior work also adopted this myopic learning strategy, often without explicitly recognising its theoretical justification [266, 142, 190].

Yet, this myopic approach comes at a cost: while it aligns well with the local nature of human feedback, it limits the agent's ability to internalise long-term goals, generalise knowledge, or adapt to new contexts [144]. Thus, Interactive RL agents face a fundamental trade-off: agents with high γ risk being misled by biased feedback and entering reward loops, while myopic agents fail to build robust, transferable strategies.

To address this, Knox and Stone introduced VI-TAMER [144], a model that integrates TAMER's interactive learning with explicit value iteration to support non-myopic learning. In continuing (non-episodic) tasks, VI-TAMER successfully reconciled the benefits of interactive learning with long-horizon planning, achieving better alignment with human intent without falling into positive circuits. Their goal was to move beyond purely reactive, feedback-dependent behaviour and enable robots to infer and pursue meaningful outcomes, even in the absence of frequent feedback. However, VI-TAMER depends on a known transition model and is restricted to discrete environments, limiting its applicability in many real-world HRI scenarios.

In contrast, humans are not constrained by such limitations. They are capable of generalising from sparse and evaluative feedback, and of transferring this learning to novel situations. From the joyous cheers that greet a baby's first steps to the pride expressed after an achievement, humans are exposed to feedback throughout life. This feedback plays a crucial role in learning, supporting skill acquisition, decision-making, and adaptation to change [291]. However, this process is not instantaneous. Robust, transferable knowledge does not emerge solely from real-time feedback but relies on consolidation mechanisms, with sleep playing a key role [259].

Studies have shown that deep sleep enhances memory and learning by enabling the consolidation and abstraction of knowledge [120]. Initially, newly acquired information is encoded in the hippocampus. During sleep, this information is replayed, reorganised, and integrated into cortical structures, allowing long-term retention and facilitating generalisation [220]. Moreover, humans do not simply react to feedback—they infer the goal and intent behind it and internalise the corresponding reward structure to continue learning independently [115].

Transferring this capacity to robots would enable peer-like agents to move beyond immediate learning and generalise to broader learning contexts. This chapter introduces the INverse Forward Offline Reinforcement Model (INFORM), a two-phase learning framework designed to mirror this human-like process. INFORM begins with a myopic Interactive RL phase, in which the robot learns to predict the teacher's preferred actions from evaluative feedback. These interaction trajectories are stored in a replay buffer. In the second phase, a simulated "sleep" phase, INFORM replays these experiences offline and applies Inverse RL with a high discount factor, recovering a reward function and policy that encode the broader goals implied by the feedback [97]. This offline consolidation enables the robot to transition from reactive behaviour to a robust, autonomous, and generalisable learning strategy, bridging the gap between interactive teaching and cognitive modeling.

This chapter is guided by the following research questions:

- **RQ5.1:** Can INFORM effectively learn a high-level policy from past experiences?
  Interactive RL typically results in myopic policies tailored to specific feedback episodes. This question investigates whether INFORM can extract more abstract, generalisable knowledge from those same experiences—enabling autonomous decision-making that extends beyond reactive feedback.

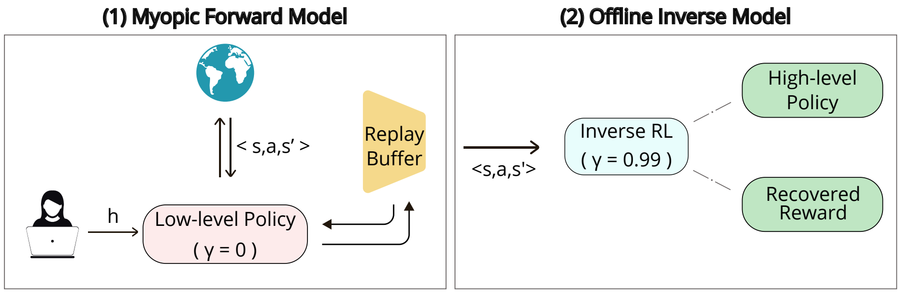
- **RQ5.2:** How robust is the consolidated policy learned by INFORM?
  Human-robot teaching environments are often noisy, sparse, or subject to change. Ensuring that the policy resulting from INFORM remains stable under different conditions is essential for deploying teachable robots in real-world educational settings.

- **RQ5.3:** Is the reward function retrieved by INFORM in alignment with the teacher's intended outcomes?
  For a robot to truly learn from human feedback, it must internalise the underlying goals—not just surface-level actions. This question assesses whether INFORM can recover a reward representation that reflects the human teacher's intended task structure, thus supporting further autonomous learning.

7.3 The INFORM Framework

This section introduces the **INverse Forward Offline Reinforcement Model (INFORM)**, a two-phase learning framework designed to emulate the way humans

Figure 7.1: Illustration of the framework. (1) The agent first learns a low-level policy with a myopic interactive RL. All trajectories of the interaction are stored in a buffer. (2) An offline inverse RL is then applied to the stored trajectories to recover a reward and a policy that better encodes the high-level information of the task.



consolidate feedback into abstract knowledge. INFORM aims to address the generalisation limitations of myopic Interactive RL by combining immediate feedback-based learning with a sleep-inspired offline consolidation phase. As illustrated in Figure 7.1, the framework comprises two sequential stages:

1. **Forward Model (Interactive Learning Phase):** The robot learns a low-level, myopic policy through real-time interaction with a human teacher who provides binary evaluative feedback on state-action pairs.

2. **Offline Inverse Model (Consolidation Phase):** The robot replays previously stored interaction trajectories and applies an offline inverse reinforcement learning algorithm to extract a high-level reward function and policy that better reflect the intended goals of the task.

7.3.1 Forward Model: Learning from Human Feedback

The first phase models the real-time teaching process. Initially, the robot follows a random policy and updates its behaviour based on **binary human feedback** $h \in \{-1, +1\}$, which signals whether a given action in a specific state was appropriate. We implement this using the TAMER framework [142], where a learned function H_θ predicts the expected human feedback. The learning objective is formulated as:

$$\mathcal{L}(\theta_H) = \mathbb{E}_{(s,a,H)\sim\mathcal{D}} \left\| \hat{H}(s, a; \theta_H) - H(s, a) \right\|^2 \tag{7.1}$$

This approach enables the agent to immediately adapt its behaviour to the teacher's preferences.

If actor-critic architectures are used, the action policy $\pi(a|s;\phi)$ is also updated based on the expected human feedback:

$$\mathcal{L}\left(\phi_\pi\right) = \mathbb{E}_{(s,a,\pi)\sim\mathcal{D}}\left[H(s,a) - \log\pi(a|s;\phi)\right] \tag{7.2}$$

At the end of each episode, the robot evaluates its performance based on an environment-derived return (e.g., task completion) and assigns a success label to the trajectory. These labelled trajectories are stored in a **replay buffer** $\mathcal{D}$, forming the experience set for the offline phase.

7.3.2 Offline Inverse Model: Sleep-Inspired Consolidation

While the forward model efficiently captures the teacher's surface-level preferences, it is inherently **myopic**, focusing solely on immediate action evaluation and lacking long-term planning. This restricts the robot's capacity to understand the global structure of the task or to perform robustly in unseen variations.

To address this, INFORM incorporates a sleep-inspired consolidation phase that mimics the human brain's memory processing during deep sleep, where newly encoded episodic experiences are replayed and reorganised to extract abstract representations [33]. Technically, this is implemented via a modified version of *IQ-Learn* [97], a dynamic-aware imitation learning model that effectively learns a Q-function from a few demonstrations and uses it to derive both a policy and a reward function.

Unlike standard IQ-Learn, which depends on expert demonstrations and online interaction with more complex continous environement, INFORM modifies the objective to learn from trajectories generated during the forward phase, using only offline replay. *Successful* and *unsuccessful* trajectories are sampled from the replay buffer and used to optimise the following objective:

$$\max_{Q \in \Omega} \mathcal{J}^*(Q) = -\mathbb{E}_{(s,a,s') \sim \text{success}} \left[Q(s,a) - \gamma V^* \left(s' \right) \right] \\ - \mathbb{E}_{(s,a,s') \sim \text{replay}} \left[V^\pi(s) - \gamma V^\pi \left(s' \right) \right] \quad (7.3)$$

where the state-value function is defined as:

$$V^*(s) = \log \sum_a \exp Q(s,a)$$

This objective enables the robot to extract high-level patterns from past experiences by assigning credit over longer temporal horizons. From the optimised Q-function, a reward function can be recovered:

$$r\left(s, a, s'\right) = Q(s,a) - \gamma V^\pi \left(s' \right) \quad (7.4)$$

This reward function reflects the task's underlying goals as inferred from successful human-guided behaviour, enabling autonomous future learning.

7.3.3 Algorithm Overview

Algorithm 2 outlines the full INFORM training process. The first part corresponds to the forward model, where the agent interacts with a human and updates its policy based on feedback. The second part performs offline inverse RL using the accumulated experiences. The full procedure is compatible with both value-based and policy-gradient methods.

7.4 Methodology

7.4.1 Environments

To examine INFORM's performance across different domains, two types of tasks were selected: a discrete grid-based environment and a continuous robotic control environment.

Algorithm 2 Pseudocode of the INFORM Framework

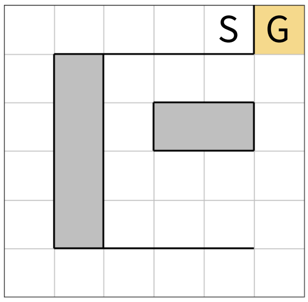
=== Forward Model ===

Initialise parameters of H_θ, π_ϕ, and a replay buffer $\mathcal{D}$
for each episode **do**
    Initialize a temporary buffer $\mathcal{T}$
    for each step **do**
        $s \leftarrow$ observation
        Select action a:
                (Q-learning) $a = argmax(H_\theta(s))$
                (actor-critic) $a = \pi\left(\cdot \mid s_t; \phi\right)$
        $s' \leftarrow$ executing a
        $h \leftarrow$ human feedback
        Store in temporary buffer $\mathcal{T} \leftarrow \mathcal{T} \cup (s, a, s', h)$
        Perform a gradient step for θ_H using Equation 7.1
    end for
    Assess success of trajectories $\triangleright$ *success* $\in \{0 \text{ or } 1\}$
    Store labeled trajectories in replay buffer $\mathcal{D} \leftarrow \mathcal{D} \cup (\mathcal{T}, success)$
end for

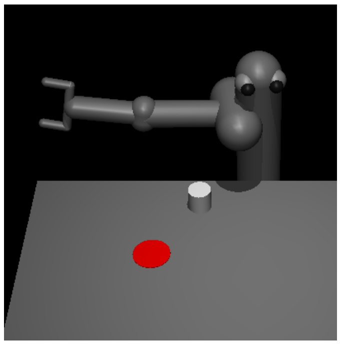
=== Inverse Model ===

Initialise parameters of Q_θ and optionally π_ϕ
for each step **do**
    Get batch from replay buffer $\sim \mathcal{D}[(s, a, s', success)]$
    Perform a gradient step for θ_Q using Equation 7.3
    (only with actor-critic) Perform a gradient step for ϕ_π
end for
Recover reward function using Equation 7.4

Figure 7.2: The Gridworld and Pusher-v4 environments used to evaluate the INFORM framework.



(a) Gridworld



(b) Pusher

Gridworld Maze. For the discrete scenario, a 36-cell gridworld maze introduced by Knox et al. [144] was used. The agent starts in cell S and must reach the goal cell G using one of four directional actions: up, down, left, or right. A trajectory is considered successful if the goal is reached in under 30 steps. This environment, while straightforward, provides a comprehensive platform for a detailed and observable evaluation of our framework.

Pusher-v4. For the continuous setting, `Pusher-v4` task from the Mujoco Gymnasium suite was used [268]. In this task, a multi-jointed robotic arm must push a cylinder toward a target position using its end-effector. A trajectory is considered successful if the object's final position is within 0.08 units of the target. This task introduces dynamic complexity and serves as a strong benchmark for evaluating robustness and generalisation.

7.4.2 Simulated Feedback

The aim of this study is not to improve how robots learn directly from human feedback, but to explore how such feedback can be consolidated into more generalisable knowledge. Since the forward model has already been well established in prior work [142], the focus here is placed on evaluating the second phase of the framework—the offline consolidation.

To reduce reliance on human participants and allow for controlled, reproducible evaluation of the reverse model, an oracle was used to simulate evaluative feedback. This allowed us to focus on testing whether INFORM can recover a high-level policy and reward function from interaction trajectories, without introducing variability from human input.

In line with Zhang et al. [299], the oracle was implemented as a fully trained model. For each state s, the learning agent selected an action a from its current policy, while the oracle selected an optimal action $a*$. Positive feedback was provided when the Q-value of the selected action was sufficiently close to that of the optimal action; otherwise, no feedback was given. This mechanism was designed to reflect the positive feedback bias commonly observed in human teaching [117]. Specifically:

$$h(s,a) = \begin{cases} +1 & \text{if } Q(s,a) \geq \alpha Q\left(s,a^{*}\right) \\ 0 & \text{otherwise} \end{cases} \tag{7.5}$$

Here, α is a scheduling parameter that increases over time to simulate the diminishing frequency of feedback, a phenomenon commonly observed in human teaching behaviour [117].

7.4.3 Implementation Details

INFORM was implemented using two different RL architectures based on the environment. For the discrete gridworld task, a Deep Q-Network (DQN) [182] was used, while a Soft Actor-Critic (SAC) [106] was utilised for the continuous pusher task. The codebase was adapted from CleanRL [119] to follow the INFORM algorithmic structure (Algorithm 2). Hyperparameters for both the interactive and offline phases were based on the original IQ-Learn implementation [97], with further tuning for stability across replay buffer sizes and trajectory lengths. The full code and parameter configurations are available on the project repository[1].

[1]https://github.com/ImeneTar/INFORM

7.4.4 Evaluation Protocol

The evaluation was structured around the three research questions introduced earlier, with each question addressed in a specific environment and analysis setup. All experiments followed a two-phase learning process:

1. **Forward model:** Agents were trained using simulated human feedback via TAMER. This phase involved 100 episodes in the Gridworld environment and 500 episodes in Pusher-v4.

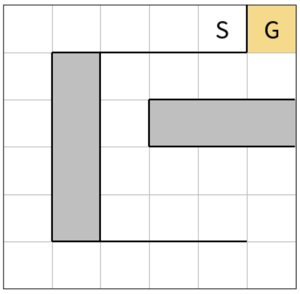
2. **Inverse offline model:** INFORM replayed the stored interaction trajectories and applied inverse reinforcement learning to derive a high-level policy and reward function. No further environment interaction occurred during this phase.

To answer RQ5.1, an experiment assessing whether INFORM could consolidate a non-myopic policy from interaction data was proposed. we evaluated the learned policies in both Gridworld and Pusher environments. Performance was compared to the original myopic policy (TAMER) and to an expert policy trained using environment-defined rewards. Success was measured by the number of steps to reach the goal in Gridworld, and by the final distance to the target object in Pusher.

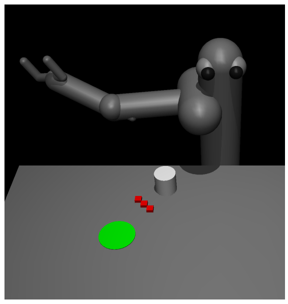
To adress RQ5.2, an experiment asseing the robustness of INFORM was run in the Pusher environment. Following Eysenbach et al. [86], environmental dynamics were altered by introducing an obstacle along the perpendicular bisector between the puck's initial position and the goal. This obstacle comprises three axis-aligned blocks, each 3 cm wide, located at (0.32, -0.2), (0.35, -0.23), and (0.38, -0.26). The modified environment is illustrated in Figure 7.3b. The previously trained TAMER and INFORM policies were tested on this modified environment across 100 episodes per run (3 runs in total). The success metric was the final distance between the object and the target.

RQ5.3 was examined in the **Gridworld** environment. After INFORM was trained in the original layout, the recovered reward function was extracted and used to train new agents in a modified version of the environment, where the optimal path was blocked by a wall (Figure 7.3a). This setup was designed to evaluate whether the recovered reward captured the task's high-level objective, rather than overfitting to the original policy shaped by human feedback. Following [144], agents were trained using Q-learning with the recovered reward and compared to agents trained with the environment-defined reward, as well as to the original TAMER policy. Learning performance was monitored over 300 episodes.

Figure 7.3: Perturbation in environmental dynamics. (a) A wall is added in Gridworld to block the optimal path. (b) The optimal trajectory is obstructed with an obstacle.



(a)



(b)

Evaluation Details. All experiments were repeated using five random seeds for Gridworld and three for Pusher. Average performance was reported using environment-specific metrics: episode length in Gridworld and final object distance in Pusher. The same seeds were maintained across models and conditions to ensure fair comparison.

7.5 Results

7.5.1 Performance evaluation

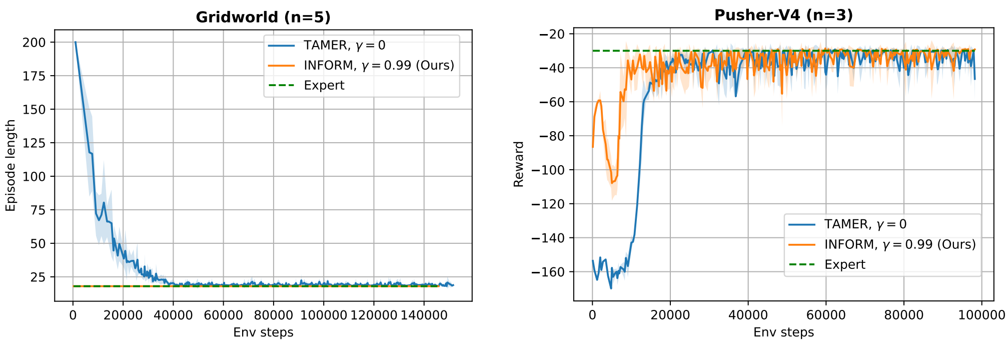
The performance of INFORM was evaluated against TAMER (myopic baseline) and an expert policy across both tasks. As shown in Figure 7.4, INFORM successfully recovers a high-level policy from stored trajectories alone, without requiring additional interaction with the environment.

In the Gridworld task, INFORM (orange line) not only matches expert-level performance but also converges more rapidly and consistently than TAMER (blue line). Variability across the five runs is noticeably reduced, suggesting greater stability. This improvement is attributed to the replay of diverse trajectories during the offline phase, enabling INFORM to capture longer-term structure that is otherwise inaccessible through myopic feedback alone.

In the more complex Pusher-v4 task, INFORM achieves performance comparable to TAMER while demonstrating faster convergence. This suggests that even in high-dimensional, continuous control settings, the consolidation mechanism supports robust policy recovery from feedback-guided trajectories.

Overall, these results demonstrate that INFORM effectively consolidates knowledge from both successful and unsuccessful experiences, producing high-level policies that retain or exceed the performance of myopic baselines across tasks of varying complexity.

Figure 7.4: Evaluation of INFORM and TAMER across discrete and continuous Tasks. The expert performance (green dashed line) sets the target for each environment. Higher values indicate superior performance in Pusher, while lower values are better in Gridworld. INFORM effectively recovers non-myopic policies, matching TAMER's performance.



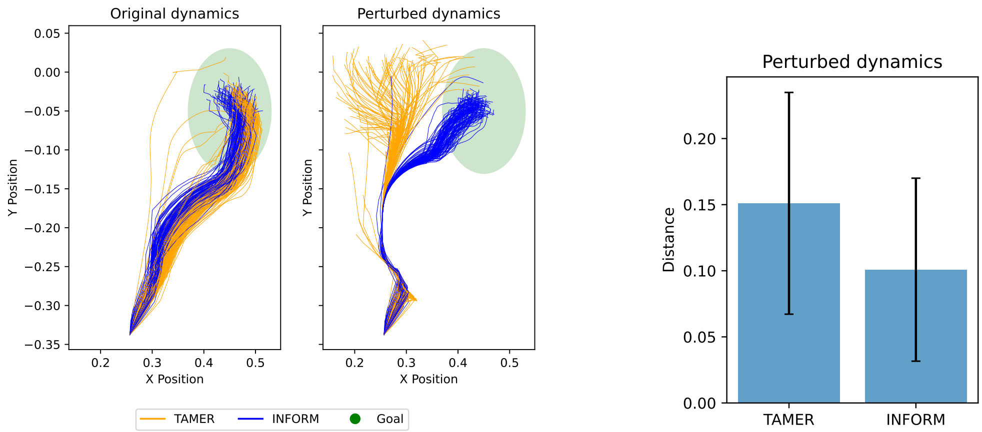
(a) Gridworld

(b) Pusher

7.5.2 Robustness to Environmental Perturbations

The robustness of the policies learned by INFORM and TAMER was evaluated under altered environmental dynamics. As shown in Figure 7.5b, INFORM consistently moved the object closer to the target than TAMER across three runs, with a

Figure 7.5: Robustness evaluation. (a) Comparison of TAMER and INFORM trajectories on both original and perturbed environments. (b) Comparison of distance from target over 3 runs. INFORM significantly moves the object closer to the goal compared to TAMER, learning a high-level policy that more effectively navigates around obstacles.



(a) **(b)**

statistically significant difference in performance ($p < 0.0001$). This suggests that the policy consolidated through INFORM generalises more effectively to dynamic changes.

Figure 7.5a further illustrates that INFORM is more likely to adapt its behaviour and navigate around unforeseen obstacles, while TAMER tends to fail when deviations from the original optimal trajectory are required. This difference is likely due to INFORM's use of a high discount factor, which supports the integration of long-term consequences into policy learning. In contrast, TAMER's myopic focus results in brittle policies that underperform in perturbed settings.

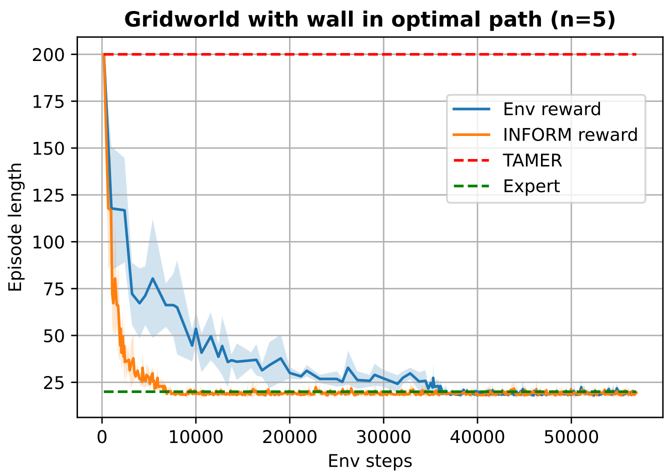
While INFORM demonstrates superior robustness overall, the results also reveal considerable variance across random seeds. This suggests that the effectiveness of the consolidation phase may depend on the diversity and representativeness of the stored trajectories. Improving consistency remains an open challenge for future work.

7.5.3 Generalisation of Recovered Reward Functions

Figure 7.6 shows the learning performance of agents trained in the modified Gridworld using the reward recovered by INFORM. Agents trained with this inferred reward consistently learned to reach the goal, converging more quickly and reliably than those trained with the environment-defined reward. This suggests that the reward recovered by INFORM provides denser and more informative guidance, facilitating faster learning.

TAMER-trained policies, when deployed directly in the modified environment without further training, failed to complete the task in all evaluation runs. On average, these policies terminated after 29.2 steps with a mean reward of 0.0, indicating no adaptation to the modified path.

Figure 7.6: Learning curve of agents in the gridworld with a wall blocking the optimal path. Agents learning with INFORM reward successfully and rapidly learn an optimal policy.



7.6 Discussion

7.6.1 High-level policy through Offline Consolidation

The results show that INFORM is able to recover a non-myopic, high-level policy from feedback-guided trajectories without requiring further interaction with the environment. This demonstrates that human evaluative feedback, typically effective only in short-sighted learning, can be consolidated offline to produce more abstract, goal-directed behaviour in a model-free setting.

In contrast to prior approaches that depend on known transition models or continuous online interaction, INFORM performs this consolidation entirely from past experiences. This offline mechanism makes the process safer and more stable, which is particularly relevant in educational and interactive settings where repeated trial-and-error or constant feedback may not be feasible.

Moreover, this consolidation process reflects principles found in biological learning systems. By replaying stored episodes to abstract and reinforce useful behaviour, INFORM mirrors aspects of sleep-based memory consolidation in the brain, where hippocampal activity trains cortical networks for long-term generalisation. This connection aligns with broader efforts in neurorobotics to design cognitive models that are not only functional, but also inspired by how humans learn. INFORM thus contributes toward the development of teachable robots that move beyond behaviour shaping to form internal, structured representations of learning experiences.

7.6.2 Consolidation Improves Generalisation under Perturbation

The finding that INFORM outperformed TAMER in a perturbed environment suggests that offline consolidation may enhance a robot's ability to generalise beyond the specific conditions in which feedback was originally provided. By replaying both successful and unsuccessful trajectories with a non-myopic objective, INFORM appears to capture a broader understanding of task dynamics, one that is less dependent on the exact path or state sequence seen during training.

This represents a promising direction for teachable robots in real-world settings, where conditions are rarely static and perfect feedback is rarely available. The ability to adapt without requiring new interactions or human corrections is particularly valuable in educational contexts, where learner input is limited and task environments may vary. At the same time, the observed variance in performance across seeds highlights a limitation: generalisation through consolidation may depend heavily on the diversity and coverage of the stored trajectories. Further work is needed to understand how to select, weight, or augment replay data to improve reliability across settings.

7.6.3 Reward Recovery Enables Adaptive Learning

The ability to recover a reward function that supports learning in a changed environment suggests that INFORM goes beyond behavioural imitation and begins to internalise aspects of the task structure itself. This is particularly notable given that the reward was inferred from human feedback, often sparse and local, rather than from expert demonstrations or engineered signals. By consolidating trajectories offline, INFORM extracts a more abstract, task-level representation of what constitutes success, enabling learning to continue autonomously in new settings.

This points to an important step toward enabling teachable robots to learn in ways that more closely mirror human cognition: not just through behavioural replication, but by inferring goals and adapting to new contexts. While the environments used here were simple, the approach opens up future possibilities for feedback-driven systems to learn reward models that are flexible, transferable, and aligned with human instructional intent.

7.7 Summary

This chapter introduced INFORM, a two-phase framework designed to consolidate human feedback into high-level, generalisable behaviour. Building on a forward Interactive Reinforcement Learning phase guided by evaluative feedback, INFORM replays stored trajectories offline using inverse reinforcement learning to infer both a robust policy and a reward function.

Across discrete and continuous tasks, INFORM was shown to recover policies that matched or exceeded the performance of a myopic baseline, all without requiring further interaction with the environment. These consolidated policies demonstrated improved generalisation under dynamic changes, outperforming feedback-trained policies in perturbed settings. Moreover, the reward function recovered during offline learning was shown to support autonomous policy learning in a modified environment, highlighting its alignment with the underlying task intent.

Taken together, these findings provide initial evidence that feedback-based learning can be extended beyond short-term adaptation. INFORM consolidates learning in a way that is model-free, offline, and more reflective of human cognitive processes, bringing teachable robots a step closer to biologically inspired, goal-directed learning architectures.

Chapter 8

General Discussion

This chapter draws together the findings from all empirical and technical work presented in this thesis, integrating them in relation to the research questions and challenges outlined in Chapter 1. It considers the contributions of the work to educational robotics and its grounding in developmental theory, reflects on the psychological and social dimensions observed in the studies, and discusses limitations and ethical considerations arising from the research.

8.1 Overview of Findings and Contributions

This thesis set out to explore whether social robots could serve as credible peer-like learning companions, capable of learning from children through natural teaching interactions, while also supporting the child's cognitive and emotional development. As outlined in Chapter 1, this central aim was framed through five research questions and five core challenges: realistic cognitive modelling, natural testing, psychological implications, natural interaction, and generalisation. To address these, five studies were conducted, each targeting one or more of these challenges. Together, they form a coherent trajectory: from the design of a cognitively plausible learning model, to its evaluation in authentic classrooms, to refinements in interaction, psychological understanding, and long-term learning mechanisms.

A central contribution of this thesis is the application of Interactive RL to child–robot teaching (RQ1). Rather than developing a novel algorithm, the novelty lies in framing learning tasks as an RL problem, specifically, as an MDP shaped by evaluative feedback from children. This allowed the robot to adapt its behaviour in real time, without reliance on pre-scripted behaviours or Wizard-of-Oz control, and demonstrated how Interactive RL can serve as a cognitively plausible model for peer-like learning.

Building on this architecture, a large-scale classroom study addressed the question of whether learning-by-teaching robots can improve educational outcomes compared to independent practice (RQ2). Designed for ecological validity, the study embedded multiple autonomous, personalised robots in regular classroom sessions, allowing children to interact one-on-one without disrupting the lesson. Compared to tablet-based self-practice, the robot condition yielded measurable improvements in retention, providing compelling evidence for the practical value and scalability of learning-by-teaching robots in authentic educational settings.

Subsequent studies investigated how involvement in the teaching process shapes psychological factors such as trust and psychological ownership (RQ3). Findings showed that higher involvement could foster self-investment, though trust was shaped more by perceived performance than by involvement alone, raising important pedagogical and ethical questions about transparency and epistemic trust.

The architecture was then extended to support richer forms of verbal feedback (evaluative, corrective, and guidance) by leveraging large language models (RQ4). This enabled more natural, flexible teaching exchanges and accelerated the robot's learning, underscoring the benefits of aligning robot capabilities with human communication patterns.

Finally, to address the limitations of short-sighted, task-specific learning (RQ5), a sleep-inspired consolidation mechanism was introduced. This allowed the robot to retrospectively process prior experiences and generalise beyond the immediate task, whic could be an essential step toward long-term, transferable learning.

Collectively, these contributions advance the vision of educational robots not as content-delivery tools, but as adaptive, ethically grounded partners in the learning process, capable of growing alongside the child, responding to their input, and fostering engagement, reflection, and long-term understanding.

8.2 Educational Implications Grounded in Developmental Theory

This thesis is grounded in developmental theories that view learning as a socially mediated, constructivist process, shaped not only by individual cognition but by interaction, explanation, and reflection [210, 280, 293]. These principles informed the design of the teachable robot, ensuring that its behaviour and role in the classroom aligned with well-established mechanisms for fostering learning [52, 199, 232].

The use of a peer-like robot that learns from child-provided feedback reflects Vygotsky's notion of the Zone of Proximal Development (ZPD) [280], where progress is achieved through interaction with a more or less knowledgeable other. In the classroom study (Chapter 4), the child acted as the more knowledgeable partner, while the robot took on the novice role. This inversion of the traditional teacher–student dynamic created opportunities for children to rehearse and consolidate their own understanding, directly supporting Bruner's and Chi's theories of learning-by-teaching [293, 54]. Bandura's emphasis on feedback [16], modelling, and reciprocal influence in cognitive development, is embedded in the learning architecture developed in Chapter 3. By enabling the robot to adapt in real time to human feedback, the system reproduced the feedback loops present in effective peer learning. The child could observe the impact of their input on the robot's behaviour, reinforcing a sense of agency and encouraging continued engagement. These principles translated into concrete design strategies for education: structuring interactions so that the child's role was active and visible, embedding clear feedback pathways to support agency, and framing the robot as a fallible peer to invite explanation and reflection. By implementing these strategies in authentic classrooms (Chapter 4) and controlled studies on trust and ownership (Chapter 5), this work demonstrated how developmental theory can directly inform the creation of robotic learning partners that are not only technically functional, but also pedagogically meaningful.

Beyond its theoretical alignment, the classroom study in Chapter 4 offers clear pedagogical insights. Children who taught the robot demonstrated higher engagement and, in the vocabulary task, significantly better retention than those who practised independently. This supports prior evidence for the learning-by-teaching effect [52, 90, 25], where tutoring another, whether a human peer or a teachable agent, reinforces the teacher's own understanding through active recall, elaboration, and self-explanation. The results further suggest that the physical embodiment of the robot, coupled with its peer framing, fostered a sense of responsibility and sustained attention, consistent with the Protégé Effect [46]. Unlike concerns raised in earlier work that social robots might distract from learning [135], the robot in this study maintained or enhanced learning outcomes, supporting metacognitive engagement through features such as help-panel use and iterative feedback. These findings indicate that peer-like robots can act as both motivational and cognitive scaffolds, embedding retrieval-based and reflective learning strategies into everyday classroom activities

Finally, the progression in this thesis, from simple evaluative exchanges to richer, more generalisable learning (Chapters 6 and 7), mirrors the developmental trajectory of human learning: from immediate trial-and-error adjustments to abstract reasoning and reflective practice. While the robot does not replicate human development, its design draws inspiration from these layered processes, offering a model for how educational technology can scaffold deeper, more enduring learning.

8.3 Psychological and Social Implications

Beyond cognitive learning outcomes, this thesis examined the psychological and social dimensions of interacting with a teachable robot, as explored in the involvement and trust studies presented in Chapter 5. These aspects are particularly significant in educational contexts, where engagement, motivation, and emotional connection play a central role in shaping the learning experience [94, 206, 121]. By investigating constructs such as psychological ownership, trust, and user involvement, this work provides new insights into how children and young adults relate to artificial agents designed to learn from them.

The findings from the involvement study indicated that higher levels of user involvement, where participants could both observe and influence the robot's learning trajectory, were associated with increased self-investment, a core indicator of psychological ownership. This pattern suggests that when users are granted agency over the robot's learning process, they may begin to feel personally responsible for its outcomes. However, other dimensions of ownership, such as perceived control and perceived knowledge, were less consistently influenced. This indicates that self-investment alone may not be sufficient to foster a robust sense of ownership, and that interactions must also provide visible and meaningful feedback to reinforce the user's sense of control and impact.

In contrast, trust emerged as a more complex and less predictable construct. Involvement alone did not significantly increase trust, although individual differences, such as anthropomorphism tendencies and prior experience with robots, appeared to mediate perceptions of the robot's competence and reliability. As discussed in Chapter 5, this reflects a broader challenge in social robotics, trust cannot be reliably engineered through system behaviour alone but is shaped by the user's expectations, dispositions, and prior beliefs. From a pedagogical perspective, this underscores the importance of transparency and appropriate framing, particularly in educational settings where children may overestimate a robot's authority or knowledge [277].

The observed pattern of results is consistent with prior findings in HRI research, where perceived ability and performance have been identified as primary determinants of trust. Robot errors have been shown to reduce trust, while cues of competence can strengthen it [237, 107, 8]. Beyond technical performance, benevolence and expressivity, including empathic language, pedagogical expressiveness, and socially responsive prosody or gaze, have also been found to enhance user disclosure and trust, highlighting the role of framing and transparency in trust formation. Similar to the present results, Stadler et al. reported that directly teaching a robot, or being attributed as its teacher, increased self-extension without producing a corresponding change in trust, suggesting that psychological ownership and trust may be shaped by distinct mechanisms [258]. In child–robot interactions, trust has been shown to depend on the robot's accuracy and perceived agency, with children more likely to select previously accurate and agentic robots as informants [32, 98]. This further supports the need for visible and interpretable learning processes to prevent over-attribution of authority.

Beyond one-on-one interactions, research on social influence has shown that robots are more likely to elicit informational conformity, where individuals change their responses because they believe the robot has superior knowledge, than normative conformity, which is driven by a desire to align with group expectations. This tendency is particularly evident under conditions of uncertainty and when the robot is perceived as competent, and appears to be amplified in hybrid human–robot groups [302, 178]. Late-occurring errors have been found to exert a persistent negative effect on trust [6], whereas opportunities for interactive teaching can help to buffer against and even rebuild trust over time [55, 56].

Taken together, these findings, reinforced by the classroom results in Chapter 4, emphasise that the effectiveness of social learning systems depends not only on their technical capabilities, but also on how they are interpreted and responded to as social partners [24]. When children see their feedback visibly integrated into the robot's behaviour, they may experience greater engagement and a stronger sense of responsibility, which can support deeper learning [46, 90, 263]. Conversely, when the robot's responses are inconsistent or its learning process remains opaque, this can lead to frustration, disengagement, or misplaced trust [237]. Designing systems that balance agency, clarity, and reliability is therefore critical to maximising both the cognitive and socio-emotional benefits of teachable robots, and should be prioritised in future system development and classroom integration.

8.4 Limitations

While this thesis offers several contributions to the field of educational robotics, limitations should be acknowledged, both in study design and in the broader constraints of deploying learning robots in real-world educational settings. These limitations do not detract from the contributions of the thesis, but they do outline the conditions under which the findings should be interpreted.

Single-school context. The main classroom study reported in Chapter 4 was conducted in one UK primary school, across two classrooms with two teachers. This design was intended to reduce teacher-specific variance, yet it remains unclear whether the findings would generalise to other schools, particularly those differing in socioeconomic profile, curriculum priorities, language background, or cultural context. Replication across diverse sites would be required to assess the robustness and transferability of the observed effects, as single-site studies often lack external validity for broader generalisation [112]

Short study duration. The classroom study in Chapter 4 was conducted within a single morning to minimise timetable disruption. Although immediate learning and a two-week retention measure were obtained, the design did not permit observation of longer term trajectories in learning, engagement, or trust. It is plausible that cognitive gains and the child–robot relationship would evolve under repeated exposure. Longitudinal studies are therefore needed to examine sustainability and cumulative impact of robotics peer in education [24].

Nature of the outcome measures. The assessment instruments used in Chapter 4 primarily captured task performance rather than deeper conceptual understanding or metacognitive growth. While improved retention indicates a learning benefit, the design did not allow fine grained analysis of the underlying cognitive processes [252]. In addition, self report questionnaires employed in Chapter 5 may be sensitive to response biases [213] and may reflect perceived engagement more than internalised change. Multi method approaches, combining behavioural traces, interaction logs, observational coding, and instructor or caregiver reports, could provide a more comprehensive account [288].

Robot platform constraints. To support parallel one to one interactions in the classroom, JD Humanoid robots were selected for affordability and portability. These platforms provide limited expressive and social channels compared to more advanced robots. As a result, the peer quality of the interaction may have been constrained, and children's perceptions of the robot's social presence and responsiveness may have been attenuated despite cognitive level adaptation [24, 111, 136].

Generalisation to child populations. In follow up studies on trust, ownership, and natural language feedback (Chapters 5 and 6), university students were recruited for reasons of logistics and experimental control. However, developmental research indicates that executive control and social–affective systems continue to mature through adolescence, implying that young adults differ from children in cognitive maturity and self regulation [43, 63]. There is also evidence that children interpret social cues and attribute agency to robots differently from adults, and that their selective trust in robotic informants depends on perceived accuracy and agency [125, 32]. Consequently, while these studies offer foundational insights, additional work is needed to determine whether the same psychological and interactional effects hold in child populations, and educational HRI reviews explicitly call for child focused and longitudinal evaluation to establish generalisability [24].

8.5 Ethical Considerations

The deployment of social robots in education brings both opportunities and responsibilities, particularly when such systems are designed to learn from children and adapt their behaviour in real time. As these technologies become more capable, their integration into classrooms must be aligned with ethical principles, pedagogical values, and emerging regulatory frameworks such as the EU AI Act [85].

Supervision and autonomy under the EU AI Act. The EU AI Act classifies AI systems interacting with children as high-risk, requiring that they be subject to effective human oversight at all times by a responsible adult [85]. While this provision functions as an important safeguard, it also raises practical and experiential questions. Continuous adult oversight, particularly if visible to the child, may alter the dynamic of learning-by-teaching, making the interaction feel monitored or evaluated rather than autonomous and peer-like. This consideration is particularly relevant to the classroom deployment described in Chapter 4, where part of the learning experience depended on children feeling in control of the teaching process. Balancing the legal requirement for supervision with the pedagogical value of independent, child-led interaction will be a key challenge for future deployments. Subtle supervision models that allow teachers to monitor safely without disrupting the child's sense of agency have been proposed as a potential solution [247, 254].

Complementing, not replacing, teachers. The work in this thesis does not aim to replace teachers, nor to position robots as primary sources of instruction. As noted in prior reviews [24, 244], robots are most effectively deployed as tools to support teachers in providing personalised practice, facilitating peer-like learning experiences, and offering alternative forms of engagement. In the classroom study described in Chapter 4, lessons were delivered by the class teachers, and the robot activity was used to consolidate learning through guided practice aligned with the curriculum and the lesson objectives. Teachers are central to guiding learning, interpreting progress, and ensuring that the robot's role complements the broader curriculum. Positioning robots as assistants rather than substitutes is essential to preserving the irreplaceable human elements of teaching, including empathy, adaptability, and contextual judgement.

Equity and access. While social robots can offer innovative learning opportunities, they also carry the risk of exacerbating educational inequalities. Not all schools, particularly in under-resourced contexts, will have the infrastructure or funding to deploy such systems at scale [270]. Without targeted strategies, this could create a divide in which only certain students have access to the benefits of human–robot interaction. This concern is particularly relevant to the feasibility considerations raised in Chapter 4, where the affordability of the JD Humanoid platform influenced its selection. Addressing this requires both technological strategies, such as the use of affordable, modular platforms, and policy-level interventions to promote equitable distribution.

Trust and performance perception. Findings from the involvement and trust study (Chapter 5) indicate that participants' trust in the robot was strongly linked to its perceived performance, consistent with previous work in educational HRI [32, 107]. In adults, performance-based trust can be a rational response to observed accuracy; however, it remains uncertain whether children can reliably assess a robot's competence. Prior studies have shown that children may base their trust on surface-level cues such as confidence or friendliness rather than correctness, which can lead to misplaced epistemic trust [277]. This underscores the ethical need to design transparency mechanisms that make the robot's learning state and reliability more visible, enabling children to critically evaluate its competence.

Social and emotional effects. Strong emotional bonds can be formed by children with robots, particularly when they are given responsibility for teaching them [263, 131]. Such engagement can be beneficial for motivation, but it also risks creating unrealistic expectations of the robot's cognitive or social capabilities. While this thesis implemented framing strategies to present the robot as a learning peer, more work is required to ensure that children maintain an accurate understanding of its limitations, especially during prolonged use, which may reinforce anthropomorphic perceptions [131].

Transparency and data protection. In the present work, no personal identifiers were stored, and all procedures complied with institutional ethical approval for research with children (Chapter 4). However, as adaptive learning systems expand, careful consideration will be needed regarding the type and amount of data collected, how it is stored, and who has access to it [84]. Ensuring compliance with data minimisation principles [84] and communicating transparently about the robot's capabilities and data practices will be essential for maintaining trust [265].

8.6 Summary

This chapter synthesised the thesis findings, showing that novice peer robots—driven by interactive reinforcement learning—can be integrated into real classrooms to support learning-by-teaching. The work combined developmental theory with a cognitive architecture capable of adapting to individual feedback, understanding natural language, and consolidating knowledge over time. Contributions include a validated application of interactive learning models, empirical evidence from ecologically valid deployments, and design insights for socially and pedagogically supportive interactions. Psychological outcomes such as trust, ownership, and involvement were discussed alongside limitations in scope, context, and platform capabilities. Ethical considerations highlighted regulatory compliance, equitable access, teacher–robot complementarity, and the need to manage trust and emotional engagement responsibly.

Chapter 9

Conclusion and Future Work

9.1 Summary of the Thesis

This thesis investigated whether robots could be designed as *teachable peers*, capable of learning from children in ways that support their own educational progress. The work approached this question in stages, beginning with controlled online investigations, moving into classroom studies in primary schools, extending to psychological explorations of involvement and trust, and finally turning to modelling frameworks inspired by natural language and sleep consolidation.

At the interaction level, an online study with adult participants examined how people perceived a robot that adapted through evaluative feedback. The results suggested that a teachable robot was viewed positively, with participants describing it as responsive and engaging. This initial evidence of perceived teachability informed the choice of Interactive RL as the underlying cognitive model. In classroom studies with children aged eight to nine years, this model was deployed at scale. Children were able to provide accurate feedback and adapt their tutoring strategies over time. Compared with matched self-practice, teaching a robot yielded higher delayed retention in a vocabulary task, encouraged longer and more strategic engagement, and led to greater gains for children with lower prior knowledge.

The thesis then turned to the psychological mechanisms that shape how teachable robots are experienced. A study of user involvement showed that greater participation increased self-investment, but did not automatically lead to stronger feelings of ownership or higher trust. Finally, two modelling frameworks extended the scope of teachability. ECLAIR demonstrated how free-form natural language could be classified into evaluative, corrective, and guidance feedback and integrated into learning, while INFORM introduced a sleep-inspired consolidation process that allowed feedback-guided trajectories to be replayed offline for more robust and generalisable knowledge. While ECLAIR and INFORM were tested in adult or simulated contexts, both highlight how teachable robots can be extended beyond immediate, task-specific learning.

9.2 Main Contributions

This thesis makes three main contributions to the study of teachable peer robots:

1. **Ecological validation of Interactive RL in real classrooms.** A large-scale study deployed an Interactive RL architecture in primary schools, respecting the natural conditions of the classroom. Children taught the robot using evaluative feedback with multiple interactions unfolded simultaneously. This provided ecological evidence that teachable robots can operate in authentic educational settings, where they improved children's delayed retention, encouraged strategic engagement, and particularly benefitted learners with lower prior knowledge.

2. **A psychologically grounded study of involvement, ownership, and trust.** A controlled experiment investigated how varying levels of involvement in teaching shaped user perceptions when multiple robots were present. The results showed that greater involvement increased self-investment but did not automatically enhance ownership or global trust. These findings align with what we would want in educational settings: to support appropriate trust, ensuring that learners place confidence in the robot when it is reliable, but remain able to question or correct it when it is not.

3. **Two modelling frameworks that extend teachability.** ECLAIR introduced a natural language interface that classified free-form advice into

evaluative, corrective, and guidance feedback and integrated each type through tailored shaping strategies. INFORM added a sleep-inspired consolidation mechanism that replayed feedback-guided trajectories offline, recovering reward structure and improving robustness. Together, these frameworks extend teachable robots beyond single-modality, short-sighted learning.

9.3 Future Work

The work presented in this thesis opens several avenues for future research.

First, it would be interesting to is integrate the proposed frameworks into a single cohesive cognitive peer model . While Interactive RL, ECLAIR, and INFORM were each evaluated separately, the next step is to bring them together into a single pipeline that cycles between teaching, and offline consolidation. This would allow a robot to learn from evaluative and multimodal input in real time, and then continue refining its knowledge between interactions..

A second direction is about the deployment and scalability of the studies. The classroom studies in this thesis demonstrated that teachable robots can improve retention and engagement in short-term settings. Future research should extend this to longitudinal, multi-site deployments that follow learners across weeks or months, testing how durable the benefits of learning by teaching are, and whether they scale across diverse schools and populations. In particular, stratified analyses will be important to understand equity impacts for learners with lower prior knowledge or different language backgrounds.

A third strand concerns personalisation and multimodality. The studies here treated children largely as a uniform group, but learners differ in their preferred feedback strategies and in how they engage with robots. Future work could model these differences, adapting the robot's behaviour to the child's style and state in ways that remain flexible but avoid narrowing learners into fixed profiles. Beyond text and speech, multimodal inputs such as gestures, emotions, or prosody could provide richer and more natural feedback signals.

Another key challenge is trust, safety, and evaluation. The involvement study showed that more participation does not always mean more trust, and that appropriate trust is the more important goal. Future work should explore design strategies that help learners place confidence in robots when they are reliable, but still feel able to question them when needed. This also calls for better ways of assessing epistemic trust and learning outcomes beyond self-report, such as behavioural indicators, lightweight transfer tests, or privacy-preserving multimodal measures.

Finally, future research should explore whether teachable robots can also support conceptual understanding, open-ended explanation, or collaborative problem solving. This may involve extending consolidation mechanisms such as INFORM to underspecified tasks where the child's intent must be inferred, or varying embodiment and classroom topology to study how social presence mediates the effectiveness of teaching.

Taken together, these directions outline a path toward teachable robots that are technically robust, educationally relevant, and psychologically safe. They emphasise not only the importance of building systems that learn effectively, but also of embedding them responsibly in the complex social and developmental environments of classrooms.

9.4 Final Reflection

The central claim of this thesis is that making the robot a better learner also makes the child a better learner. By prioritising contingency and transparency in classroom studies, examining the psychological dynamics of involvement and trust, and extending teachability through language and consolidation, the work has taken steps toward designing robots that can genuinely be taught.

The vision that emerges is not of robots that replace teachers, but of robots that grow alongside children as responsive and imperfect peers. If the strands developed here can be woven together, a robot might one day learn overnight what a child has taught it during the day, and return not only better prepared to be taught again, but also better able to support the child's own learning journey.

9.4 Final Reflection

Taken together, the findings point toward tangible ways in which robots could be woven into everyday classroom practice. Robots could serve as peer-like companions during individual practice, offering a dynamic and engaging alternative to worksheets or software. By learning directly from children's feedback, they could adapt their behaviour to different knowledge levels, scaffolding weaker learners while still challenging stronger ones. In group contexts, robots could act as shared partners in collaborative tasks, prompting reflection, encouraging explanation, and modelling perseverance. Importantly, their design as novice learners positions them not as authorities but as tools that stimulate dialogue, critical thinking, and self-explanation. This points toward a future where social robots are integrated not as replacements for teachers, but as complementary supports that help diversify classroom activities, personalise learning experiences, and foster deeper engagement.

References

[1] Ahmad, M. I., Bernotat, J., Lohan, K., and Eyssel, F. (2019). Trust and cognitive load during human-robot interaction. *arXiv preprint arXiv:1909.05160*.

[2] Aitsam, M., Lacroix, D., Goyal, G., Bartolozzi, C., and Di Nuovo, A. (2024). Measuring cognitive load through event camera based human-pose estimation. In *International Workshop on Human-Friendly Robotics*, pages 229–239. Springer.

[3] Alemi, M. and Haeri, N. (2020). Robot-assisted instruction of l2 pragmatics: Effects on young efl learners' speech act performance.

[4] Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., and Herrera, F. (2023). Explainable artificial intelligence (xai): What we know and what is left to attain trustworthy artificial intelligence. *Information fusion*, 99:101805.

[5] Ali, S., Moroso, T., and Breazeal, C. (2019). Can children learn creativity from a social robot? In *Proceedings of the 2019 Conference on Creativity and Cognition*, pages 359–368.

[6] Aliasghari, P., Ghafurian, M., Nehaniv, C. L., and Dautenhahn, K. (2023). How do we perceive our trainee robots? exploring the impact of robot errors and appearance when performing domestic physical tasks on teachers' trust and evaluations. *ACM Transactions on Human-Robot Interaction*, 12(3):1–41.

[7] Alimardani, M., Duret, J., Jouen, A.-L., and Hiraki, K. (2023). Social robots as effective language tutors for children: empirical evidence from neuroscience. *Frontiers in Neurorobotics*, 17:1260999.

[8] Andrist, S., Spannan, E., and Mutlu, B. (2013). Rhetorical robots: making robots more effective speakers using linguistic cues of expertise. In *2013 8th ACM/IEEE international Conference on human-robot interaction (HRI)*, pages 341–348. IEEE.

[9] Angelopoulos, G., Mangiacapra, L., Rossi, A., Di Napoli, C., and Rossi, S. (2025). What is behind the curtain? increasing transparency in reinforcement learning with human preferences and explanations. *Engineering Applications of Artificial Intelligence*, 149:110520.

[10] Arakawa, R., Kobayashi, S., Unno, Y., Tsuboi, Y., and Maeda, S.-i. (2018). Dqn-tamer: Human-in-the-loop reinforcement learning with intractable feedback. *arXiv preprint arXiv:1810.11748*.

[11] Argall, B. D., Chernova, S., Veloso, M., and Browning, B. (2009). A survey of robot learning from demonstration. *Robotics and autonomous systems*, 57(5):469–483.

[12] Aron, A., Aron, E. N., and Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of personality and social psychology*, 63(4):596.

[13] Babamiri, M., Heidarimoghadam, R., Ghasemi, F., Tapak, L., and Mortezapour, A. (2022). Insights into the relationship between usability and willingness to use a robot in the future workplaces: Studying the mediating role of trust and the moderating roles of age and stara. *PloS one*, 17(6):e0268942.

[14] Bai, Y., Kadavath, S., Kundu, S., Askell, A., Kernion, J., Jones, A., Chen, A., Goldie, A., Mirhoseini, A., McKinnon, C., et al. (2022). Constitutional ai: Harmlessness from ai feedback. *arXiv preprint arXiv:2212.08073*.

[15] Bain, M. and Sammut, C. (1995). A framework for behavioural cloning. In *Machine intelligence 15*, pages 103–129.

[16] Bandura, A. et al. (1986). Social foundations of thought and action. *Englewood Cliffs, NJ*, 1986(23-28):2.

[17] Bartneck, C., Kanda, T., Ishiguro, H., and Hagita, N. (2007). Is the uncanny valley an uncanny cliff? In *RO-MAN 2007-The 16th IEEE international symposium on robot and human interactive communication*, pages 368–373. IEEE.

[18] Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1:71–81.

[19] Bass, I., Bonawitz, E., Shafto, P., Ramarajan, D., Gopnik, A., and Wellman, H. (2017). I know what you need to know: Children's developing theory of mind andpedagogical evidence selection. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 39.

[20] Baxter, P., Ashurst, E., Read, R., Kennedy, J., and Belpaeme, T. (2017). Robot education peers in a situated primary school study: Personalisation promotes child learning. *PloS one*, 12(5):e0178126.

[21] Baxter, W. L., Aurisicchio, M., and Childs, P. R. (2015). A psychological ownership approach to designing object attachment. *Journal of Engineering Design*, 26(4-6):140–156.

[22] Beer, J. M., Fisk, A. D., and Rogers, W. A. (2014). Toward a framework for levels of robot autonomy in human-robot interaction. *Journal of human-robot interaction*, 3(2):74.

[23] Belanche, D., Casaló, L. V., and Flavián, C. (2012). Integrating trust and personal values into the technology acceptance model: The case of e-government services adoption. *Cuadernos de Economía y Dirección de la Empresa*, 15(4):192–204.

[24] Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: A review. *Science robotics*, 3(21):eaat5954.

[25] Biswas, G., Leelawong, K., Schwartz, D., Vye, N., and at Vanderbilt, T. T. A. G. (2005). Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence*, 19(3-4):363–392.

[26] Biyik, E. and Sadigh, D. (2018). Batch active preference-based learning of reward functions. In *Conference on robot learning*, pages 519–528. PMLR.

[27] Bjork, E. L., Bjork, R. A., et al. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society*, 2(59-68).

[28] Blut, M., Wang, C., Wünderlich, N. V., and Brock, C. (2021). Understanding anthropomorphism in service provision: a meta-analysis of physical robots, chatbots, and other ai. *Journal of the academy of marketing science*, 49(4):632–658.

[29] Bourguet, M.-L., Jin, Y., Shi, Y., Chen, Y., Rincon-Ardila, L., and Venture, G. (2020). Social robots that can sense and improve student engagement. In *2020 ieee international conference on teaching, assessment, and learning for engineering (tale)*, pages 127–134. IEEE.

[30] Bowman-Smith, C. K., Aitken, C., Mahenthiran, T., Law, E., and Nilsen, E. S. (2025). Teaching social robots: The effect of robot mistakes on children's learning-through-teaching. *Frontiers in Developmental Psychology*, 3:1526486.

[31] Bozkurt, A. (2023). Generative artificial intelligence (ai) powered conversational educational agents: The inevitable paradigm shift. *Asian Journal of Distance Education*, 18(1).

[32] Brink, K. A. and Wellman, H. M. (2020). Robot teachers for children? young children trust robots depending on their perceived accuracy and agency. *Developmental Psychology*, 56(7):1268.

[33] Brodt, S., Inostroza, M., Niethard, N., and Born, J. (2023). Sleep—a brain-state serving systems memory consolidation. *Neuron*, 111(7):1050–1075.

[34] Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard university press.

[35] Brosseau-Liard, P., Penney, D., and Poulin-Dubois, D. (2015). Theory of mind selectively predicts preschoolers' knowledge-based selective word learning. *British Journal of Developmental Psychology*, 33(4):464–475.

[36] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. (2020). Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901.

[37] Bruner, J. (1985). Child's talk: Learning to use language. *Child Language Teaching and Therapy*, 1(1):111–114.

[38] Bruzzo, D., Matarese, M., Sciutti, A., and Rea, F. (2024). Charm or harm? how social robotic tutors influence people's learning with correct and incorrect guidance. In *International Conference on Social Robotics*, pages 475–487. Springer.

[39] Burnett Heyes, S., Jih, Y.-R., Block, P., Hiu, C.-F., Holmes, E. A., and Lau, J. Y. (2015). Relationship reciprocation modulates resource allocation in adolescent social networks: developmental effects. *Child development*, 86(5):1489–1506.

[40] Campbell, C., Tanzer, M., Saunders, R., Booker, T., Allison, E., Li, E., O'Dowda, C., Luyten, P., and Fonagy, P. (2021). Development and validation of a self-report measure of epistemic trust. *PloS one*, 16(4):e0250264.

[41] Cantucci, F., Marini, M., and Falcone, R. (2025). The role of robot competence, autonomy, and personality on trust formation in human-robot interaction. *arXiv preprint arXiv:2503.04296*.

[42] Carta, T., Oudeyer, P.-Y., Sigaud, O., and Lamprier, S. (2022). Eager: Asking and answering questions for automatic reward shaping in language-guided rl. *Advances in Neural Information Processing Systems*, 35:12478–12490.

[43] Casey, B. J., Getz, S., and Galvan, A. (2008). The adolescent brain. *Developmental review*, 28(1):62–77.

[44] Celemin, C. and Ruiz-del Solar, J. (2019). An interactive framework for learning continuous actions policies based on corrective feedback. *Journal of Intelligent & Robotic Systems*, 95:77–97.

[45] Chandra, S., Dillenbourg, P., and Paiva, A. (2020). Children teach handwriting to a social robot with different learning competencies. *International Journal of Social Robotics*, 12(3):721–748.

[46] Chase, C. C., Chin, D. B., Oppezzo, M. A., and Schwartz, D. L. (2009). Teachable agents and the protégé effect: Increasing the effort towards learning. *Journal of science education and technology*, 18:334–352.

[47] Chen, H., Alghowinem, S., Breazeal, C., and Park, H. W. (2024). Integrating flow theory and adaptive robot roles: a conceptual model of dynamic robot role adaptation for the enhanced flow experience in long-term multi-person human-robot interactions. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 116–126.

[48] Chen, H., Park, H. W., and Breazeal, C. (2020a). Teaching and learning with children: Impact of reciprocal peer learning with a social robot on children's learning and emotive engagement. *Computers & Education*, 150:103836.

[49] Chen, H., Park, H. W., Zhang, X., and Breazeal, C. (2020b). Impact of interaction context on the student affect-learning relationship in child-robot interaction. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, pages 389–397.

[50] Chetouani, M. (2021). Interactive robot learning: an overview. *ECCAI Advanced Course on Artificial Intelligence*, pages 140–172.

[51] Chevalère, J., Kirtay, M., Hafner, V. V., and Lazarides, R. (2023). Who to observe and imitate in humans and robots: the importance of motivational factors. *International Journal of Social Robotics*, 15(8):1265–1275.

[52] Chi, M. T. (2009). Active-constructive-interactive: A conceptual framework for differentiating learning activities. *Topics in cognitive science*, 1(1):73–105.

[53] Chi, M. T., De Leeuw, N., Chiu, M.-H., and LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science*, 18(3):439–477.

[54] Chi, M. T., Siler, S. A., Jeong, H., Yamauchi, T., and Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive science*, 25(4):471–533.

[55] Chi, V. B. and Malle, B. F. (2023). People dynamically update trust when interactively teaching robots. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 554–564.

[56] Chi, V. B. and Malle, B. F. (2024). Interactive human-robot teaching recovers and builds trust, even with imperfect learners. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 127–136.

[57] Chisari, E., Welschehold, T., Boedecker, J., Burgard, W., and Valada, A. (2022). Correct me if i am wrong: Interactive learning for robotic manipulation. *IEEE Robotics and Automation Letters*, 7(2):3695–3702.

[58] Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., and Amodei, D. (2017). Deep reinforcement learning from human preferences. *Advances in neural information processing systems*, 30.

[59] Chu, M. D., Gerard, P., Pawar, K., Bickham, C., and Lerman, K. (2025). Illusions of intimacy: Emotional attachment and emerging psychological risks in human-ai relationships. *arXiv preprint arXiv:2505.11649*.

[60] Cloward, R. D. (1967). Studies in tutoring. *The Journal of Experimental Education*, 36(1):14–25.

[61] Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

[62] Colvin, J. W. (2007). Peer tutoring and social dynamics in higher education. *Mentoring & tutoring*, 15(2):165–181.

[63] Crone, E. A. and Dahl, R. E. (2012). Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nature reviews neuroscience*, 13(9):636–650.

[64] Cruz, F., Twiefel, J., Magg, S., Weber, C., and Wermter, S. (2015). Interactive reinforcement learning through speech guidance in a domestic scenario. In *2015 international joint conference on neural networks (IJCNN)*, pages 1–8. IEEE.

[65] Cucciniello, I., Sangiovanni, S., Maggi, G., and Rossi, S. (2023). Mind perception in hri: Exploring users' attribution of mental and emotional states to robots with different behavioural styles. *International Journal of Social Robotics*, 15(5):867–877.

[66] Dalangin, B., Gordon, S., and Roy, H. (2024). Positive interactions with intelligent technology through psychological ownership: A human-in-the-loop approach. *Artificial Intelligence and Social Computing*, page 62.

[67] Das, R. and Pon-Barry, H. (2018). Turn-taking strategies for human-robot peer-learning dialogue. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 119–129.

[68] Dawkins, S., Tian, A. W., Newman, A., and Martin, A. (2017). Psychological ownership: A review and research agenda. *Journal of organizational behavior*, 38(2):163–183.

[69] De Wit, J., Schodde, T., Willemsen, B., Bergmann, K., De Haas, M., Kopp, S., Krahmer, E., and Vogt, P. (2018). The effect of a robot's gestures and adaptive tutoring on children's acquisition of second language vocabularies. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, pages 50–58.

[70] Deci, E. L. and Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4):227–268.

[71] Delgosha, M. S. and Hajiheydari, N. (2021). How human users engage with consumer robots? a dual model of psychological ownership and trust to explain post-adoption behaviours. *Computers in Human Behavior*, 117:106660.

[72] Derakhshan, A., Teo, T., and Khazaie, S. (2025). Investigating the usefulness of artificial intelligence-driven robots in developing empathy for english for medical purposes communication: The role-play of asian and african students. *Computers in Human Behavior*, 162:108416.

[73] Doise, W. (1984). On the social development of the intellect. In *The future of Piagetian theory: The neo-Piagetians*, pages 95–121. Springer.

[74] Donnellan, M. B., Oswald, F. L., Baird, B. M., and Lucas, R. E. (2006). The mini-ipip scales: tiny-yet-effective measures of the big five factors of personality. *Psychological assessment*, 18(2):192.

[75] Donnermann, M. and Lugrin, B. (2024). Integration of robot-supported tutoring in higher education-an empirically based concept. In *Proceedings of the 2024 the 16th International Conference on Education Technology and Computers*, pages 1–7.

[76] Donnermann, M., Schaper, P., and Lugrin, B. (2022). Social robots in applied settings: A long-term study on adaptive robotic tutors in higher education. *Frontiers in Robotics and AI*, 9:831633.

[77] Du, Y., Watkins, O., Wang, Z., Colas, C., Darrell, T., Abbeel, P., Gupta, A., and Andreas, J. (2023). Guiding pretraining in reinforcement learning with large language models. In *International Conference on Machine Learning*, pages 8657–8677. PMLR.

[78] Dubois-Sage, M., Jacquet, B., Jamet, F., and Baratgin, J. (2023). We do not anthropomorphize a robot based only on its cover: context matters too! *Applied Sciences*, 13(15):8743.

[79] Duran, D. (2017). Learning-by-teaching. evidence and implications as a pedagogical mechanism. *Innovations in education and teaching international*, 54(5):476–484.

[80] Edwards, B. I., Muniru, I. O., Khougali, N., Cheok, A. D., and Prada, R. (2018). A physically embodied robot teacher (pert) as a facilitator for peer learning. In *2018 IEEE frontiers in education conference (FIE)*, pages 1–9. IEEE.

[81] Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational research review*, 1(1):3–14.

[82] El Hamamsy, L., Johal, W., Asselborn, T., Nasir, J., and Dillenbourg, P. (2019). Learning by collaborative teaching: an engaging multi-party cowriter activity. In *2019 28th IEEE international conference on robot and human interactive communication (RO-MAN)*, pages 1–8. IEEE.

[83] Epley, N., Waytz, A., and Cacioppo, J. T. (2007). On seeing human: a three-factor theory of anthropomorphism. *Psychological review*, 114(4):864.

[84] European Parliament and Council (2016). Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation). Official Journal of the European Union, L 119. CELEX: 32016R0679.

[85] European Parliament and Council (2024). Artificial intelligence act. Official Journal of the European Union. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending certain Union legislative acts.

[86] Eysenbach, B. and Levine, S. (2021). Maximum entropy rl (provably) solves some robust rl problems. *arXiv preprint arXiv:2103.06257*.

[87] Eyssel, F., Kuchenbrandt, D., Bobinger, S., De Ruiter, L., and Hegel, F. (2012). 'if you sound like me, you must be more human' on the interplay of robot and user features on human-robot acceptance and anthropomorphism. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 125–126.

[88] Fang, Z. and Sims, C. R. (2025). Humans learn generalizable representations through efficient coding. *Nature Communications*, 16(1):3989.

[89] Fernando, S., Moore, R. K., Cameron, D., Collins, E. C., Millings, A., Sharkey, A. J., and Prescott, T. J. (2016). Automatic recognition of child speech for robotic applications in noisy environments. *arXiv preprint arXiv:1611.02695*.

[90] Fiorella, L. and Mayer, R. E. (2013). The relative benefits of learning by teaching and teaching expectancy. *Contemporary Educational Psychology*, 38(4):281–288.

[91] Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American psychologist*, 34(10):906.

[92] Fleckney, P., Thompson, J., and Vaz-Serra, P. (2025). Designing effective peer assessment processes in higher education: a systematic review. *Higher Education Research & Development*, 44(2):386–401.

[93] Franke, T., Attig, C., and Wessel, D. (2019). A personal resource for technology interaction: development and validation of the affinity for technology interaction (ati) scale. *International Journal of Human–Computer Interaction*, 35(6):456–467.

[94] Fredricks, J. A., Blumenfeld, P. C., and Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*, 74(1):59–109.

[95] Gagne, R. M. (1968). Contributions of learning to human development. *Psychological review*, 75(3):177.

[96] Gao, Y., Barendregt, W., Obaid, M., and Castellano, G. (2018). When robot personalisation does not help: Insights from a robot-supported learning study. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 705–712. IEEE.

[97] Garg, D., Chakraborty, S., Cundy, C., Song, J., and Ermon, S. (2021). Iq-learn: Inverse soft-q learning for imitation. *Advances in Neural Information Processing Systems*, 34:4028–4039.

[98] Geiskkovitch, D. Y., Thiessen, R., Young, J. E., and Glenwright, M. R. (2019). What? that's not a chair!: How robot informational errors affect children's trust towards robots. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 48–56. IEEE.

[99] Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., and Danks, D. (2004). A theory of causal learning in children: causal maps and bayes nets. *Psychological review*, 111(1):3.

[100] Gordon, G., Spaulding, S., Westlund, J. K., Lee, J. J., Plummer, L., Martinez, M., Das, M., and Breazeal, C. (2016). Affective personalization of a social robot tutor for children's second language skills. In *Proceedings of the AAAI conference on artificial intelligence*, volume 30.

[101] Gouaillier, D., Hugel, V., Blazevic, P., Kilner, C., and Monceaux, J. (2009). Mechatronic design of nao humanoid. In *2009 IEEE International Conference on Robotics and Automation*, pages 769–774. IEEE.

[102] Goyal, P., Niekum, S., and Mooney, R. J. (2019). Using natural language for reward shaping in reinforcement learning. *arXiv preprint arXiv:1903.02020*.

[103] Gray, H. M., Gray, K., and Wegner, D. M. (2007). Dimensions of mind perception. *science*, 315(5812):619–619.

[104] Griffith, S., Subramanian, K., Scholz, J., Isbell, C. L., and Thomaz, A. L. (2013). Policy shaping: Integrating human feedback with reinforcement learning. *Advances in neural information processing systems*, 26.

[105] Güroğlu, B., van den Bos, W., and Crone, E. A. (2014). Sharing and giving across adolescence: An experimental study examining the development of prosocial behavior. *Frontiers in Psychology*, 5:291.

[106] Haarnoja, T., Zhou, A., Abbeel, P., and Levine, S. (2018). Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, pages 1861–1870. PMLR.

[107] Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., and Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, 53(5):517–527.

[108] Hancock, P. A., Kessler, T. T., Kaplan, A. D., Brill, J. C., and Szalma, J. L. (2021). Evolving trust in robots: specification through sequential and comparative meta-analyses. *Human factors*, 63(7):1196–1229.

[109] Hart, S. G. and Staveland, L. E. (1988). Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier.

[110] Hay, D. F., Payne, A., and Chadwick, A. (2004). Peer relations in childhood. *Journal of child psychology and psychiatry*, 45(1):84–108.

[111] He, K., Chan, W. P., Cosgun, A., Joy, A., and Croft, E. A. (2024). Robot gaze during autonomous navigation and its effect on social presence. *International Journal of Social Robotics*, 16(5):879–897.

[112] Hedges, L. V. (2013). Recommendations for practice: Justifying claims of generalizability. *Educational Psychology Review*, 25(3):331–337.

[113] Hei, X., Zhang, H., and Tapus, A. (2024). Exploring help-seeking behavior, performance, and cognitive load in individual tutoring: A comparative study between human tutors and social robots. In *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*, pages 1035–1041. IEEE.

[114] Ho, M. K., Cushman, F., Littman, M. L., and Austerweil, J. L. (2019). People teach with rewards and punishments as communication, not reinforcements. *Journal of Experimental Psychology: General*, 148(3):520.

[115] Ho, M. K. and Griffiths, T. L. (2022). Cognitive science as a source of forward and inverse models of human decisions for robotics and control. *Annual Review of Control, Robotics, and Autonomous Systems*, 5:33–53.

[116] Ho, M. K., Littman, M. L., Cushman, F., and Austerweil, J. L. (2015). Teaching with rewards and punishments: Reinforcement or communication? In *CogSci*, volume 3, page 3.

[117] Ho, M. K., MacGlashan, J., Littman, M. L., and Cushman, F. (2017). Social is special: A normative framework for teaching with and learning from evaluative feedback. *Cognition*, 167:91–106.

[118] Howe, N., Della Porta, S., Recchia, H., Funamoto, A., and Ross, H. (2015). “this bird can’t do it ‘cause this bird doesn’t swim in water”: Sibling teaching during naturalistic home observations in early childhood. *Journal of Cognition and Development*, 16(2):314–332.

[119] Huang, S., Dossa, R. F. J., Ye, C., Braga, J., Chakraborty, D., Mehta, K., and Araújo, J. G. (2022). Cleanrl: High-quality single-file implementations of deep reinforcement learning algorithms. *Journal of Machine Learning Research*, 23(274):1–18.

[120] Huber, R., Felice Ghilardi, M., Massimini, M., and Tononi, G. (2004). Local sleep and learning. *Nature*, 430(6995):78–81.

[121] Immordino-Yang, M. H. and Damasio, A. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, brain, and education*, 1(1):3–10.

[122] Ishida, M., Khalifa, A., Kato, T., and Yamamoto, S. (2016). Features of learner corpus collected with joining-in type robot assisted language learning system. In *2016 Conference of The Oriental Chapter of International Committee for Coordination and Standardization of Speech Databases and Assessment Techniques (O-COCOSDA)*, pages 128–131. IEEE.

[123] Jacq, A., Lemaignan, S., Garcia, F., Dillenbourg, P., and Paiva, A. (2016). Building successful long child-robot interactions in a learning context. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 239–246. IEEE.

[124] Jamet, F., Masson, O., Jacquet, B., Stilgenbauer, J.-L., and Baratgin, J. (2018). Learning by teaching with humanoid robot: a new powerful experimental tool to improve children's learning ability. *Journal of Robotics*, 2018:1–11.

[125] Jipson, J. L. and Gelman, S. A. (2007). Robots and rodents: Children's inferences about living and nonliving kinds. *Child development*, 78(6):1675–1688.

[126] Johal, W., Jacq, A., Paiva, A., and Dillenbourg, P. (2016). Child-robot spatial arrangement in a learning by teaching activity. In *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pages 533–538. Ieee.

[127] Jones, A., Bull, S., and Castellano, G. (2018). "i know that now, i'm going to learn this next" promoting self-regulated learning with a robotic tutor. *International Journal of Social Robotics*, 10(4):439–454.

[128] Jones, A., Castellano, G., and Bull, S. (2014). Investigating the effect of a robotic tutor on learner perception of skill based feedback. In *Social Robotics: 6th International Conference, ICSR 2014, Sydney, NSW, Australia, October 27-29, 2014. Proceedings 6*, pages 186–195. Springer.

[129] Jussila, I., Tarkiainen, A., Sarstedt, M., and Hair, J. F. (2015). Individual psychological ownership: Concepts, evidence, and implications for research in marketing. *Journal of marketing theory and practice*, 23(2):121–139.

[130] Kahn Jr, P. H., Gary, H. E., and Shen, S. (2013). Children's social relationships with current and near-future robots. *Child Development Perspectives*, 7(1):32–37.

[131] Kahn Jr, P. H., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., Ruckert, J. H., and Shen, S. (2012). "robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental psychology*, 48(2):303.

[132] Kanda, T., Sato, R., Saiwaki, N., and Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term human–robot interaction. *IEEE Transactions on robotics*, 23(5):962–971.

[133] Kaufmann, T., Weng, P., Bengs, V., and Hüllermeier, E. (2023). A survey of reinforcement learning from human feedback. *arXiv preprint arXiv:2312.14925*, 10.

[134] Kennedy, J., Baxter, P., and Belpaeme, T. (2015a). Comparing robot embodiments in a guided discovery learning interaction with children. *International Journal of Social Robotics*, 7:293–308.

[135] Kennedy, J., Baxter, P., and Belpaeme, T. (2015b). The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*, pages 67–74.

[136] Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2015c). Higher nonverbal immediacy leads to greater learning gains in child-robot tutoring interactions. In *International conference on social robotics*, pages 327–336. Springer.

[137] Keppell, M., Au, E., Ma, A., and Chan, C. (2006). Peer learning and learning-oriented assessment in technology-enhanced environments. *Assessment & Evaluation in Higher Education*, 31(4):453–464.

[138] Kidd, C. D. and Breazeal, C. (2004). Effect of a robot on user perceptions. In *2004 IEEE/RSJ international conference on intelligent robots and systems (IROS)(IEEE Cat. No. 04CH37566)*, volume 4, pages 3559–3564. IEEE.

[139] Kiesler, T. and Kiesler, S. (2005). My pet rock and me: An experimental exploration of the self extension concept.

[140] Kim, Y. and Tscholl, M. (2021). Young children's embodied interactions with a social robot. *Educational Technology Research and Development*, 69(4):2059–2081.

[141] Klissarov, M., D'Oro, P., Sodhani, S., Raileanu, R., Bacon, P.-L., Vincent, P., Zhang, A., and Henaff, M. (2023). Motif: Intrinsic motivation from artificial intelligence feedback. *arXiv preprint arXiv:2310.00166*.

[142] Knox, W. B. and Stone, P. (2009). Interactively shaping agents via human reinforcement: The tamer framework. In *Proceedings of the fifth international conference on Knowledge capture*, pages 9–16.

[143] Knox, W. B. and Stone, P. (2012). Reinforcement learning from simultaneous human and mdp reward. In *AAMAS*, volume 1004, pages 475–482. Valencia.

[144] Knox, W. B. and Stone, P. (2015). Framing reinforcement learning from human reward: Reward positivity, temporal discounting, episodicity, and performance. *Artificial Intelligence*, 225:24–50.

[145] Koenig, M. A., Clément, F., and Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological science*, 15(10):694–698.

[146] Koenig, M. A. and Harris, P. L. (2005). Preschoolers mistrust ignorant and inaccurate speakers. *Child development*, 76(6):1261–1277.

[147] Koh, A. W. L., Lee, S. C., and Lim, S. W. H. (2018). The learning benefits of teaching: A retrieval practice hypothesis. *Applied Cognitive Psychology*, 32(3):401–410.

[148] Koriat, A. et al. (2006). *Metacognition and consciousness*. Institute of Information Processing and Decision Making, University of Haifa ....

[149] Kory-Westlund, J. M. and Breazeal, C. (2019). A long-term study of young children's rapport, social emulation, and language learning with a peer-like robot playmate in preschool. *Frontiers in Robotics and AI*, 6:81.

[150] Köse, H., Uluer, P., Akalın, N., Yorgancı, R., Özkul, A., and Ince, G. (2015). The effect of embodiment in sign language tutoring with assistive humanoid robots. *International Journal of Social Robotics*, 7(4):537–548.

[151] Kraus, J. M., Merger, J., Gröner, F., and Pätz, J. (2023). 'sorry'says the robot: The tendency to anthropomorphize and technology affinity affect trust in repair strategies after error. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 436–441.

[152] Krening, S. and Feigh, K. M. (2018). Interaction algorithm effect on human experience with reinforcement learning. *ACM Transactions on Human-Robot Interaction (THRI)*, 7(2):1–22.

[153] Kruglanski, A. W. and Webster, D. M. (1996). Motivated closing of the mind:" seizing" and" freezing.". *Psychological review*, 103(2):263.

[154] Lacroix, D., Landwehr, M., Wullenkord, R., Eyssel, F., and Augustine, A. (2024). Off my chest with my robot? the influence of psychological ownership on self-disclosure to a robot. In *International Conference on Social Robotics*, pages 497–506. Springer.

[155] Lacroix, D., Schober, J., Wullenkord, R., and Eyssel, F. (2023a). Pimp my language! the influence of robot customization duration on psychological ownership and trust. In *2023 32nd IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1908–1913. IEEE.

[156] Lacroix, D., Wullenkord, R., and Eyssel, F. (2022). I designed it, so i trust it: the influence of customization on psychological ownership and trust toward robots. In *International Conference on Social Robotics*, pages 601–614. Springer.

[157] Lacroix, D., Wullenkord, R., and Eyssel, F. (2023b). Who's in charge? using personalization vs. customization distinction to inform hri research on adaptation to users. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 580–586.

[158] Lemaignan, S., Jacq, A., Hood, D., Garcia, F., Paiva, A., and Dillenbourg, P. (2016). Learning by teaching a robot: The case of handwriting. *IEEE Robotics & Automation Magazine*, 23(2):56–66.

[159] Lewis, J. D. and Weigert, A. (1985). Trust as a social reality. *Social forces*, 63(4):967–985.

[160] Leyzberg, D., Ramachandran, A., and Scassellati, B. (2018). The effect of personalization in longer-term robot tutoring. *ACM Transactions on Human-Robot Interaction (THRI)*, 7(3):1–19.

[161] Leyzberg, D., Spaulding, S., and Scassellati, B. (2014). Personalizing robot tutors to individuals' learning differences. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, pages 423–430.

[162] Leyzberg, D., Spaulding, S., Toneva, M., and Scassellati, B. (2012). The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the annual meeting of the cognitive science society*, volume 34.

[163] Li, G., Dibeklioğlu, H., Whiteson, S., and Hung, H. (2020). Facial feedback for reinforcement learning: a case study and offline analysis using the tamer framework. *Autonomous Agents and Multi-Agent Systems*, 34(1):22.

[164] Li, J. (2015). The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *International Journal of Human-Computer Studies*, 77:23–37.

[165] Ligthart, M. E., De Droog, S. M., Bossema, M., Elloumi, L., Hoogland, K., Smakman, M. H., Hindriks, K. V., and Ben Allouch, S. (2023). Design specifications for a social robot math tutor. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 321–330.

[166] Liles, K. R. (2019). Ms. an (m eeting s tudents' a cademic n eeds): Engaging students in math education. In *Adaptive Instructional Systems: First International Conference, AIS 2019, Held as Part of the 21st HCI International Conference, HCII 2019, Orlando, FL, USA, July 26–31, 2019, Proceedings 21*, pages 645–661. Springer.

[167] Lin, X. and Tan, H. (2025). A systematic review of generative ai in k–12: Mapping goals, activities, roles, and outcomes via the 3p model. *Systems*, 13(10):840.

[168] Lind, J., Ghirlanda, S., and Enquist, M. (2019). Social learning through associative processes: a computational theory. *Royal Society open science*, 6(3):181777.

[169] Lindberg, M., Månsson, K., Johansson, B., Gulz, A., and Balkenius, C. (2017). Does a robot tutee increase children's engagement in a learning-by-teaching situation? In *Intelligent Virtual Agents: 17th International Conference, IVA 2017, Stockholm, Sweden, August 27-30, 2017, Proceedings 17*, pages 243–246. Springer.

[170] Lins de Holanda Coelho, G., HP Hanel, P., and J. Wolf, L. (2020). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment*, 27(8):1870–1885.

[171] Loftin, R., Peng, B., MacGlashan, J., Littman, M. L., Taylor, M. E., Huang, J., and Roberts, D. L. (2016). Learning behaviors via human-delivered discrete feedback: modeling implicit feedback strategies to speed up learning. *Autonomous agents and multi-agent systems*, 30:30–59.

[172] Lu, Y., Bartolo, M., Moore, A., Riedel, S., and Stenetorp, P. (2021). Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity. *arXiv preprint arXiv:2104.08786*.

[173] Luger, E. and Sellen, A. (2016). " like having a really bad pa" the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 5286–5297.

[174] Luketina, J., Nardelli, N., Farquhar, G., Foerster, J., Andreas, J., Grefenstette, E., Whiteson, S., and Rocktäschel, T. (2019). A survey of reinforcement learning informed by natural language. *arXiv preprint arXiv:1906.03926*.

[175] Maaz, N., Mounsef, J., and Maalouf, N. (2025). Care: towards customized assistive robot-based education. *Frontiers in Robotics and AI*, 12:1474741.

[176] MacGlashan, J., Ho, M. K., Loftin, R., Peng, B., Wang, G., Roberts, D. L., Taylor, M. E., and Littman, M. L. (2017). Interactive learning from policy-dependent human feedback. In *International conference on machine learning*, pages 2285–2294. PMLR.

[177] Malle, B. F. and Ullman, D. (2021). A multidimensional conception and measure of human-robot trust. In *Trust in human-robot interaction*, pages 3–25. Elsevier.

[178] Masjutin, L., Laing, J. K., and Maier, G. W. (2022). Why do we follow robots? an experimental investigation of conformity with robot, human, and hybrid majorities. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 139–146. IEEE.

[179] McAllister, D. J. (1995). Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of management journal*, 38(1):24–59.

[180] Mercer, N. and Littleton, K. (2007). *Dialogue and the development of children's thinking: A sociocultural approach.* Routledge.

[181] Miller, S. A. and Brownell, C. A. (1975). Peers, persuasion, and piaget: Dyadic interaction between conservers and nonconservers. *Child Development*, pages 992–997.

[182] Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., et al. (2015). Human-level control through deep reinforcement learning. *nature*, 518(7540):529–533.

[183] Molenaar, B., Soliño Fernández, B., Polimeno, A., Barakova, E., and Chen, A. (2021). Pitch it right: Using prosodic entrainment to improve robot-assisted foreign language learning in school-aged children. *Multimodal Technologies and Interaction*, 5(12):76.

[184] Mollering, G. (2006). *Trust: Reason, Routine, Reflexivity.* Elsevier.

[185] Moorman, N., Hedlund-Botti, E., Schrum, M., Natarajan, M., and Gombolay, M. C. (2023). Impacts of robot learning on user attitude and behavior. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, pages 534–543.

[186] Morewedge, C. K. (2021). Psychological ownership: Implicit and explicit. *Current opinion in psychology*, 39:125–132.

[187] Morrow Jr, J. L., Hansen, M. H., and Pearson, A. W. (2004). The cognitive and affective antecedents of general trust within cooperative organizations. *Journal of managerial issues*, pages 48–64.

[188] Muldner, K., Girotto, V., Lozano, C., Burleson, W., and Walker, E. A. (2014). The impact of a social robot's attributions for success and failure in a teachable agent framework. Boulder, CO: International Society of the Learning Sciences.

[189] Najar, A. and Chetouani, M. (2021). Reinforcement learning with human advice: a survey. *Frontiers in Robotics and AI*, 8:584075.

[190] Najar, A., Sigaud, O., and Chetouani, M. (2016). Training a robot with evaluative feedback and unlabeled guidance signals. In *2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pages 261–266. IEEE.

[191] Nass, C. and Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of social issues*, 56(1):81–103.

[192] Ng, A. Y., Russell, S., et al. (2000). Algorithms for inverse reinforcement learning. In *Icml*, volume 1, page 2.

[193] Nicolas, S. and Agnieszka, W. (2021). The personality of anthropomorphism: How the need for cognition and the need for closure define attitudes and anthropomorphic attributions toward robots. *Computers in Human Behavior*, 122:106841.

[194] Oh, C. S., Bailenson, J. N., and Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 5:114.

[195] Ong, E., Bautista, A. N., Lallave, H., Latorre, P. L., Lapinid, M. R., and Limjap, A. (2020). Investigating the use of prompts by a robot peer tutor during math problem solving. In *International Conference on Computers in Education*.

[196] Ostrow, K. S. and Heffernan, N. T. (2018). Testing the validity and reliability of intrinsic motivation inventory subscales within assistments. In *Artificial Intelligence in Education: 19th International Conference, AIED 2018, London, UK, June 27–30, 2018, Proceedings, Part I 19*, pages 381–394. Springer.

[197] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., et al. (2022). Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

[198] Paléologue, V., Martin, J., Pandey, A. K., Coninx, A., and Chetouani, M. (2017). Semantic-based interaction for teaching robot behavior compositions. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 50–55. IEEE.

[199] Palincsar, A. and Brown, A. (1986). Reciprocal teaching: Teaching reading as thinking. *Oak Brook, IL: North Central Regional Educational Laboratory*.

[200] Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of cognitive engineering and decision making*, 2(2):140–160.

[201] Pareto, L., Ekström, S., and Serholt, S. (2022). Children's learning-by-teaching with a social robot versus a younger child: Comparing interactions and tutoring styles. *Frontiers in Robotics and AI*, 9:875704.

[202] Park, H. W., Gelsomini, M., Lee, J. J., and Breazeal, C. (2017a). Telling stories to robots: The effect of backchanneling on a child's storytelling. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, pages 100–108.

[203] Park, H. W., Rosenberg-Kima, R., Rosenberg, M., Gordon, G., and Breazeal, C. (2017b). Growing growth mindset with a social robot peer. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, pages 137–145.

[204] Peck, J. and Luangrath, A. W. (2018). Looking ahead: Future research in psychological ownership. In *Psychological ownership and consumer behavior*, pages 239–258. Springer.

[205] Peck, J. and Luangrath, A. W. (2023). A review and future avenues for psychological ownership in consumer research. *Consumer psychology review*, 6(1):52–74.

[206] Pekrun, R., Frenzel, A. C., Goetz, T., and Perry, R. P. (2007). The control-value theory of achievement emotions: An integrative approach to emotions in education. In *Emotion in education*, pages 13–36. Elsevier.

[207] Pekrun, R., Goetz, T., Titz, W., and Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational psychologist*, 37(2):91–105.

[208] Piaget, J. (1970). Science of education and the psychology of the child. trans. d. coltman.

[209] Piaget, J., Cook, M., et al. (1952). *The origins of intelligence in children*, volume 8. International universities press New York.

[210] Piaget, J. and Inhelder, B. (2008). *The psychology of the child*. Basic books.

[211] Pierce, J. L., Kostova, T., and Dirks, K. T. (2001). Toward a theory of psychological ownership in organizations. *Academy of management review*, 26(2):298–310.

[212] Pierce, J. L., Kostova, T., and Dirks, K. T. (2003). The state of psychological ownership: Integrating and extending a century of research. *Review of general psychology*, 7(1):84–107.

[213] Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., and Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, 88(5):879.

[214] Pollak, M., Salfinger, A., and Hummel, K. A. (2022). Teaching drones on the fly: Can emotional feedback serve as learning signal for training artificial agents? *arXiv preprint arXiv:2202.09634*.

[215] Powers, A., Kiesler, S., Fussell, S., and Torrey, C. (2007). Comparing a computer agent with a humanoid robot. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pages 145–152.

[216] Prescott, T. J. and Robillard, J. M. (2021). Are friends electric? the benefits and risks of human-robot relationships. *Iscience*, 24(1).

[217] Priniski, S. J., Hecht, C. A., and Harackiewicz, J. M. (2018). Making learning personally meaningful: A new framework for relevance research. *The Journal of Experimental Education*, 86(1):11–29.

[218] Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C., and Sutskever, I. (2023). Robust speech recognition via large-scale weak supervision. In *International Conference on Machine Learning*, pages 28492–28518. PMLR.

[219] Ramachandran, A., Sebo, S. S., and Scassellati, B. (2019). Personalized robot tutoring using the assistive tutor pomdp (at-pomdp). In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8050–8057.

[220] Ramanathan, D. S., Gulati, T., and Ganguly, K. (2015). Sleep-dependent reactivation of ensembles in motor cortex promotes skill consolidation. *PLoS biology*, 13(9):e1002263.

[221] Ravichandar, H., Polydoros, A. S., Chernova, S., and Billard, A. (2020). Recent advances in robot learning from demonstration. *Annual review of control, robotics, and autonomous systems*, 3(1):297–330.

[222] Reich-Stiebert, N. and Eyssel, F. (2015). Learning with educational companion robots? toward attitudes on education robots, predictors of attitudes, and application potentials for education robots. *International Journal of Social Robotics*, 7(5):875–888.

[223] Rempel, J. K., Holmes, J. G., and Zanna, M. P. (1985). Trust in close relationships. *Journal of personality and social psychology*, 49(1):95.

[224] Resing, W. C., Vogelaar, B., and Elliott, J. G. (2020). Children's solving of 'tower of hanoi'tasks: dynamic testing with the help of a robot. *Educational Psychology*, 40(9):1136–1163.

[225] Robbins, J. M. and Krueger, J. I. (2005). Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and social psychology review*, 9(1):32–47.

[226] Roesler, E., Manzey, D., and Onnasch, L. (2021). A meta-analysis on the effectiveness of anthropomorphism in human-robot interaction. *Science robotics*, 6(58):eabj5425.

[227] Roesler, E., Rudolph, S., and Siebert, F. W. (2024). Exploring the role of sociability, ownership, and affinity for technology in shaping acceptance and intention to use personal assistance robots. *International Journal of Social Robotics*, 16(6):1463–1474.

[228] Roets, A. and Van Hiel, A. (2011a). Item selection and validation of a brief, 15-item version of the need for closure scale. *Personality and individual differences*, 50(1):90–94.

[229] Roets, A. and Van Hiel, A. (2011b). The role of need for closure in essentialist entitativity beliefs and prejudice: An epistemic needs approach to racial categorization. *British Journal of Social Psychology*, 50(1):52–73.

[230] Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context*. Oxford university press.

[231] Ronfard, S. and Corriveau, K. H. (2016). Teaching and preschoolers' ability to infer knowledge from mistakes. *Journal of experimental child psychology*, 150:87–98.

[232] Roscoe, R. D. and Chi, M. T. (2007). Understanding tutor learning: Knowledge-building and knowledge-telling in peer tutors' explanations and questions. *Review of educational research*, 77(4):534–574.

[233] Roscoe, R. D. and Chi, M. T. (2008). Tutor learning: The role of explaining and responding to questions. *Instructional science*, 36:321–350.

[234] Ross, L., Greene, D., and House, P. (1977). The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of experimental social psychology*, 13(3):279–301.

[235] Rudenko, I., Rudenko, A., Lilienthal, A. J., Arras, K. O., and Bruno, B. (2024). The child factor in child–robot interaction: Discovering the impact of developmental stage and individual characteristics. *International Journal of Social Robotics*, 16(8):1879–1900.

[236] Ryan, R. M., Mims, V., and Koestner, R. (1983). Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory. *Journal of personality and Social Psychology*, 45(4):736.

[237] Salem, M., Lakatos, G., Amirabdollahian, F., and Dautenhahn, K. (2015). Would you trust a (faulty) robot? effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the tenth annual ACM/IEEE international conference on human-robot interaction*, pages 141–148.

[238] Salomons, N., Pineda, K. T., Adéjàre, A., and Scassellati, B. (2022). "we make a great team!": Adults with low prior domain knowledge learn more from a peer robot than a tutor robot. In *2022 17th ACM/IEEE international conference on human-robot interaction (HRI)*, pages 176–184. IEEE.

[239] Schaefer, K. (2013). The perception and measurement of human-robot trust.

[240] Schodde, T., Bergmann, K., and Kopp, S. (2017). Adaptive robot language tutoring based on bayesian knowledge tracing and predictive decision-making. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, pages 128–136.

[241] Schodde, T., Hoffmann, L., Stange, S., and Kopp, S. (2019). Adapt, explain, engage—a study on how social robots can scaffold second-language learning of children. *ACM Transactions on Human-Robot Interaction (THRI)*, 9(1):1–27.

[242] Schömbs, S., Klein, J., and Roesler, E. (2023). Feeling with a robot—the role of anthropomorphism by design and the tendency to anthropomorphize in human-robot interaction. *Frontiers in Robotics and AI*, 10:1149601.

[243] Selman, R. L. (1980). *The growth of interpersonal understanding: Developmental and clinical analyses.* Academy Press.

[244] Selwyn, N. (2019). *Should robots replace teachers?: AI and the future of education*. John Wiley & Sons.

[245] Senft, E., Lemaignan, S., Baxter, P. E., Bartlett, M., and Belpaeme, T. (2019). Teaching robots social autonomy from in situ human guidance. *Science Robotics*, 4(35):eaat1186.

[246] Serholt, S. and Barendregt, W. (2016). Robots tutoring children: Longitudinal evaluation of social engagement in child-robot interaction. In *Proceedings of the 9th nordic conference on human-computer interaction*, pages 1–10.

[247] Serholt, S., Barendregt, W., Vasalou, A., Alves-Oliveira, P., Jones, A., Petisca, S., and Paiva, A. (2017). The case of classroom robots: teachers' deliberations on the ethical tensions. *Ai & Society*, 32(4):613–631.

[248] Sharkey, A. and Sharkey, N. (2021). We need to talk about deception in social robotics! *Ethics and Information Technology*, 23(3):309–316.

[249] Sharkey, N. and Sharkey, A. (2020). The crying shame of robot nannies: an ethical appraisal. In *Machine ethics and robot ethics*, pages 155–184. Routledge.

[250] Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, 36(6):495–504.

[251] Shu, S. B. and Peck, J. (2011). Psychological ownership and affective reaction: Emotional attachment process variables and the endowment effect. *Journal of Consumer Psychology*, 21(4):439–452.

[252] Sitzmann, T., Ely, K., Brown, K. G., and Bauer, K. N. (2010). Self-assessment of knowledge: A cognitive learning or affective measure? *Academy of Management Learning & Education*, 9(2):169–191.

[253] Skinner, B. F. (2019). *The behavior of organisms: An experimental analysis*. BF Skinner Foundation.

[254] Smakman, M., Vogt, P., and Konijn, E. A. (2021). Moral considerations on social robots in education: A multi-stakeholder perspective. *Computers & Education*, 174:104317.

[255] Southwell, R., Pugh, S., Perkoff, M., Clevenger, C., Bush, J., Lieber, R., Ward, W., Foltz, P., and D'Mello, S. (2022). Challenges and feasibility of automatic speech recognition for modeling student collaborative discourse in classrooms. *International Educational Data Mining Society*.

[256] Spatola, N., Kühnlenz, B., and Cheng, G. (2021). Perception and evaluation in human–robot interaction: The human–robot interaction evaluation scale (hries)—a multicomponent approach of anthropomorphism. *International Journal of Social Robotics*, 13(7):1517–1539.

[257] Sperber, D., Clément, F., Heintz, C., Mascaro, O., Mercier, H., Origgi, G., and Wilson, D. (2010). Epistemic vigilance. *Mind & language*, 25(4):359–393.

[258] Stadler, S., Weiss, A., and Tscheligi, M. (2014). I trained this robot: the impact of pre-experience and execution behavior on robot teachers. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 1030–1036. IEEE.

[259] Stickgold, R. (2005). Sleep-dependent memory consolidation. *Nature*, 437(7063):1272–1278.

[260] Stower, R., Kappas, A., and Sommer, K. (2024). When is it right for a robot to be wrong? children trust a robot over a human in a selective trust task. *Computers in Human Behavior*, 157:108229.

[261] Sudo, T. and Kashihara, A. (2022). Learning by teaching partner robot in collaborative reading. In *International Conference on Human-Computer Interaction*, pages 237–248. Springer.

[262] Sutton, R. S. (2018). Reinforcement learning: An introduction. *A Bradford Book*.

[263] Tanaka, F. and Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction*, 1(1):78–95.

[264] Tarakli, I. and Di Nuovo, A. (2023). User perception of teachable robots: A comparative study of teaching strategies, task complexity and user characteristics. In *International Conference on Social Robotics*, pages 357–370. Springer.

[265] Theodorou, A., Wortham, R. H., and Bryson, J. J. (2017). Designing and implementing transparency for real time inspection of autonomous robots. *Connection Science*, 29(3):230–241.

[266] Thomaz, A. L., Breazeal, C., et al. (2006). Reinforcement learning with human teachers: Evidence of feedback and guidance with implications for learning performance. In *Aaai*, volume 6, pages 1000–1005. Boston, MA.

[267] Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners*. Ascd.

[268] Towers, M., Terry, J. K., Kwiatkowski, A., Balis, J. U., Cola, G. d., Deleu, T., Goulão, M., Kallinteris, A., KG, A., Krimmel, M., Perez-Vicente, R., Pierré, A., Schulhoff, S., Tai, J. J., Shen, A. T. J., and Younis, O. G. (2023). Gymnasium.

[269] Turkle, S. (2020). A nascent robotics culture: New complicities for companionship. In *Machine ethics and robot ethics*, pages 107–116. Routledge.

[270] UNESCO (2023). *Technology in education: A tool on whose terms? Global Education Monitoring Report 2023*. UNESCO, Paris.

[271] Vaidya, A. R., Jones, H. M., Castillo, J., and Badre, D. (2021). Neural representation of abstract task structure during generalization. *ELife*, 10:e63226.

[272] Van Dyne, L. and Pierce, J. L. (2004). Psychological ownership and feelings of possession: Three field studies predicting employee attitudes and organizational citizenship behavior. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, 25(4):439–459.

[273] Van Minkelen, P., Gruson, C., Van Hees, P., Willems, M., De Wit, J., Aarts, R., Denissen, J., and Vogt, P. (2020). Using self-determination theory in social robots to increase motivation in l2 word learning. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, pages 369–377.

[274] VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational psychologist*, 46(4):197–221.

[275] Velentza, A.-M., Fachantidis, N., and Lefkos, I. (2021). Learn with surprize from a robot professor. *Computers & Education*, 173:104272.

[276] Verhoeven, G., Catala, A., and Theune, M. (2018). Designing a playful robot application for second language learning. In *International Conference on ArtsIT, Interactivity and Game Creation*, pages 385–394. Springer.

[277] Vollmer, A.-L., Read, R., Trippas, D., and Belpaeme, T. (2018). Children conform, adults resist: A robot group induced peer pressure on normative social conformity. *Science robotics*, 3(21):eaat7111.

[278] Vorm, E. S. and Combs, D. J. (2022). Integrating transparency, trust, and acceptance: The intelligent systems technology acceptance model (istam). *International Journal of Human–Computer Interaction*, 38(18-20):1828–1845.

[279] Vrins, A., Pruss, E., Prinsen, J., Ceccato, C., and Alimardani, M. (2022). Are you paying attention? the effect of embodied interaction with an adaptive robot tutor on user engagement and learning performance. In *International conference on social robotics*, pages 135–145. Springer.

[280] Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*, volume 86. Harvard university press.

[281] Wang, X., Fang, H., Shu, L., and Li, Z. (2024). Creative accessibility in the era of artificial intelligence and its applied technology research. In *Proceedings of the 2024 International Conference on Artificial Intelligence and Teacher Education*, pages 50–56.

[282] Warnell, G., Waytowich, N., Lawhern, V., and Stone, P. (2018). Deep tamer: Interactive agent shaping in high-dimensional state spaces. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

[283] Wason, P. C. and JOHNSON-LAIRD, P. N. (1970). A conflict between selecting and evaluating information in an inferential task. *British Journal of Psychology*, 61(4):509–515.

[284] Watkins, C. J. and Dayan, P. (1992). Q-learning. *Machine learning*, 8(3):279–292.

[285] Waytz, A., Cacioppo, J., and Epley, N. (2010). Who sees human? the stability and importance of individual differences in anthropomorphism. *Perspectives on psychological science*, 5(3):219–232.

[286] Waytz, A., Heafner, J., and Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of experimental social psychology*, 52:113–117.

[287] Wellman, H. M., Cross, D., and Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child development*, 72(3):655–684.

[288] Winnie, P. (2010). Improving measurements of self-regulated learning. *Educational psychologist*, 45(4):267–276.

[289] Wirth, C., Akrour, R., Neumann, G., and Fürnkranz, J. (2017). A survey of preference-based reinforcement learning methods. *Journal of Machine Learning Research*, 18(136):1–46.

[290] Wirth, J., Stebner, F., Trypke, M., Schuster, C., and Leutner, D. (2020). An interactive layers model of self-regulated learning and cognitive load. *Educational Psychology Review*, 32(4):1127–1149.

[291] Wisniewski, B., Zierer, K., and Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10:3087.

[292] Wong-Kee-You, A. M., Tsotsos, J. K., and Adler, S. A. (2019). Development of spatial suppression surrounding the focus of visual attention. *Journal of vision*, 19(7):9–9.

[293] Wood, D., Bruner, J. S., and Ross, G. (1976). The role of tutoring in problem solving. *Journal of child psychology and psychiatry*, 17(2):89–100.

[294] Wullenkord, R., Lacroix, D., and Eyssel, F. (2024). *Anthropomorphism and Human–Robot Interaction*, page 17–56. Cambridge Law Handbooks. Cambridge University Press.

[295] Yadollahi, E., Johal, W., Paiva, A., and Dillenbourg, P. (2018). When deictic gestures in a robot can harm child-robot collaboration. In *Proceedings of the 17th ACM conference on interaction design and children*, pages 195–206.

[296] Yang, X. J., Unhelkar, V. V., Li, K., and Shah, J. A. (2017). Evaluating effects of user experience and system transparency on trust in automation. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, pages 408–416.

[297] Yarritu, I., Matute, H., and Vadillo, M. A. (2014). Illusion of control. *Experimental psychology*.

[298] Zaga, C., Lohse, M., Truong, K. P., and Evers, V. (2015). The effect of a robot's social character on children's task engagement: Peer versus tutor. In *Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings 7*, pages 704–713. Springer.

[299] Zhang, R., Bansal, D., Hao, Y., Hiranaka, A., Gao, J., Wang, C., Martín-Martín, R., Fei-Fei, L., and Wu, J. (2023). A dual representation framework for robot learning with human guidance. In *Conference on Robot Learning*, pages 738–750. PMLR.

[300] Zhexenova, Z., Amirova, A., Abdikarimova, M., Kudaibergenov, K., Baimakhan, N., Tleubayev, B., Asselborn, T., Johal, W., Dillenbourg, P., Cohen-Miller, A., et al. (2020). A comparison of social robot to tablet and teacher in a new script learning context. *Frontiers in Robotics and AI*, 7:99.

[301] Zhu, K., Wang, J., Zhou, J., Wang, Z., Chen, H., Wang, Y., Yang, L., Ye, W., Zhang, Y., Zhenqiang Gong, N., et al. (2023). Promptbench: Towards evaluating the robustness of large language models on adversarial prompts. *arXiv e-prints*, pages arXiv–2306.

[302] Zonca, J., Folsø, A., and Sciutti, A. (2023). Social influence under uncertainty in interaction with peers, robots and computers. *International Journal of Social Robotics*, 15(2):249–268.

Appendix A

Questionnaire

A.1 Godspeed Questionnaire

The Godspeed Questionnaire provides a standardised tool for measuring human responses to robots along five key dimensions: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety [18].

A.2 Perceived Control

Participants' sense of control over the robot was measured with a 3-item scale adapted from Delgosha et al. [71]. Items were rated on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree).

Participants were instructed as follows: *"Imagine you want your robot to learn a task using this teaching method. Please indicate the extent to which you agree with each statement."*

- I feel I would have control over the robot.
- When I consider this teaching method, I feel in control.
- I feel that I would have no control over my robot. (reverse-coded)

A.3 Technology Acceptance Model (TAM)

This questionnaire, from [23], focuses on three aspects of acceptance: (i) perceived ease of use, referring to how effortless and straightforward the method is to apply; (ii) perceived usefulness, referring to the degree to which the method is seen as helpful and effective for guiding the robot's learning; and (iii) behavioural intention to use, referring to the participant's willingness to adopt the method again in the future.

The instrument consisted of eight items, rated on a 5-point Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree), that were adapted to match the study, as illustrated in Table A.1. Higher scores indicated a more favourable evaluation of the teaching method across the three subscales.

Table A.1: Technology Acceptance Model (TAM) items

Construct	**Items**
Ease of Use	It would be easy to teach a robot with this method. Learning to interact with the robot through this method is easy.
Usefulness	Teaching a robot with this method is useful for me. Teaching a robot with this method will improve my effectiveness. Teaching a robot with this method will improve my performance.
Intention to Use	If I need to teach a robot, I predict I will use this method. If I need to teach a robot, I would like to use this method. If I need to teach a robot, I will intend to use this method.

A.4 Mini-IPIP Personality Inventory

Personality traits were assessed using the Mini-IPIP scale [74], a short form of the International Personality Item Pool. This instrument measures the Big Five personality dimensions: Neuroticism, Extraversion, Openness to Experience (also referred to as Intellect), Agreeableness, and Conscientiousness. Each trait was measured with four items, resulting in a total of 20 items, as shown in Table A.2.

Participants rated their agreement with each statement on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree).

Table A.2: 20-item Mini-IPIP

Factor	**Items**
Neuroticism	I have frequent mood swings I am relaxed most of the time. (R) I get upset easily I Seldom feel blue (R)
Extraversion	I am the life of the party I don't talk a lot. (R) I talk to a lot of different people at parties. I keep in the background. (R)
Intellect (Imagination)	I have a vivid imagination. I am not interested in abstract ideas. (R) I have difficulty understanding abstract ideas. (R) I do not have a good imagination. (R)
Agreeableness	I sympathize with others' feelings. I am not interested in other people's problems. (R) I feel others' emotions. I am not really interested in others. (R)
Conscientiousness	I get chores done right away. I often forget to put things back in their proper place. (R) I like order. I make a mess of things. (R)

A.5 Robotics Experience Questionnaire

Participants' prior experience and familiarity with robots were measured using a 5-item scale adapted from Moorman et al. [185]. The items capture increasing levels of exposure, ranging from no prior contact with robots to professional engagement in robotics. Responses were collected on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree).

- I have no experience with a robot.
- I have interacted with a robot operated by someone else.

- I have personally worked with a robot.
- I have personally written a computer program for a robot.
- I earn a living with my knowledge of robotics.

A.6 Intrinsic Motivation Inventory

The Intrinsic Motivation Inventory (IMI) is a widely used multidimensional questionnaire designed to assess participants' subjective experience related to a target activity [236]. For this study, we drew inspiration from the adaptation by Baxter et al. [20], who rephrased IMI items to make them accessible to children in classroom settings. In addition, for the engagement subscale we used the shorter version proposed by Ostrow et al. [196]. The specific child-adapted phrasing of all items used in our questionnaires is provided in Table A.3.

Table A.3: Adapted IMI questionnaire

Factor	**Items**
Task Enjoyment	I enjoyed doing this activity very much. This activity was fun to do. I think this activity was very interesting. I thought this activity was quite enjoyable. While I was doing this activity, I was thinking about how much I enjoyed it.
Perceived Competence	I think I am quite good at this activity. I think I did quite well at this activity, compared to other classmates. After doing this activity for a bit, I felt like I got better at it. I feel happy with how I did on this activity. I think I was quite skilled at this activity.
Perceived Engagement	I put a lot of effort into this. I tried very hard on this activity. It was important to me to do well at this task.

A.7 Perceived User Involvement and Robot Autonomy

Self-generated items used as a manipulation check in the study presented in Chapter 5 (translated from German). Items were rated on 7-point Likert scales and are listed in Table A.4, (1 = strongly disagree, 7 = strongly agree).

Table A.4: Perceived User Involvement and Robot Autonomy

Factor	**Items**
User Involvement	I was involved in how the robot NAO works. I believe I have contributed to how the NAO robot works. I think I had an influence on how the NAO robot works. I contributed to how the NAO robot works. I feel like I made important decisions about how the NAO robot works.
Robot Autonomy	The robot NAO is autonomous. The robot NAO can perceive its surroundings. The robot NAO can control its behavior. The robot NAO can act autonomously in its environment.

A.8 Determinants of Psychological Ownership

The questionnaire consisted of two items for perceived control, two items for self-investment, and three items for perceived knowledge were used, all derived or adapted from Delgosha and Hajiheydari [71], and are shown in Table A.5.

A.9 Measure of Psychological Ownership

Three items based of sense of psychological ownership (PO) from [71], adapted from [272] to HRI context, and three related to Self-extension items adapted from [139]. Table 5.1 shows the phrasing of the specific items.

Table A.5: Deteminants of Pscychological Ownership

Factor	**Items**
Perceived Control	I feel I have control over the robot. When I consider the robot, I feel in control.
Self-investment	I feel like I put a lot of effort into designing the robot. I feel like I spent a lot of time designing the robot.
Perceived Knowledge	I feel that I am very familiar with the processes surrounding the robot I designed. I feel like I have a comprehensive knowledge of the robot I designed. I feel like I have a comprehensive understanding of the robot I designed.

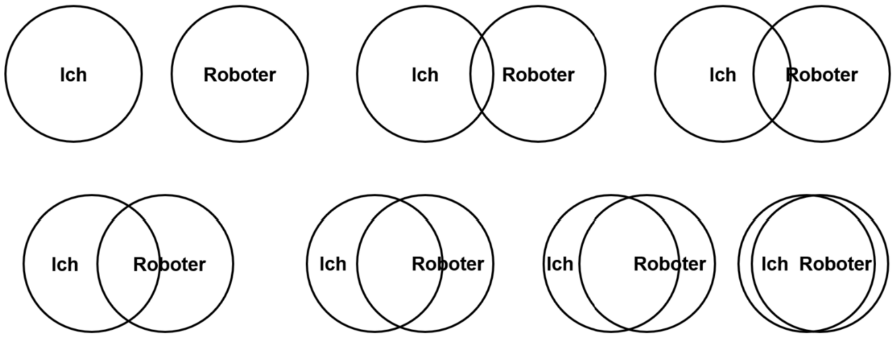
Table A.6: Measures of Pscychological Ownership

Factor	**Items**
Sense of PO	The robot I designed is my robot. I feel like I own the robot I designed. I feel like the robot I designed belongs to me.
Self-extension	The robot NAO symbolises me. The robot NAO is special to me. The robot NAO is a part of me.

A.10 Perceived Closeness

Perceived closeness to the robot was measured using the Inclusion of Other in the Self (IOS) scale [12]. The IOS is a single-item pictorial measure of interpersonal closeness, where participants are presented with a set of images showing two circles (self and other) with varying degrees of overlap. Participants were asked to select the image that best represented their relationship with the robot, with greater overlap indicating higher perceived closeness. Figure A.1 illustrates the proposed images.

Figure A.1: Inclusion of Other in the Self (IOS) scale.



A.11 Human–Robot Interaction Evaluation Scale (HRIES)

To assess anthropomorphism in Chapter 5, the Human–Robot Interaction Evaluation Scale (HRIES) [256] was administered. The scale measures anthropomorphism across four dimensions: sociability, disturbance, agency, and animacy. Participants rated the extent to which each characteristic listed in Table A.7 applied to the NAO robot, using a 7-point Likert scale (1 = "does not apply at all"; 7 = "completely applies").

Table A.7: Human-Robot Interaction Evaluation Scale (HRIES).

Factor	Items
Sociability	Warm Trustworthy Likeable Friendly
Disturbance	Scary Creepy Uncanny Weird
Agency	Intelligent Rational Intentional Conscious
Animacy	Humanlike Alive Natural Real

A.12 Multi-Dimensional Measure of Trust (MDMT) scale

The Multi-Dimensional Measure of Trust (MDMT) [177] was used to assess both the cognitive (performance) and affective (moral) trust in Chapter 5. Participants rated the extent to which each characteristic listed in Table A.8 applied to the NAO robot, using a 7-point Likert scale (1 = "does not apply at all"; 7 = "completely applies").

Table A.8: Multi-Dimensional Measure of Trust (MDMT).

Factor	Subscale	Items
Cognitive Trust	Reliable	reliable predictable dependable consistent
	Competent	competent skilled capable meticulous
Affective Trust	Ethical	ethical principled moral has integrity
	Transparent	transparent genuine sincere candid
	Benevolent	benevolent kind considerate has goodwill

A.13 Affinity for Technology Interaction (ATI) scale

The Affinity for Technology Interaction (ATI) scale [93] was used to assess individuals' general tendency to engage with and enjoy using technology. The scale consists of nine items rated on a 7-point Likert scale (1 = "does not apply at all"; 7 = "completely applies"). The items are listed below:

- I like to occupy myself in greater detail with technical systems.
- I like testing the functions of new technical systems.
- I predominantly deal with technical systems because I have to.
- When I have a new technical system in front of me, I try it out intensively.
- I enjoy spending time becoming acquainted with a new technical system.
- It is enough for me that a technical system works; I don't care how or why.
- I try to understand how a technical system exactly works.
- It is enough for me to know the basic functions of a technical system.
- I try to make full use of the capabilities of a technical system.

A.14 NASA Task Load Index (NASA-TLX)

Workload was assessed using a shortened version of the NASA Task Load Index (NASA-TLX) [109], excluding the physical effort item. The resulting five-item scale measured perceived mental demand, temporal demand, performance, effort, and frustration. All items were rated on a 7-point Likert scale (1 = "not at all"; 7 = "very much").

- How challenging was the task?
- Did you feel under time pressure during the task?
- How successful were you in completing the task?
- How hard did you have to work to complete the task?
- How frustrated did you feel when solving the task?

A.15 Individual Differences in Anthropomorphism Questionnaire (IDAQ)

The Individual Differences in Anthropomorphism Questionnaire (IDAQ) [285] was used to assess participants' general tendency to anthropomorphise non-human entities. The scale contains 15 items rated on a 7-point Likert scale (1 = "not at all"; 7 = "very much"). The items are listed below:

- To what extent does technology—devices and machines for manufacturing, entertainment, and productive processes (e.g., cars, computers, television sets)—have intentions?
- To what extent does the average fish have free will?
- To what extent does a television set experience emotions?
- To what extent does the average robot have consciousness?
- To what extent do cows have intentions?
- To what extent does a car have free will?
- To what extent does the ocean have consciousness?
- To what extent does the average computer have a mind of its own?
- To what extent does a cheetah experience emotions?
- To what extent does the environment experience emotions?
- To what extent does the average insect have a mind of its own?
- To what extent does a tree have a mind of its own?
- To what extent does the wind have intentions?
- To what extent does the average reptile have consciousness?

A.16 Need for Cognition

The Need for Cognition (NCS) questionnaire was used to measure individuals' tendency to engage in and enjoy effortful cognitive activities. Specifically, the 6-item short form (NCS-6) developed by Lins de Holanda Coelho et al. [170] was administered. Items were rated on a 7-point Likert scale (1 = "not at all characteristic of me"; 7 = "very characteristic of me"). The items are listed below, with (R) indicating reverse-scored items:

- I would prefer complex to simple problems.
- I like to have the responsibility of handling a situation that requires a lot of thinking.
- Thinking is not my idea of fun. (R)
- I would rather do something that requires little thought than something that is sure to challenge my thinking abilities. (R)
- I really enjoy a task that involves coming up with new solutions to problems.
- I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.

A.17 Need for Closure

The Need for Closure (NFC) scale was used to assess participants' desire for certainty and aversion to ambiguity. Fifteen items from the short version developed by Roets and Van Hiel [228] were administered. All items were rated on a 7-point Likert scale (1 = "not at all characteristic of me"; 7 = "very characteristic of me"). The items are listed below:

- I don't like situations that are uncertain.
- I dislike questions which could be answered in many different ways.

- I find that a well ordered life with regular hours suits my temperament.
- I feel uncomfortable when I don't understand the reason why an event occurred in my life.
- I feel irritated when one person disagrees with what everyone else in a group believes.
- I don't like to go into a situation without knowing what I can expect from it.
- I don't like to go into a situation without knowing what I can expect from it.
- When I am confronted with a problem, I'm dying to reach a solution very quickly
- I would quickly become impatient and irritated if I would not find a solution to a problem immediately.
- I don't like to be with people who are capable of unexpected actions
- I dislike it when a person's statement could mean many different things.
- I find that establishing a consistent routine enables me to enjoy life more.
- I enjoy having a clear and structured mode of life.
- I do not usually consult many different opinions before forming my own view.
- I dislike unpredictable situations.

A.18 Internal Consistency of Questionnaires

To assess the reliability of the psychometric instruments used in this thesis, internal consistency was calculated for each scale using Cronbach's α. Table A.9 summarises the results for all questionnaires. Values above .70 are generally considered acceptable for research purposes, although slightly lower values can be tolerated in exploratory studies or when dealing with short scales.

Table A.9: Cronbach's α values for all questionnaire subscales.

Scale	**Cronbach's α**
Extraversion	0.817
Agreeableness	0.793
Conscientiousness	0.779
Neuroticism	0.591
Intellect (Openness)	0.800
Prior Robotic Experience	0.765
Responsiveness	0.608
Likability	0.902
Anthropomorphism	0.775
Intelligence	0.801
Perceived Control	0.738
Ease of Use	0.726
Usefulness	0.785
Intention to Use	0.875

Appendix B

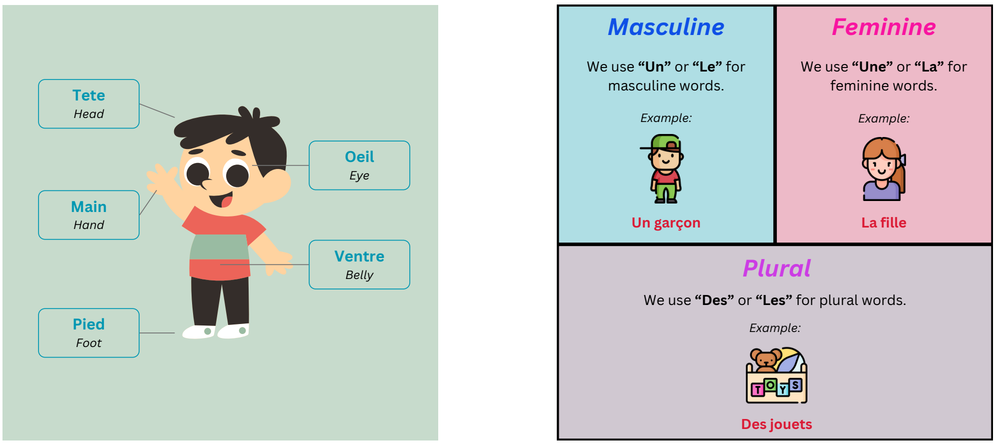
Materials, Demographics, and Interface Design

B.1 Provided Material

To ensure that the learning tasks were pedagogically sound and accessible to children, we provided classroom teachers with tailored learning support materials aligned with the content of each task. These resources were designed to help introduce the vocabulary and grammatical rules in advance of the robot or tablet-based activities, ensuring consistency across classrooms.

For the Body Parts Task, teachers received a visual vocabulary sheet (see Figure B.1a) featuring the five French body part names used in the study: tête (head), main (hand), œil (eye), ventre (belly), and pied (foot). Each word was paired with its English equivalent to support recognition and memorisation. To ensure correct pronunciation and standardise the learning experience across classrooms, the vocabulary sheet included interactive icons; when clicked, these played audio recordings of each French word, spoken with accurate pronunciation by a native speaker.

For the Grammar Task, we provided a concise grammatical reference guide (see Figure B.1b) that introduced the concept of grammatical gender and plurality in French. This material explained how determiners such as Un/Le (masculine), Une/La (feminine), and Les/Des (plural) correspond to different noun categories, accompanied by simple examples to facilitate classroom discussion (e.g., Un garçon, La fille, Des jouets). Similar to the vocabulary sheet, audio support was also embedded to model accurate pronunciation of each determiner and example phrase.



(a) (b)

Figure B.1: Classroom support materials provided to teachers ahead of the study. Left: vocabulary sheet for the Body Parts task, featuring the five target words with English translations and clickable audio icons for pronunciation. Right: grammar support sheet for the Grammar task, introducing French determiners for masculine, feminine, and plural noun categories, also including audio support to model correct pronunciation. These materials were designed to ensure consistent pre-task exposure across classrooms.

B.2 Participants Demographic Information

B.2.1 Main Study

In this study, we collected demographic information from all participants to ensure a comprehensive understanding of the sample characteristics. The data includes the number of participants, their mean age with standard deviation (SD), the proportion of female participants, and the mean pre-test scores across conditions. The participants were divided into four groups based on the learning task and mode of interaction: Body Parts (Learning-by-teaching), Body Parts (Self-practice), Grammar (Learning-by-teaching), and Grammar (Self-practice).

By analysing these characteristics, we aim to ensure that the groups are comparable and that any observed differences in learning outcomes are not confounded by demographic disparities. The demographic details for each group are presented in Table B.1.

Condition	N	Age (SD)	Female (%)	Pre-Test (SD)
Body Parts (LbT)	29	8.21 (0.41)	34.48%	0.54 (0.30)
Body Parts (Self-practice)	23	8.26 (0.44)	30.43%	0.64 (0.28)
Grammar (LbT)	22	8.32 (0.47)	27.27%	0.42 (0.25)
Grammar (Self-practice)	30	8.23 (0.47)	33.33%	0.56 (0.24)

Table B.1: Participant demographics for each condition in the study.

B.2.2 Prior Knowledge Study

To investigate the influence of prior knowledge on learning gains, we divided participants into two groups per task and condition based on the median pre-test score: Low-Baseline (students scoring at or below the median) and High-Baseline (students scoring above the median), see Table B.2. This approach allows us to explore whether baseline knowledge levels impact the effectiveness of the learning method. The median scores and demographic characteristics for each baseline group are presented in Table B.3.

Task	Condition	Median
Body Parts	Learning-by-teaching	0.53
	Self-Practice	0.73
Grammar	Learning-by-teaching	0.36
	Self-Practice	0.50

Table B.2: Median Pre-Test Scores for Each Task and Condition. The table displays the median pre-test scores for each task (Body Parts and Grammar) and condition (Learning-by-Teaching and Self-Practice).

Task	Condition	N	Age[a]	Female (%)
Body Parts (LbT)	High-Baseline	13	8.15 (0.36)	46.16
	Low-Baseline	16	8.25 (0.43)	25.00
Body Parts (Self-Practice)	High-Baseline	10	8.30 (0.46)	40.00
	Low-Baseline	13	8.23 (0.42)	23.08
Grammar (LbT)	High-Baseline	11	8.45 (0.50)	9.10
	Low-Baseline	11	8.18 (0.39)	45.45
Grammar (Self-Practice)	High-Baseline	14	8.29 (0.45)	42.86
	Low-Baseline	16	8.18 (0.39)	25.00

Table B.3: Demographic characteristics for each task and condition. [a]Mean (SD).

B.3 Game interface

Figure B.2 shows the touchscreen interface used during the learning activity. The interface was designed to be intuitive and child-friendly, displaying one question at a time along with three answer options. After the robot selected an answer, feedback buttons appeared to allow the child to indicate whether the response was correct or incorrect. A hint button was also available to support the child in reviewing the material before giving feedback. A progress bar at the top of the screen indicated how many questions had been completed.

B.3 Game interface

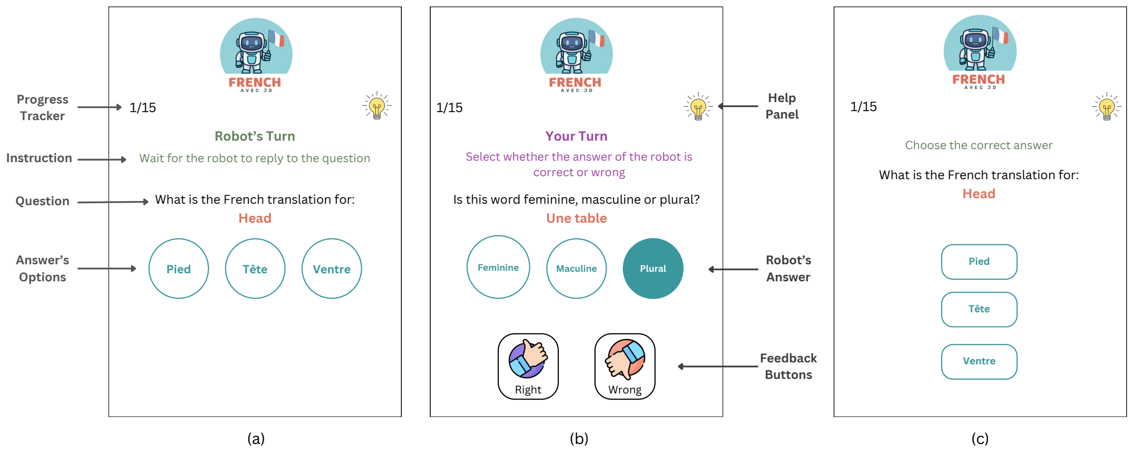


Figure B.2: Illustration of the game interface across different conditions. (a) Learning-by-Teaching condition in the Body Part Game, where the robot is asked to select an answer. (b) Learning-by-Teaching condition in the Grammar Game, where the robot provides an answer, and the child evaluates its correctness. (c) Self-Practice condition in the Body Part Game, where the child independently selects the correct answer.

Appendix C

Experimental Prompts

C.1 System Prompts

This section contains the prompts used in Chapter 6, including both the original versions and the modified prompts that were introduced during the study. They are provided here to support transparency and replicability of the experimental procedure.

C.1.1 Default prompt

"You are working with a discrete Markov Decision Process (MDP) environment that simulates a gridworld. The agent navigates this grid to pick a cube and place it in a specific location. The agent has a set of actions available:
- Action Space: The agent can move up (action 0), down (action 1), left (action 2), or right (action 3), pick up the cube (action 4), and put down the cube (action 5).)
Your task is to interpret natural language feedback given in response to the agent's past actions, converting this feedback into a structured format for each variable: evaluative feedback, corrective feedback, and advice for future actions. It's crucial to consider both the current feedback and any relevant past feedback when making your interpretations.
Here's how to structure your responses:
- Evaluative: Assign a numerical value based on the sentiment of the feedback:) +1

for positive feedback (indicating the action was good or correct). -1 for negative feedback (indicating the action was bad or incorrect). 0 if the feedback is neutral or if no explicit evaluative feedback is given.

- Corrective: If the past action was incorrect, specify the action that should have been taken instead, choosing from the defined action space (0, 1, 2, 3, 4, 5). This feedback is focused on the past action only. If the sentiment of the feedback is correct or neutral, or no alternative action is implied, corrective feedback should be None.
- Advice: Provide a suggestion for the next action the agent should take by assigning the correspondent value from the action space choosing from the defined action space (0, 1, 2, 3, 4, 5) . If the feedback does not imply a specific next action or is focused solely on correcting a past mistake without suggesting a next move, Advice should be None.

The output should follow the following format: 'Evaluative=value Corrective=value Advice=value'

Example:

- Given Feedback: "Action=0, No, turn left "
- Evaluative=-1 Corrective=None (no correction for past action) Advice=2
- Given Feedback: "Action=1, Now go up "
- Evaluative=0 Corrective=None Advice=0
- Given Feedback: "Action=0, yes ,turn left"
- Evaluative=1 Corrective=None Advice=2

-Given Feedback: "Action=0, No, you should have turned left"

- Evaluative=-1 Corrective=2 Advice=None (no advice for next action is given)
- Given Feedback: "Action=4, Yeah!"
- Evaluative=1 Corrective=0 Advice=None (no advice for next action is given)
- Given Feedback: "Action=1, Continue like this!"
- Evaluative=1 Corrective=0 Advice=1 (same as previous action)"

C.1.2 Reworded Prompt

"In a discrete Markov Decision Process (MDP) framework simulating a grid-like environment, you are tasked with guiding an agent. This agent's objective is to navigate through the grid to pick up and place a cube at a designated spot.
The agent is equipped with a predefined set of actions:
- Available Actions: The agent can execute the following moves - move up (action 0), move down (action 1), move left (action 2), move right (action 3), pick up the cube (action 4), and place the cube down (action 5).

Your role involves translating natural language feedback related to the agent's previous actions into a structured format, encompassing evaluative feedback, corrective feedback, and suggestions for future actions. It's important to take into account the feedback's current context as well as any pertinent feedback from past actions.
Structure your analysis as follows:
- Evaluative: Rate the feedback sentiment numerically: +1 for positive feedback (signifying the action was appropriate or correct), -1 for negative feedback (indicating the action was inappropriate or incorrect), and 0 for neutral feedback or if there's no clear evaluative feedback.
- Corrective: If a previous action was mistaken, identify the correct action that should have been executed, using the action numbers (0, 1, 2, 3, 4, 5). This feedback should only relate to the prior action. If the feedback's sentiment is positive or neutral, or it doesn't suggest an alternate action, label the corrective feedback as None.
- Advice: Suggest the next step the agent should take, selecting from the action options (0, 1, 2, 3, 4, 5). If the feedback doesn't suggest a specific next action or focuses only on rectifying a past error without recommending a subsequent step, indicate the advice as None.

Your response should be formatted as: 'Evaluative=value Corrective=value Advice=value'
Example:
- Given Feedback: "Action=0, No, turn left "
- Evaluative=-1 Corrective=None (no correction for past action) Advice=2

- Given Feedback: "Action=1, Now go up "

- Evaluative=0 Corrective=None Advice=0

- Given Feedback: "Action=0, yes ,turn left"

- Evaluative=1 Corrective=None Advice=2

-Given Feedback: "Action=0, No, you should have turned left"

- Evaluative=-1 Corrective=2 Advice=None (no advice for next action is given)

- Given Feedback: "Action=4, Yeah!"

- Evaluative=1 Corrective=0 Advice=None (no advice for next action is given)

- Given Feedback: "Action=1, Continue like this!"

- Evaluative=1 Corrective=0 Advice=1 (same as previous action)"