



Empathic Robot for Group Learning: A Field Study

PATRÍCIA ALVES-OLIVEIRA, Instituto Universitário de Lisboa (ISCTE-IUL), CIS-IUL, and Group on Artificial Intelligence for People and Society, GAIPS, from INESC-ID, Portugal

PEDRO SEQUEIRA, Northeastern University, College of Computer and Information Science, USA

FRANCISCO S. MELO, Instituto Superior Técnico, Universidade de Lisboa, and Group on Artificial Intelligence for People and Society, GAIPS, from INESC-ID, Portugal

GINEVRA CASTELLANO, Department of Information Technology, Uppsala University, Sweden

ANA PAIVA, Instituto Superior Técnico, Universidade de Lisboa, and Group on Artificial Intelligence for People and Society, GAIPS, from INESC-ID, Portugal

3

This work explores a group learning scenario with an autonomous empathic robot. We address two research questions: (1) Can an autonomous robot designed with empathic competencies foster collaborative learning in a group context? (2) Can an empathic robot sustain positive educational outcomes in long-term collaborative learning interactions with groups of students? To answer these questions, we developed an autonomous robot with empathic competencies that is able to interact with a group of students in a learning activity about sustainable development. Two studies were conducted. The first study compares learning outcomes in children across three conditions: learning with an empathic robot; learning with a robot without empathic capabilities; and learning without a robot. The results show that the autonomous robot with empathy fosters meaningful discussions about sustainability, which is a learning outcome in sustainability education. The second study features groups of students who interact with the robot in a school classroom for 2 months. The long-term educational interaction did not seem to provide significant learning gains, although there was a change in game-actions to achieve more sustainability during game-play. This result reflects the need to perform more long-term research in the field of educational robots for group learning.

CCS Concepts: • **Human-centered computing** → **Collaborative interaction; Empirical studies in HCI; User studies; Field studies;**

This work was partially supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2019, through the project AMIGOS Grant No. PTDC/EEISII/7174/2014, the Carnegie Mellon Portugal Program and its Information and Communications Technologies Institute, under Project No. CMUP-ERI/HCI/0051/2013, and by the EU-FP7 Project EMOTE under Grant Agreement No. 317923. P. Alves-Oliveira acknowledges an FCT Ph.D. Grant No. SFRH/BD/110223/2015.

Authors' addresses: P. Alves-Oliveira, Instituto Universitário de Lisboa (ISCTE-IUL), CIS-IUL, and Group on Artificial Intelligence for People and Society, GAIPS, from INESC-ID, Av. das Forças Armadas 376, Lisbon, 1600-077, Portugal; email: patricia_alves_oliveira@iscte.pt; P. Sequeira, Northeastern University, College of Computer and Information Science, 202 W VH 360 Huntington Avenue, Boston, MA, 02115, USA; email: psequeira@ccs.neu.edu; F. S. Melo, Instituto Superior Técnico, Universidade de Lisboa, and Group on Artificial Intelligence for People and Society, GAIPS, from INESC-ID, IST Taguspark, Av. Prof. Dr. Cavaco Silva, Porto Salvo, 2744-016, Portugal; email: fmelo@inesc-id.pt; G. Castellano, Department of Information Technology, Uppsala University, Lägerhyddsvägen 2, 751-05, Uppsala, Sweden; email: ginevra.castellano@it.uu.se; A. Paiva, Instituto Superior Técnico, Universidade de Lisboa, and Group on Artificial Intelligence for People and Society, GAIPS, from INESC-ID, IST Taguspark, Av. Prof. Dr. Cavaco Silva, Porto Salvo, Portugal; email: ana.paiva@inesc-id.pt.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2019 Association for Computing Machinery.

2573-9522/2019/03-ART3

<https://doi.org/10.1145/3300188>

Additional Key Words and Phrases: Social robotics, human-robot interaction, empathy, collaborative learning, group learning, learning gains, education

ACM Reference format:

Patrícia Alves-Oliveira, Pedro Sequeira, Francisco S. Melo, Ginevra Castellano, and Ana Paiva. 2019. Empathic Robot for Group Learning: A Field Study. *Trans. Hum.-Robot Interact.* 8, 1, Article 3 (March 2019), 34 pages. <https://doi.org/10.1145/3300188>

1 INTRODUCTION

Learning is an intrinsic human ability. We learn both from our own experience and from our peers. However, learning is not just about improving performance or assimilating new knowledge. It is also about analyzing new situations, understanding different perspectives, using knowledge to find commonalities between distinct situations, discussing, and even becoming competent at solving conflicts. Social learning, whereby we learn with and from our peers, teachers, parents, and others, plays a fundamental role in most of these broad forms of learning (Piaget 2013). It is not surprising then that the educational environment stimulates social collaboration between peers during learning. Collaborative learning has been associated with improving attitudes toward school, fostering achievement, developing thinking skills, promoting interpersonal and intergroup relations (Blumenfeld et al. 1996).

However, as technology evolves, so do learning paradigms. Currently, technology offers novel learning tools that complement the classical learning paradigms. Massive open online courses (MOOCs) (Li et al. 2014) and intelligent tutoring systems (ITS) (Nwana 1990) are just two illustrative examples of new learning tools that are available through technology. Another change in the teaching and learning paradigm comes from serious games (Ritterfeld et al. 2009). Serious games are designed for a purpose other than pure entertainment. They have been around since at least the 1950s, and their applications in education are well-documented (De Gloria et al. 2014).¹ Collaborative serious games, in particular, combine the advantages of serious games with social learning, and some studies have suggested that they support learners in articulating the knowledge that would otherwise have remained intuitive (van der Meij et al. 2011). Although some researchers questioned the effectiveness of collaborative serious games, research in this area is still scarce and often ambiguous (Wouters et al. 2013).

In this article, we describe the use of a *robotic tutor in the context of a collaborative serious game*. In particular, the research described herein seeks to combine the positive aspects of intelligent and peer tutoring, such as personalization and adaptation to the learner, with collaborative serious games, such as social learning and gamification. This offers a new role for robots as technological tools in education.

Previous studies have explored the use of artificial robotic tutors and peers in both classroom environments (Kanda et al. 2004; Miyake and Okita 2012; Tanaka et al. 2007) and as entertainment partners (Michalowski et al. 2007). However, such use is often limited to one-robot-one-user interactions, which significantly limits the social component of learning. By adopting a collaborative serious game as the interaction domain, we move from such typical one-robot-one-user interaction to a much richer scenario that involves one robot and two users, thus, a group interaction. In addition, some studies have investigated the effect of long-term interactions between children and robots regarding the quality of the relationship (Westlund et al. 2018) and other variables (Leite

¹Interestingly, research suggests that while serious games may be more effective in terms of learning, they are not always more motivating than conventional instruction methods (Wouters et al. 2013).

et al. 2013a); however, few studies have looked at the impact of long-term interactions and their effects on learning gains, which is also a contribution of this article.

Our ambitious application scenario poses several technological challenges, particularly in endowing the robot with a behavior that is both socially plausible and able to successfully accomplish the pedagogical goals of the activity. Our robotic tutor must be both socially competent and successful in teaching. It should also be able to perceive the difficulties faced by the students, not only from their explicit behavior but more importantly from their implicit behavior. It should be able to understand their individual affective state and the “emotional climate” between them to intervene in a way that is adequate. In other words, a robot should be able to *empathize* with the human users, individually and as a group.

A recent survey discussed the importance of empathy for an artificial agent (robot or virtual) that interacts with humans (Paiva et al. 2017). In the context of education, the empathy of the teacher toward the students was also shown to impact the learning process and outcomes (Feshbach and Feshbach 2009). In a meta-analysis conducted by Cornelius-White (2007), empathy was one of the variables that if present in teachers is associated with positive students outcomes. In a subsequent meta-analysis conducted by Roorda et al. (2011), empathy was included as one of the variables in the teacher-student’s relationship that is associated with their engagement and achievements in school. This demonstrates the importance of empathy in tutor-student relationships and how it influences students’ educational outcomes. However, few studies have measured this impact on learning outcomes. This article describes an autonomous and empathic robot tutor that interacts with multiple learners via collaborative serious game and investigates the impact, both immediate and long-term, of the robot on the student’s learning performance. Two field studies were performed, leading to a deeper understanding of empathy in learning and the long-term effects of having an autonomous robot in school for educational purposes.

1.1 Contributions

The work presented herein was developed in the context of the EMOTE project.² We investigated the use of an *autonomous and empathic robot tutor for collaborative group learning*. We explored the impact of a robot interacting and teaching groups of students in their classroom in relation to learning gains. The contributions of this article include the following:

- (1) We conducted a short-term evaluation study to investigate the immediate impact of the empathic robot in fostering collaborative learning.
- (2) We conducted a long-term term study to investigate the long-term impact of the empathic robot in terms of learning.
- (3) We deployed an autonomous robot in the context of real-world classrooms for group learning.

The road-map of this work starts with Section 2 presenting the state of the art regarding robots in education, groups of humans and robots, and empathy in social robots; Section 3 details on the collaborative learning activity, including the learning goals and the game-play dynamics; Section 4 refers to the development of the educational and social behaviors for the robotic tutor; Section 5 explains the study method, hypothesis, measures, and materials used; Section 6 describes the short-term study and Section 7 the long-term study; Section 8 presents the general discussion and conclusion, including limitations and future work.

²EMOTE project: <http://www.emote-project.eu>, which stands for EMbodied Perceptive Tutors for Empathy-based learning project.

1.2 Our Previous Work

This article focuses on the evaluation of the robotics system with respect to learning gains in students. The details of the design and implementation of the robot are not a main contribution of this article and can be found in previous publications. In particular, the robot behaviors and the development of the empathy module of the autonomous robot are detailed in Alves-Oliveira et al. (2015) and are summarized in Section 4.3.1 of this article. The development of the collaborative Artificial Intelligence (AI) that sustains the robot's game-playing and pedagogical decision-making abilities is detailed in Sequeira et al. (2015) and is overviewed here in Section 4.3.2. The educational dialogue dimensions for collaborative learning designed for the robot are detailed in Alves-Oliveira et al. (2014) and summarized in Section 4.3.4. Finally, the design, development, and validation of the "Restricted-perception Wizard-of-Oz" methodology, which allowed the successful development of the social and educational behaviors of the autonomous robot, are detailed in Sequeira et al. (2016), and we summarized the main steps in Section 4.2.³

2 STATE OF THE ART

The experiences we have during our childhood shape to some extent the way we think, grow, feel, and behave. It is thus important to surround children with nurturing and safe learning environments. The way children learn, however, is being transformed with new technologies for education, such as computers and tablets, that are enhanced with serious games (Savage and Sterry 1990) or computer-supported learning activities, to foster learning acquisitions. Robots in particular hold promise to facilitate learning outcomes and promote enjoyment during learning (Kennedy et al. 2015a). Recent events, such as the R4L (or Robots 4 Learning) series of workshops,⁴ are an example of the interest that this area is capturing in research and the potential that robots have in education. In this article, we present research that advances the state of the art in the area of social and empathic robots for education, in particular for collaborative group learning scenarios. Thus, we now provide an overview of these areas to contextualize the contributions of the article.

2.1 Robots in Education

Educational robots are a subset of educational technologies in which robots are used as platforms or tools for students' learning, usually in subjects such as maths, problem solving, chemistry, and so on. The use of robots as a medium to learn and understand curricular subjects started in the 60s with Papert's work in which he introduced the concept of "robots as manipulatives" (or physical objects) that are specifically designed to foster learning (Papert 1980). An example is the LEGO Mindstorms, which are used to teach STEAM-related curricular topics, providing a new tool for education (Hendler 2000; Khine 2017). Furthermore, as robot technology matures, robots can be used as social actors in a classroom as a way to deliver educational content, instruct, foster discussion, challenge, scaffold, and support the learning of children in a social-intelligent manner. In fact, reviews on the applicability and potential of robots in education shows that robots are being developed for children across different learning domains (Belpaeme et al. 2018a; Mubin et al. 2013). Despite the potential of robots in learning, a systematic review has shown that to have an impact on learning gains, they need to be skilfully used by teachers to attend to the students' needs. Otherwise, learning gains are not visible (Spolaôr and Benitti 2017).

³The source code of all the robot's components is publicly available and can be retrieved at: <https://github.com/emote-project/Scenario2>. Further details on each component can be found in the "Downloads/Components" and "Publications/Deliverables" sections of the project's website at <http://www.emote-project.eu/>.

⁴The workshops and special issue can be found in <https://r4l.epfl.ch/>.

There is investment in the field of Human-Robot Interaction (HRI) to develop robots that are used for learning. Pioneering research by Kanda et al. (2004) features ROBOVIE, a robot for teaching English to Japanese children in an elementary school context. This was one of the first field studies in educational HRI, conducted over a period of 18 consecutive days in a school. The results showed that children who exhibited a lot of interest during the starting phase had a significantly elevated English score, and the robot indeed acted as a motivational factor for learning in these cases. In South Korea, the IROBI robot was also endowed with didactic content to support young children with learning English as a second language (Han et al. 2008). The robot was placed in a class with very young children over a long period of time. In the same application domain, the EU H2020 L2TOR project⁵ aims to study if robots can be used as tutors to support teaching preschool children a second language (Kennedy et al. 2016). A review of the applicability of robots for second language acquisition was performed by Chang et al. (2010) in which they have characterized the existing robots used and the instructional media explored for this type of task. Other projects with robots for children have dedicated attention to the investigation of robots as tools for learning and supporting positive interactions, such as the Socially Assistive Robotics, an NSF Expedition in Computing project.⁶

Robots have also been used to support handwriting abilities using the paradigm of “learning-by-teaching” (Frager and Stern 1970) in which children act as tutors of the robot, providing it with feedback for a better writing performance. This paradigm is known to benefit children’s self-esteem, provide practice with hand-writing without noticing, and provide engagement in a so-called “protégé effect,” which is a sense of responsibility over the robot’s performance (since they are instructed to be the robot’s teachers) (Chase et al. 2009). For example, Tanaka and Matsuzoe (2012) conducted a 6-day field trial with young students and a robot in school, and the results showed that a robot can help children efficiently learn new English verbs, when children give instructions to the robot. Additionally, projects such as the Co-writer project⁷ have studied the role of students as teachers of the robot. A study in the scope of the Co-writer project was conducted in school in which a group of children gathered around the NAO robot to provide feedback on its mistakes as the robot improved according to children’s feedback (Lemaignan et al. 2016). While providing feedback to the robot, children are put into small groups who are responsible for teaching the robot to write better. In a study in which the NAO robot acted as a student that needed to learn how to write and the children served as teachers who helped it write better, results showed high levels of commitment and engagement from children embracing this task (Hood et al. 2015). This demonstrated the promising results of using this educational system. Another study investigated interpersonal distancing of children both toward a human adult or a robot facilitator within a collaborative activity (Chandra et al. 2015). The scenario involved two children performing a collaborative learning activity following a learning-by-teaching approach, which included writing a word/letter on a tactile tablet. The study showed that children felt more responsible and provided more corrective feedback when the robot was present than when a human mediator was present (replacing the robot). This suggests that the role a robot can play can have an impact in the type of interactions that emerge, particularly corrective feedback.

Beyond the typical curricular activities, social robots are also being used for social and emotional learning. The research by Jimenez et al. (2014) described a study in which the behavior of a robot prompts constructive interaction with a learner when compared with two students learning the same task. The types of prompts and interactions built into the robot as it learns together with the

⁵L2TOR project: <http://www.l2tor.eu/>.

⁶Socially Assistive Robotics, an NSF Expedition in Computing project: <https://robotshelpingkids.yale.edu/>.

⁷Co-writer project: <http://chili.epfl.ch/cowriter>.

children lead to better performance of the robotic condition. Another study focused on fostering children to develop a growing mindset as part of social and emotional training (Park et al. 2017b). Therefore, in a scenario featuring a child playing puzzle games with a robot, a fully autonomous robot was built with the capability to exhibit “behaviors suggestive of it having either a growth mindset or a neutral mindset” (Park et al. 2017b). The results of a study that compared two types of robots in the same scenario have shown that children who played with a growth mindset robot self-reported having a stronger growth mindset and tried harder during the task.

In a long-term study, Serholt (2018) investigated interaction breakdowns between children and a robot in a learning task. In this individual learning scenario, breakdowns in the interaction were associated with “the robot’s inability to evoke initial engagement and identify misunderstandings, confusing scaffolding, lack of consistency and fairness, and controller problems.” In another long-term study in which a robot acted as an agent for learning, Jones and Castellano (2018) concluded that if a robot provides personalized and adapted scaffolding to students based on their learning, they can better regulate their own learning (Jones et al. 2017). Furthermore, in a study of two consecutive weeks in a school, a peer-robot that exhibited behavioural personalization was found to have a positive influence on learning when interacting with children, compared to a robot that exhibited non-personalization. Personalization of the robot was defined in terms of non-verbal behaviour, linguistic content, and performance alignment. Specifically, the results from this study showed that children exhibited significantly increased learning only in the novel learning task in the personalized condition (Baxter et al. 2017). Although these three scenarios present themselves as extremely rich and challenging for social learning, especially because a robot was deployed in school for long-term educational gains, they were built for one-robot-one-student interactions. This article goes beyond the current state of the art by exploring collaboration in a group context using a robot designed to support learning. In addition, we evaluated the learning gains, an aspect rarely measured in studies that use robots for education. In fact, usually studies evaluate other variables, such as likability and engagement between children and robots.

2.2 Groups of Humans and Robots

As mentioned, the majority of the application scenarios developed thus far to study humans and robots are designed for one-on-one interactions in which one robot interacts with one person. Even in scenarios in which the robot is placed in a classroom with many children, the type of interactions often designed consider one-on-one interactions, e.g., Belpaeme et al. (2018b). For this study, we were interested in scenarios using groups of two or three students who can learn together with the support of a social robot. According to Du and Wang (2016), dyads and triads are considered groups, with dyads being considered the smallest type of a group if they share common and dependent elements. In the case of our research, dyads of students share the same learning context as part of the group.

Groups in HRI can be studied according to different perspectives, such as (1) groups of humans interacting with one robot, (2) groups of robots interacting with one human, or (3) groups of humans interacting with groups of robots. In addition to workshops organized regarding this topic (Jung et al. 2017), relevant studies have been conducted. For example, a study by Fraune et al. (2015b) showed that different number of robots (namely, a single robot or a group of robots) and the type of robots (anthropomorphic, zoomorphic, or mechanomorphic) determine the attitudes, emotions, and stereotypes that people hold when interacting with them, with anthropomorphic robots in groups being one of the preferred choices. In a field study at a university, Fraune et al. (2015a) studied how participants respond when robots (individually and in groups) enter their common space. The results showed that although participants reported enjoying interacting with both individual and robots in groups, they interacted more with groups of robots. In addition to

this, the characteristics of social robots can affect the interactions. For example, entitative groups of robots (robots designed to look and act similar to each other, such as those that have a similar appearance and a shared goal) compared to single robots, were evaluated more negatively (Fraune et al. 2017). In another study, the frequency for which robots acted as moderators affected the social and task-related features, namely, group cohesion and task performance, in multi-party interactions (Short and Mataric 2017). These studies show that a robot's behavior in a group should be carefully designed as their behavior and appearance influence interactions.

Nonetheless, the perception of robots in groups not only depends on the robots' behaviors but is also influenced by the characteristics of the people. A study conducted by Correia et al. (2018) showed that when people only observe robots (before any direct interaction with them), they tend to choose robots that exhibit relationship behaviors (e.g., robots that foster a group climate) over competitive robots (e.g., robots that are more focused in succeeding and winning). However, after a direct interaction, the results seem to change, with participants who were competitive preferring a more competitive robot, and participants who were not as competitive preferring robots with relationship-driven characteristics. This study reflects that membership preferences between groups of humans and groups of robots goes beyond robot's characteristics, extending to the characteristics of each person. In another study regarding group membership, Sembroski et al. (2017) found that in-group and out-group perception between humans and robots can lead to more conformist behaviors from people, depending on the request type and level of authority. A different study investigated a robot's potential to shape trust in groups of humans, concluding that the robot that exhibited vulnerable behavior (in comparison with a neutral robot) promoted more group engagement and social signs of trust, with people providing more support in times of failure (Strohkorb Sebo et al. 2018). Furthermore, a study dedicated to the investigation of non-verbal behavior between a robot and multiple people concluded that the gaze of the robot influences people's perception of the motion of the robot and that, in turn, affects the perception of the robot's gaze (Vázquez et al. 2017). Additionally, when incorporating robots in unstructured social environments, such as in shopping malls or crowded city streets, it is important for the robot to exhibit adequate motion for collision avoidance to navigate fluently between groups of people (Mavrogiannis et al. 2018). Therefore, considering mutual dependency between the behavior of robots and the behavior of the humans is important when designing and evaluating robotic systems aimed to be social in group contexts.

In relation to educational contexts and groups, Strohkorb et al. (2015) used interactive role-playing to help children improve their emotional intelligence skills. By focusing on finding ways to analyze non-verbal and verbal behavior to detect if a child was high or low in social dominance, the scenario featured groups of children, and the robot helped them in their social relations. Despite being a group scenario, the collaborative nature of the task was not the focus of the work. In fact, according to Dillenbourg (1999), collaborative learning pertains to a situation in which particular forms of interaction among people are expected to occur that trigger learning. In robots applied to education, social robots can play roles, such as of a peer (learning together with a group of students); a tutor (teaching a group of students); a facilitator/mediator (mediating the learning interactions and interventions and helping the group to follow productive learning); a supervisor (supervising the work being done by students and providing feedback); or even a friend (supporting the students emotionally) (Alves-Oliveira et al. 2016; Broadbent et al. 2018; Zaga et al. 2015). This demonstrates the richness of interactions that can be explored in the educational domain when using robots to foster learning. Although there is a wide variety of work in HRI exploring robots as tutors in a classroom context, thus far, the type of activity and interactions established with robots have been fundamentally individual interactions.

2.3 Empathy in Social Robots

Empathy is an essential ingredient for any successful relationship. When attempting to define empathy, we come across various definitions. Empathy has been related to affective responses toward others (affective empathy) and a cognitive understanding of the emotional state of others (cognitive empathy). Hoffman (2001) brought attention to the processes that underlie empathic responses, defining empathy as the psychological process that leads to feelings that are congruent to the other person's situation more than to one's own. From a more behavioral perspective, Davis (2018) associated empathy with the responses of someone to the experiences of another person. Holistically, Preston and De Waal (2002) considered empathy as a concept that relates sympathy and emotional contagion. The expression of empathy is foundational with respect to interpersonal relationships and with our ability to communicate. Indeed, it is what connects us emotionally and intellectually.

In an educational context, there is a general understanding from teachers that empathy promotes positive interactions, a more supportive classroom climate, and enhances student-centered practices (McAllister and Irvine 2002). In a meta-analysis conducted by Cornelius-White (2007), empathy was one of the variables associated with positive students outcomes when present in teachers. When robots are placed as collaborative partners within a group, features such as their ability to communicate and relate become relevant. Interactions with robots that are more open-ended and involve a high degree of collaboration that must be natural and engaging; thus, empathy needs to be addressed and modeled in those robots (Paiva et al. 2017). Recent research has been devoted to the implementation of empathy in robots in diverse application areas, such as in health care with socially assistive robotics (Tapus and Mataric 2007). In these cases, robots are perceived as having more empathy if they provided appropriate answers during a dialogue (Riek et al. 2010). In this line of research, robots that display accurate empathic behaviors enable the perception of a closer relationship with a human (Cramer et al. 2010). Additionally, robots with empathy were perceived as being friendlier during an entertaining scenario (Leite et al. 2013b).

Empathy also plays a role in child-robot interactions. In a study conducted by Leite et al. (2014), they explored whether a robot with empathic capabilities can influence the perception that children have toward the robot, specifically social presence, engagement and support. Indeed, in a chess game scenario, the robot that displayed empathic behaviors was perceived positively by the children, and the perceived levels of social presence, engagement and support were stable and high during a long-term interaction (Leite et al. 2014). This research was important for the study of robots with empathy and children in real-world settings. In a broader view of empathy in robots, Paiva et al. (2018) dedicated a book chapter on the creation of more humane machines in which successful interactions between humans and robots is associated with robots that are endowed with emotional processing and responses. Our research goes beyond the current state of the art as it considers how a robot can model empathy in a group setting in a collaborative learning activity.

3 COLLABORATIVE LEARNING ABOUT SUSTAINABLE DEVELOPMENT

A well-designed scenario for collaborative learning increases the probability that some positive interactions occur, thus leading to learning experiences (Dillenbourg 1999). The design of the scenario for collaborative learning is of utmost importance, since the interactions between the members are crucial. As stated before, thus far social robots have been used extensively in domains in which the robot is able to establish one-on-one interactions with the learner, thus supporting problem solving, personalization, feedback, and scaffolding. However, in collaborative learning, other elements come into play, such as perspective taking and understanding the consequences of actions. For the scenario in our studies, we used a collaborative multi-player game targeting issues of sustainable energy consumption. This game, which we shall refer to as M-Energities (for

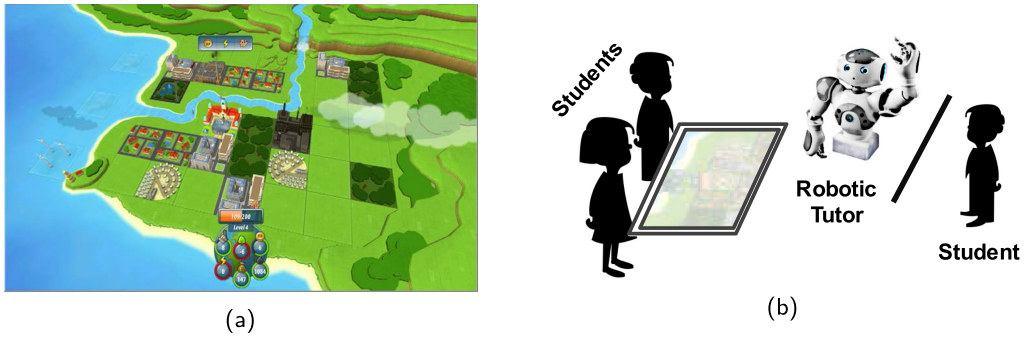


Fig. 1. Learning activity used in this study: (a) a screenshot of the M-Enercities game used in the study, which is a multi-player, turn-based, collaborative version of the Enercities game about sustainable development; (b) the learning space shows the interaction of either two students with a robotic tutor or three students playing the M-Enercities game over an interactive touch-table.

Table 1. Learning Goals Supported by the Collaborative Group Learning Game, M-Enercities

Learning goal 1	<i>Factual knowledge</i> about different energy sources
Learning goal 2	Raise awareness about <i>trade-offs</i> and the existence of multiple and possibly conflicting <i>perspectives</i> in relation to sustainable development
Learning goal 3	Raise awareness of <i>personal values</i> in relation to sustainable development

Multi-player Enercities), corresponds to a multi-player, collaborative version of the original game Enercities (Knol and De Vries 2011) (see an image of the game’s interface in Figure 1(a)).

3.1 Learning Goals

The game was adjusted to the current schools’ curriculum and was easily integrated into the school activities, allowing for three children to play the game, two children and a teacher, or two children and a robot (see Figure 1(b) for an overview of the group learning setting). The group of three work together to build a sustainable city, thus learning about sustainable development. This game is in line with Constructivism’s basic idea that knowledge and meaning are built during social interaction and cooperation (Steffe and Gale 1995). As a collaborative group learning activity, M-Enercities motivates players to discuss and decide together how to build a sustainable city, providing exactly the type of social interaction dynamic to foster communication and cooperation as means to learn about sustainability. In fact, according to scholars, sustainable education is all about having meaningful discussions in which multiple perspectives and trade-offs are exchanged by expressing personal values (Antle et al. 2015; Fior et al. 2010). In light of how sustainable education should be taught, the M-Enercities game supports learning goals related with factual knowledge about sustainable development, raised awareness for trade-offs and multiple perspectives concerning sustainability, and the role of personal values. These learning goals are elaborated in Table 1.

3.2 Game-play

The game is structured in levels that players need to master to continue their city. For example, to proceed to the second level in the game, players must make the population of the city grow to a

certain amount by building residential areas. However, at the same time, if the city runs out of non-renewable resources, players may get into trouble, because the city is not sustainable. Additionally, the game-play is based on a turn-taking dynamic where, at each each turn, one player decides what to do and the rest of the group assists in this decision process, discussing how they would like to build their city. For them to decide, they have to take into consideration the city indicators and how their actions influence the sustainability of the city. Each decision can have positive and negative effects on the “environment,” “economy,” and citizens’ “well-being,” which is indicated by a score for each of these factors. Players can choose among a set of actions that advances the game:

Making a construction—Players can build a construction, such as a park, a market, or an industry. The different constructions influence the city indicators differently. Thus, a park positively affects the environmental indicator of the city, an industry positively influences the economy, and a market positively influences the well-being of the citizens. However, at the same time, these constructions can have negative effects on other indicators. Players can also invest in the city’s energy by building energy supplying constructions (e.g., windmills) or increase the city’s size and thereby the population by constructing more residential areas.

Performing an upgrade—Players can invest in making an upgrade on previously created constructions. They can, e.g., add a recycling facility in an already built industry to upgrade it and improve the environment without wasting additional resources. Depending on the upgrade performed, it can influence different indicators.

Applying a policy—Players can implement new policies in their city, such as implementing an energy education program. Policies are applied in the city-major building as a way to represent how the real world functions.

Skipping a turn—Players can choose to skip a turn for the constructions and policies to come into full effect. When skipping a turn, constructions, upgrades, and policies present in the city augment their effect, e.g., if the players have chosen to build an industry to increase the economy of the city, the city will be richer when more turns have passed. Consider turns passing as time passing (or as a time indicator). Therefore, the more the players skip turns, more benefits over previous constructions, upgrades, or policies will occur.

To support turn-taking and to make every member of the group participate, there are virtual buttons located at different sides of the table, allowing players to perform different actions in turns. Since each decision can have both positive and negative effects, players must be aware of the trade-offs and the impossibility of creating a perfect city without any sacrifices. After reaching a group decision, the player closer to each button presses it, making a game action on their city.

4 ROBOTIC TUTOR

To fulfill its role as an autonomous empathic tutor in the collaborative M-Energities game, the behavior of our robot can be seen at two distinct levels: an *activity-related* level and an *social interactive* level. Activity-related behaviors concern the decisions of the robot in the pedagogical activity itself. In our case, it involves all game actions for M-Energities and the necessary game-play adaptation mechanisms. The social interactive behaviors involve all verbal and non-verbal expressive behaviors of the robot during interactions with the group of students, such as the dialogue management and physical animations. Both types of behaviors are selected based on information about the state of the game and contextual information about the physical environment provided by different audio-visual capture devices. Given the intended empathic nature of the robot, its behavior also depends on the individual and collective affective state of the students. This section overviews the design and implementation of the behavioral components of the robot, including

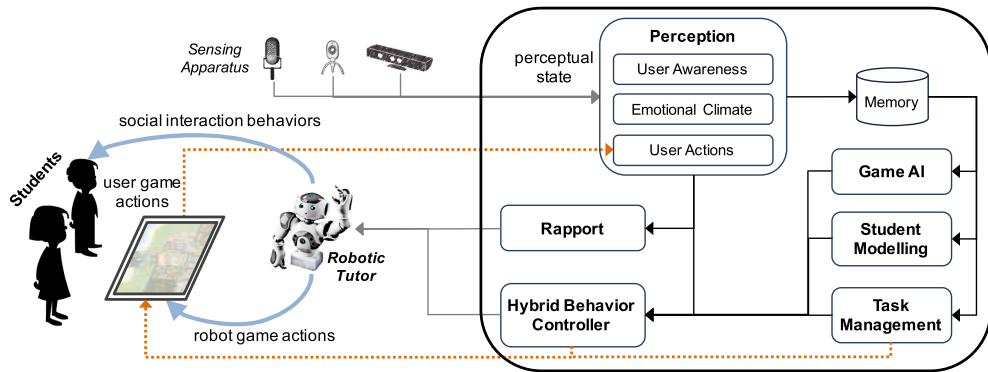


Fig. 2. The overall robot system architecture implemented to autonomously control the robotic tutor during the interaction with the students in our studies. Light-blue arrows denote the robot's actions, and dotted orange lines represent communication between the system and the M-Energities game application.

task-specific behaviors, empathic behaviors, adaptation to the students, and behaviors driven by the learning goals.

4.1 Architecture Overview

The overall architecture supporting the robot used in our studies is depicted in Figure 2. The system comprises the following main modules⁸:

Perception—Module responsible for processing all information from the robot's environment, including information on the student's emotional state, social behavior and actions performed during the learning activity. It comprises three main components: the *User Awareness* component processes the students' state and actions; the *User Actions* component manages the students' M-Energities game input, providing the robot with the necessary task-related context for the interaction; finally, an *Emotional Climate* component is responsible for detecting group-level emotional state.

Memory—This module keeps track of past events, and is subdivided in *recent* and *past event* memories. *Recent Event Memory* stores the recent actions of the students in the game, collected directly from the User Actions perception component. Long-term interaction, however, requires the robot to track information from previous sessions. For this reason, the information stored in the recent memory is moved to the *Past Event Memory* at the end of each session.

Task Management—This module manages the execution of the learning activity itself, e.g., starting and ending the activity and specifying which students are interacting.

Student Modeling—This module is responsible for collecting information regarding the student's performance during the game session, using it to track possible changes in the learning and emotional state of students. Such information is used by the robot in later sessions to address specific learning challenges.

Rapport—This module regulates the robot's rapport during the interaction with the students. For example, it adjusts the robot's speech volume to the student's to ensure a smooth communication; it shifts the robot's gaze toward the active speaker to provide a more natural

⁸Technical details about each component can be found at www.emote-project.eu/components.

interaction; it interrupts the robot's speech when a student is speaking; and performs back-channeling behaviors after the users' responses.

Game AI—This module manages all the robot's actions in the M-Enercities game. It uses the student's past actions (retrieved from Memory) to learn the strategies being used by them during the activity, and generates possible actions according to the current state of the game.

Hybrid Behavior Controller—This is the core module controlling the robot's social interaction behavior. It decides which actions the robot should play in the M-Enercities game (informed by the Game AI) and also how to structure the dialogue with the students (informed by the Emotional Climate and the other modules).

While some of the aforementioned modules are standard in HRI domains, some are specific to our scenario, namely in what refers to the interaction with multiple students. Below, we discuss in greater detail the methodological and technological considerations that drove the design of such components.

4.2 Designing the Interaction Behaviors of the Robot

As an empathic tutor, the robot should be able to play the game and to interact in a social and empathic manner with the students, raising awareness for personal values when considering sustainability. The robot should do so by setting a good example when choosing its game actions, explaining the reason behind each play in light of its "personal values," namely, achieving a balanced development of the city. We note that, in the context of our scenario, no personal values are wrong or right, and the students are not expected to adopt the robot's personal values. Instead, it is simply a means to open up the discussion and raise awareness of others' perspectives, as active participation is a sign of learning in sustainable development education (Lave and Wenger 1991). Therefore, to design the social behavior of the robot, we adopted the restricted-perception Wizard-of-Oz (WoZ) methodology detailed in Sequeira et al. (2016), which can be summarized in the following steps:

- (1) We gathered data from mock-up sessions where two students collaboratively interacted with a school teacher playing M-Enercities. The goal was to gain insight in common pedagogical and empathic strategies used by real teachers in our learning activity.
- (2) Data collected during the mock-ups was used to build the modules responsible for the perception, basic game behavior, and interaction behavior of the robot.
- (3) We conducted Restricted-Perception WoZ studies, in which experts remotely controlled the robot during the interaction with students in the M-Enercities game. Unlike the standard WoZ paradigm in HRI, when applying a Restricted-Perception WoZ method, the experts controlling the robot have access only to processed observations from the interaction, similar to those which will drive the robot's autonomous behavior. This means that the decision-making of the wizard will be limited to the same perceptions that the autonomous robot will have over the interaction. This paradigm allows the operator to focus on the relevant aspects of the robot's social interaction.
- (4) Using the data collected during the Restricted-Perception WoZ studies, we created the Hybrid Behavior Controller. The controller was built from two key elements: *interaction strategy rules*, in which we encoded expert domain knowledge, e.g., explicit behavior patterns observed during the mock-up sessions and WoZ studies; and a *mapping function*, which identifies more complex behavior patterns discovered from the data using a Machine Learning (ML) approach.

4.3 Implementing the Interaction Behaviors of the Robot

Although this section describes work that is not a direct contribution of this article, it is crucial to understand the design and implementation decisions that were taken regarding the behavior of the autonomous robotic tutor used in the studies reported here. In particular, we discuss the core modules that support the robot's social and empathic interaction with multiple students in the context of our learning activity (see Figure 2).

4.3.1 Emotional Climate. For the robot to interact with a group of students in an empathic and emotionally intelligent way, it is fundamental that it is able to *detect* and *recognize the emotional state of the group*. Emotional climate is a central element in social group interactions between humans and has been studied in many group contexts. We consider emotional climate to be the valence state of a group-level emotion at a given time. Following the discussion in Alves-Oliveira et al. (2015), in our scenario, the behavioral and emotional state of the students at any given time changes the emotional climate of the group at that time. This means that a positive emotional can be detected from the students expressing positive facial expressions, demonstrating joint attention in the educational task by looking at the table where the interaction takes place, and so on. Conversely, a negative emotional is detected whenever students look away from the task and seem distracted or bored, and so on. Emotional climate influences the behavior of the robotic tutor by changing the content and the way that certain utterances are performed. For example, if students are taking more time than the usual to decide what to do and a negative emotional climate valence is detected (e.g., boredom), the robot intervenes to maintain engagement by saying: “*What should we do now?*” However, if a positive emotional climate valence is detected (e.g., engagement), the robot may say: “*We are playing well*” as a positive reinforcement. These constitute examples of empathic behaviors designed for the robot.

4.3.2 Game AI. The Game AI module is detailed in Sequeira et al. (2015) and ensures that the robot tutor is not only able to play the game competently but also discusses the impact of each action performed by the group in the construction of a sustainable city. The robot's game-play promotes collaboration within the group and comprises a *game-playing* and a *social* component. The *game playing component* (planner) is designed to accommodate a specific educational strategy, e.g., if the goal is to achieve a “balanced” strategy, it favors actions leading to game states where all scores (environment, economy and well-being) are as high and even as possible. It also detects game situations with the potential to provide interesting pedagogical opportunities, e.g., when the level of natural resources is low it suggests game actions that spend fewer resources. The *social component* uses information about recent plays to build a model of the students' game strategies. It allows the robot to intervene during and after the players' actions in the game, e.g., the robot is able to suggest more suitable alternative plays in certain situations and explain the desired effect of such decisions over the city's development. Such game-play model also allow the robot to adapt its own strategy so to follow the perceived group strategy, a fundamental aspect of its game behavior due to the collaborative nature of the activity. For example, if the students are playing in an environmentally aware fashion, the robot's strategy will also be more environment-friendly.

4.3.3 Student Modeling. The Student Modeling component in Figure 2 follows the discussion in Jones et al. (2015). It provides *task and domain knowledge*, i.e., information both on the learning activity and on the knowledge and skills that the student is expected to acquire. The Student Modeling uses the information from the User Action module to update a high-level description of each student. Such description includes the student's task performance and estimated domain knowledge. Both are stored in the Memory module to be used in a pedagogical manner during the interaction, e.g., the robot can show its support with respect to the users' difficulties, and by

```

Category: Strategy
Sub-category: Wellbeing
Behavior:
{
    <gaze>           game-ui
    <animate>        sadness
    <animate>        slow-eye-blink
    <speech>         "Our population is not very happy."
    <glance>         subject-1
    <speech>         "This worries me."
    <glance>         subject-2
    <speech>         "What can we do?"
    <gaze>           game-ui
}

```

Listing 1. Example of a robot behavior.

summarizing the main results achieved at the end of each learning session. In the long-term study, the students' performance is especially useful to revisit specific tasks within M-Enercities that were completed or not, as well as information about how they were completed. This allows the tutor to "recall" previous sessions, highlighting learning gains and discussing specific challenges that the students went through. It also allows the tutor to provide the group with hints on how to address such challenges, thereby adapting to their learning needs.

4.3.4 Hybrid Behavior Controller. The interaction behaviors of the robot are governed by the Hybrid Behavior Controller module, whose design and implementation details can be found in Sequeira et al. (2016). The controller comprises a set of *interaction strategy rules* and a *mapping function*. The input of the controller is a set of *perceptual features*, namely, facial features, encoding the students' expressive information; auditory features, identifying the active speaker and detecting a set of keywords spoken by the students that are relevant for the learning task; and game-related features, providing information about critical moments of the game, such as when a level changes or some resource of the city becomes scarce. All features are automatically extracted and encoded from raw data captured from microphones, cameras and other sensing devices strategically positioned in the environment. The output of the controller module is a *social interaction behavior*, including all the animations, gaze functions and utterances spoken by the robot during the interaction with the students. The design of the behaviors, as discussed in Alves-Oliveira et al. (2014), was inspired by observed teacher-students interactions during the aforementioned mock-up sessions. In addition to the dialog of the robot, each interaction behavior encodes the non-verbal behavior of the robot that was also inspired in the way real teachers and the students interacted, e.g, by shifting the robot's gaze between the game and the players to drive their focus of attention toward relevant aspects of the task. An example of a full behavior definition is specified in Listing 1, designed to address a situation where the well-being of the city's population in M-Enercities was low.⁹ The Hybrid Behavior Controller is then composed of:

Interaction Strategy Rules—Correspond to manually encoded *behavior rules* in the form If-perceptual state-Then-interaction behavior. The idea is that when the features have the values as specified in the rule's If statement, the rule becomes active, which in turn

⁹The list of encoded behaviors can be retrieved from "Publications/Deliverables" sections of the project's website at <http://www.emote-project.eu/>.

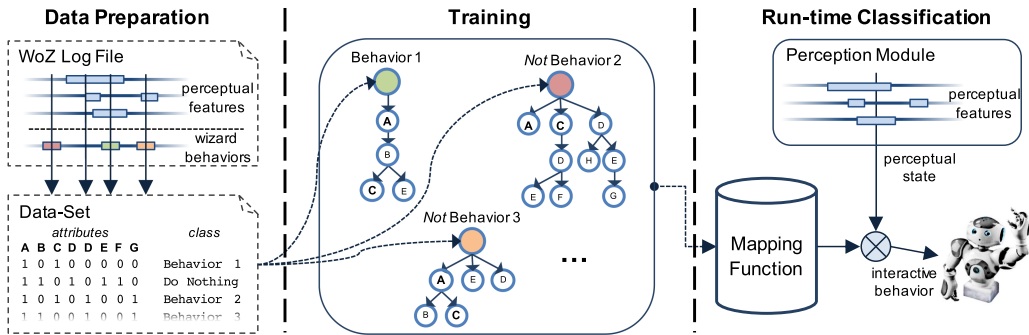


Fig. 3. A depiction of the ML procedure involved in creating the robot controller's *Mapping Function*. Adapted from Sequeira et al. (2016).

automatically triggers the associated interaction behavior of the robot. A set of rules was defined to encode domain knowledge that is relevant to improve the students' comprehension of the task and to understand their learning progress. Some rules were inspired by pedagogical strategies observed in teachers during the mock-up sessions. We also performed informal interviews with the teachers to understand their reasoning process and gather information about interaction dynamics and common strategies used during the several interaction studies. This led to the design of rules such that whenever the game starts, the robot gives a short tutorial explaining the game rules, and after the game ends a rule is triggered that "wraps-up" by summarizing the main achievements and analyzing the group's performance. Other rules encode interaction-management functions, such as announcing the next player or other game-related information. The rationale was that the behaviors in these rules occurred at well-defined moments and in a consistent manner; hence, we do not need to learn interaction strategies for such cases.

Mapping Function—To endow the robot with a more robust behavioral repertoire, the hand-designed strategy rules were complemented by interaction strategies discovered using a ML technique. In particular, we used ML to identify behavioral patterns that are less common and, therefore, harder to explicitly identify by the experts or through observation and annotation. An important aspect of the Restricted-Perception WoZ method is that the behavior data from the operator is dependent on the same perceptual features that will drive the behavior of the robot during autonomous interaction (Sequeira et al. 2016). For this reason, such data is particularly amenable to a ML analysis. We used the data from these studies to train a classifier that maps perceived situations to the robot's actions, i.e., a model of *which* behaviors should be triggered and *when* to trigger them. The procedure is illustrated in Figure 3. It starts with a *Data Preparation* phase involving the transformation of the collected demonstrations into a data-set of state features-behavior pairs, which are referred to as training instances. The *Training* phase learns a mapping function encoding the observed interaction strategies from the given data-set.¹⁰ Specifically, we used a technique based on an associative metric within frequent-pattern mining that is detailed in Sequeira and Antunes (2010) and in Sequeira et al. (2013). As illustrated in Figure 3, for each wizard behavior sampled from the log file, the corresponding

¹⁰We note that the interaction controller module is agnostic to the ML algorithm that is used to learn the Mapping Function. In that regard, standard ML classification algorithms may be suitable to learn interaction strategies based on the collected WoZ data.

“Behavior” tree is updated according to the perceptual features that were active at that time. This indicates that states where those perceptions are active are an example of *when* to execute the behavior. For all other behaviors, the corresponding “Not Behavior” trees are updated, indicating that the features are an example of *when not* to execute them. By the end of training, each “(Not) Behavior” tree stores patterns that indicate the perceptual states on which the corresponding behavior should (not) be executed. After having learned the mapping function, the system can choose an appropriate interaction behavior at run-time upon request, given the robot’s perceptual state. We note that while the Rules module covers the question regarding when and when not to execute some behavior (the rules were handcrafted to ensure this), the ML module had to be designed such that behaviors are not automatically triggered incorrectly and at the wrong times, which could potentially “break” the interaction flow between the robot and the students.

5 METHOD

This section presents the design of the method to meet the proposed goal of this work.

5.1 Hypothesis

A few points can be highlighted to frame our studies: teachers with empathy competencies are associated with positive students’ outcomes; collaborative learning environments can be more beneficial, depending on the educational topic that is being taught; social robots have been used for children in educational applications, with positive impact on students’ engagement during learning. Therefore, we have formulated the following study hypothesis:

Hypothesis 1—In a collaborative group learning environment an empathic robot improves students’ learning outcomes. We have performed a between-subjects design study in which an empathic robotic tutor interacts with groups of children in a school classroom, performing a collaborative activity about sustainable development. This was a short-term study in which groups of children performed one session with the robot, distributed randomly across one of the three study conditions: (1) two children learn with an empathic robot, (2) two children learn with a non-empathic robot, or (3) three children learn without the presence of a robot. We hypothesize that students will have higher learning achievements in the condition in which they perform the learning activity with the empathic robot.

Hypothesis 2—In a collaborative learning environment groups of children learn over time with an empathic robot. This hypothesis concerns a deeper understanding about empathy in robots, as it concerns a long-term study. Thus, we have performed a within-subjects design study in which groups of students interacted with the empathic robot over a period of two months (four sessions, one session every other week), and have evaluated their learning outcomes. The learning content in our research is related with sustainable development curricula, a domain of knowledge that requires group discussions and understanding of others people’s opinions and perspectives to make sustainable decisions. We hypothesize that learning gains will increase over time.

5.2 Ethical Considerations

We developed a robotic tutor that forms a social bond with lower-secondary students to promote learning in a personalized way. As Fridin (2014) and Serholt et al. (2017) describe, this entails ethical concerns specially related to long-term interactions. These ethical concerns include attachment to the robot, deception about the robot’s abilities, and robot autonomy and authority. Regarding the attachment to the robot, it was explained to all children that participated in the study exactly how long the robotic tutor will be present in their school and when it will be removed, similar to

Table 2. Learning Goals Supported by the Robotic Tutor and by the M-Enercities Game, Matched with the Measurement Media Used to Evaluate Sustainable Development Learning Outcomes

Learning goals in sustainable education	Measurement media	Section
Factual knowledge	Multiple-choice questionnaire	Section 5.3.1
Trade-offs and multiple perspectives	Writing assessments: (1) trade-offs were measured according to the number of options considered to solve a sustainable problem; (2) multiple perspectives were measured according to the number of arguments.	Section 5.3.2
Personal values	Behavioral analysis about: (1) Scores comments, (2) In-depth discussions, (3) Meaningful conversations	Section 5.3.3

introducing a temporary teacher. We have explained to children the robot's workings and answered any questions about it to avoid deception over the robot's abilities. In relation to the robot's authority, as children are aware of the balance between expertise and authority (Serholt et al. 2016), we explained that while the robot is trying to help them to accomplish learning tasks it will not be responsible for grading and does not have the authority to keep them engaged in the task. Moreover, all participants provided written informed consent from their caregivers prior to participation and assented to participate in the study when asked before the starting of each session. The guidelines of the Declaration of Helsinki and the standards of the American Psychological Association were followed.

5.3 Measures

We have designed, developed, and evaluated two different assessment media to measure learning goals in sustainable development education. The two assessment media used were *writing assignments* and *behavioral analysis*. Their full description is described below and summarized in Table 2.¹¹

5.3.1 Factual Knowledge Measure. A multiple-choice questionnaire was designed as a measure of Factual Knowledge about energy sources. The questions were designed according to the knowledge available in the M-Enercities game by a researcher of the EMOTE project who was also a teacher in school. The multiple-choice questionnaire about sustainability was piloted to determine whether the difficulty level in the pre- and post-tests assignments was appropriate, which would mean no statistical difference between pre- and post-tests scores. The pilot test was performed with 48 children from grades 4 and 5 (the same age-group as the target students from our main study) and the difficulty level was evaluated based on the percentage of correct answers to each question. Results from the pilot test showed no significant difference between pre-test ($M = 5.0$, $SD = 0.16$) and the post-test ($M = 4.9$, $SD = 0.16$), $p > 0.05$, therefore showing a similar level of difficulty. Both pre- and post-tests comprised 12 items each (24 items in total).

5.3.2 Trade-offs and Multiple Perspectives Measure. To test students' ability to understand that there are many different perspectives to debate sustainable development, we created a writing exercise that reflects a sustainable problem in which different stances can be taken. Students were

¹¹All writing assignments used in the work are made available online on Deliverable 7.2 of the EMOTE project at <http://www.emote-project.eu/publications/deliverables>.

instructed to provide two types of answers: (1) chose one or more solutions to solve the problem, as a measure of Trade-offs; (2) argue for the chosen solution(s) as a measure of Multiple Perspectives. We piloted the exercises with the same 48 children. Two researchers coded the data and the reliability score (Cohen's kappa) for the number of perspectives mentioned in the argument was .86, denoting a strong agreement between coders. Results from the pilot study indicated no significant differences between the pre- ($N = 23$, $M = 0.70$, $SD = 0.93$) and post-test ($N = 25$, $M = 0.52$, $SD = 0.77$) sustainable problems, enabling their use as a measure for the study.

5.3.3 Personal Values. Learning about sustainability is not a straightforward process. According to Fior et al. (2010), sustainable development education is not primarily about changing attitudes, instead “*environmental learning in the presence of complexity, uncertainty, risk and necessity [they argue] must be accepting of multiple perspectives supportive of meta-learning across perspectives, and detached from the making of decisions in its (and learners’) own immediate context*” (Fior et al. 2010). Thus, instead of changing attitudes and behaviors, the learning measures were designed to capture children’s awareness of different perspectives and the ever-present trade-offs around sustainability. Furthermore, according to Antle et al. (2015), sustainability curricula for elementary schools “*often focus on key concepts such as balancing conservation and consumption*” (Antle et al. 2015, p. 37), while ignoring the important role that children’s personal values have in learning. They argue that sustainability education for elementary school-aged children should be modeled according to the Emergent Dialogue Model, especially in the area of digital media games such as Youtopia (Antle et al. 2013). The core idea of the Emergent Dialogue Model is that children should be invited to participate in *personally meaningful dialogues* during the game play. In the case of our study, we have used an autonomous robot that interacts with children as a way to foster meaningful dialogues about sustainable development during the game-play of the M-Enercities. Thus, Personal Values were measured by analyzing the video recording collected during the study for behavioral analysis. Using the dedicated software ELAN v4.8.1 (Wittenburg et al. 2006), each study session was coded for behavioral analysis using the video recordings of the participants while performing the collaborative group learning activity. We have based our coding scheme on the one created by Antle et al. (2013), for the analysis of the Emergent Dialogue Model. We were interested in the analysis of the verbal behavior of the participants during the learning activity to be able to gain insights into their meaningful participation in discussions of personal values as a way to measure learning outcomes. The coding scheme used was the following:

- **Scores comment**—Discussion or comments about the game scores. This category relates to children’s comments or observations of the impact of their game actions on the game scores. It is related with an increase or decrease of the scores on any game parameters. An inclusion example would be: “*We are running out of money.*”
- **In-depth discussions**—Includes events in which one or more children talk about decisions on what resources and developments to use. An in-depth event involves a sense of the world or individual values, which differs from simple preferences. It must also involve reasoning using those values, typically around trade-offs between human and natural needs. As such, statements like “*I think we should have houses not trees*” is a preference and was not coded. However, statements such as “*No, let’s build houses instead of apartments, because they use less lumber, and we can make more trees into nature reserves*” was coded as in-depth discussion, because it involves values in the context of reasoning about trade-offs related with sustainability.
- **Meaningful conversations**—Includes verbal and/or physical disagreement with another’s action(s), or utterances related to the sustainability domain. Meaningful conversations require an objection or stance on an issue, and therefore presenting available options or



Fig. 4. Two students playing the M-Enercities with our autonomous robot in a school classroom.

suggestions is not considered. Meaningful conversations may result in resolution, abandonment (unresolved) or unilateral decision-making. Inclusion example: *“I disagree, without industry you cannot progress.”*

5.4 Materials

The list of materials used in the set-up of both the short-term and long-term studies is listed below. Consider also Figure 4 for a picture of the real set-up.

- (1) NAO torso robot from Aldebaran Robotics;
- (2) Large interactive multi-touch table running the M-Enercities;
- (3) Four video cameras for a full interaction recording;
- (4) Two lavalier microphones for voice pitch recognition (no voice recognition was used; additional details can be found in Section 4.1);
- (5) Voice recorder for behavioral analysis.

6 SHORT-TERM STUDY

An experimental study was conducted to evaluate the impact of a robot with empathy competencies on students sustainability education outcomes in a collaborative group learning environment. This relates with our Hypothesis 1, detailed in Section 5.1. To achieve the proposed goal, the study was designed considering three experimental conditions:

- **Condition 1**—Two children interacted with a robotic tutor with empathy competencies while playing the M-Enercities game.
- **Condition 2**—Two children interacted with a robotic tutor without empathy competencies while playing the M-Enercities game.
- **Condition 3**—Three children played the M-Enercities game without the presence of a robotic tutor.

6.1 Empathic vs. Non-empathic Robot: Impact on Behavior

To create Conditions 1 and 2, featuring the robot with and without empathic competencies, we have made a choice about which modules should be activated/deactivated. We refer to

Table 3. Overview of Activation of All the Modules in the Empathic and Non-empathic Version of the Robotic Tutor

Module	Activated	
	Empathic	Non-Empathic
<i>Rapport</i>	Yes	Partially
<i>Game-AI</i>	Yes	Yes
<i>Emotional Climate</i>	Yes	No
<i>Past Event Memory</i>	Yes	No
<i>Recent Event Memory</i>	Yes	Yes
<i>Hybrid Behavior Controller</i>	Yes	Partially
<i>Sustainable learning dialogue</i>	Yes	Yes

Section 4.1 for the technical architecture overview of the robotic system. Table 3 details which modules from the overall system architecture, depicted in Figure 2, are fully or partially (de)activated in each version of the robotic tutor. Despite some modules being deactivated or partially activated when comparing the versions of the robot, it is crucial to note that during the learning activity, the percentage of interventions by the robot toward the students was similar. This was ensured as the behaviors of the robot were designed and developed to be balanced between conditions. In practice, this means that the robot talks and gestures the same amount of time in both empathic and non-empathic versions. Additionally, basic idle behavior, animations, and speech capabilities remain intact; the non-empathic robot will, however, appear less aware of the students.

As we can see from Table 3, the *Past Event Memory module* is deactivated in the non-empathic condition, which means that the robot is unable to recall previous learning sessions and summarize activities that occurred therein. The rationale is that using memory of other people's past experiences is a way to simulate how they feel in situations similar to those that they are currently facing, which consequently leads to empathic behavior (Ciamelli et al. 2013). Notwithstanding, the robot has the *Recent Event Memory module* activated in both conditions, which means it remembers the performance of students *during* each learning session in both versions.

The *Emotional Climate module* is also deactivated in the non-empathic condition. This module is responsible for perceiving the emotional state of the group of students during the learning activity, an inherent perception for performing empathic behaviors. With this module deactivated in the non-empathic condition, the robot provides more generic suggestions for students during the game that are not related with emotional perceptions. Nonetheless, the *Sustainable Learning Dialogue module* is activated in both empathic and non-empathic conditions and is responsible for ensuring that the robot performs similar dialogues about sustainable learning in both of the study conditions.

The *Rapport module* is partially deactivated in the non-empathic condition, meaning that some contingent behaviors are still activated, but not all of them. An example that can illustrate the impact on the behavior of the robot concerns gaze, an important social signal (Emery 2000). In the empathic condition, the robot moves his eye-gaze to the student currently speaking by accurately locating the student using the sound coming from the microphone (each student is using a microphone, as can be seen in Figure 4, for the robot to accurately follow the student, however, no speech recognition system is used). In the non-empathic condition, the robot still looks at the student who is speaking but uses predefined coordinates of the students' locations in front of the table. This means that subtle changes in the students' locations, especially of their faces, are not tracked by the robot, resulting in a less context-aware and contingency behavior toward the students. However, it is important to note that this does not translate a random gaze behavior (Park

et al. 2017a) but a less precise gaze orientation that does not lead to drastic effects on the perception of the robot between conditions. Another feature of the Rapport module that is deactivated in the non-empathic condition concerns the voice volume and pitch of the robot not being adjusted to the perceived volume and pitch of the students' voices, contrarily to what occurs in the empathic condition. As these adjustments are related with empathy behaviors (Imel et al. 2014), the characteristics of the robot's voice are kept constant in the non-empathic version.

As for the *Hybrid Behavior Controller module*, only the Interaction Strategy Rules are active for the non-empathic version, which means that there are no behaviors being triggered by the Mapping Function. Although this part of the controller does not necessarily lead to empathic behavior, we note that it is the result of applying a ML algorithm aimed at discovering more subtle interaction strategies used by the wizard in the Restricted-Perception WoZ, which cannot be hand-crafted and put in the Interaction Strategy Rules list. Notwithstanding, this does not mean that the robot will intervene inappropriately and/or at the wrong times, as all behavior is still controlled by manually encoded rules. Thus, the robot was kept as a knowledgeable and informative interlocutor in all study conditions, as previous studies have shown that children can easily distinguish between reliable robots as information sources (Breazeal et al. 2016).

Overall, the deactivated modules concern perceptions of the cognitive and emotional states of the students. Notwithstanding, the social and pedagogical behaviors are similar in both study conditions, with the robot having similar frequency of interventions. This ensured the social and intelligent tutoring autonomous behavior of the robot.

6.2 Sample

A total of 63 children ($M = 13.74$, $SD = 0.73$, 25 female) participated in this study. Participants were grouped by their school teachers according to groups of students they knew worked well together in a learning context, and were randomly assigned across study conditions. Therefore, 22 participants interacted with the empathic robotic tutor, in a total of 11 learning sessions consisting each of 2 children and 1 robot; 20 participants interacted with a non-empathic robotic tutor, in a total of 10 learning sessions consisting each of 2 children and 1 robot; 21 participants were allocated in the condition with no robot, in a total of 7 learning sessions consisting of 3 children. Two researchers were responsible for the study in the school: a psychologist, who interacted with the participants and acted as a leading researcher, and a computer scientist, who was responsible for the technical equipment.

6.3 Procedure

Each group of students arrived to the designated classroom where the study took place. The leading researcher provided an explanation about the study they would undergo. Participants were invited to fill-in the pre-tests about sustainability explained in Section 5.3. After completing the pre-tests, participants were introduced to the robotic tutor (in case of conditions 1 and 2) and to the M-Energities game, while they performed an initial trial round of the game to have a hands-on experience with the activity. When the trial round of the game was over, the researcher left the room, leaving the participants performing the learning activity with the robot (in case of conditions 1 and 2) or by themselves (in case of condition 3). Although the students were left alone in the classroom performing the learning activity, they had permanent indirect supervision by both researchers. This was ensured, since the classroom had a large window to an external room, which allowed to monitor the progress and at the same time providing privacy to their learning process. Furthermore, this set-up enabled participants to reach out to the researcher to ask for help, e.g., if technical problems with the learning activity or with the robot occurred. Finally, when the learning activity was over, the researcher entered the room and closed the learning activity application,

thus ending the activity. Participants were able to say goodbye to the robot and afterwards were invited to fill-in the post-tests about sustainability knowledge. At the end, some time was given to discuss their experience during the study, providing an open space for children to ask questions or share thoughts. Each session had a duration of 1h, in which 30min were allocated to the activity of playing the M-Enercities game, and the remaining 30min were dedicated to pre- and post-tests.

6.4 Results

We present the learning gains for the different measures used about sustainable education.

6.4.1 Factual Knowledge. We compared the results from the pre- and post-tests across the three study conditions to compare the learning outcomes in the participants' factual knowledge about sustainable learning. According to the Mixed-ANOVA test, there was no significant difference between the study conditions for learning outcomes on factual knowledge about sustainable education, $F(2, 59) = 0.586, p = 0.560$. The means for the pre-test result were the following: $M = 9.29, SD = 1.42$; $M = 9.20, SD = 1.06$; $M = 9.33, SD = 1.02$, corresponding to the interaction with the empathic robot, non-empathic robot, and no-robot conditions, respectively. While the means for the post-tests were the following: $M = 7.90, SD = 1.67$; $M = 8.30, SD = 1.46$; $M = 8.38, SD = 1.24$, corresponding to the interaction with the empathic robot, non-empathic robot, and no-robot conditions, respectively.

6.4.2 Trade-offs and Multiple Perspectives. The assessment of participants' understanding of trade-offs and perspectives was performed, as illustrated in Table 2.

Trade-offs (or Number of Solutions). We ran a Mixed-ANOVA test to analyze if there were differences in the number of solutions chosen by the participants across the three study conditions to solve the sustainable exercise problems. We took into account their performance in the pre-test compared to their performance in the post-test and the condition they were allocated in. Results showed no significant differences across conditions when comparing the results from the pre- and post-tests, $F(2, 60) = 1.726, p = 0.187$.

Multiple Perspectives (or Number of Arguments). The number of arguments mentioned by the participants to justify their solutions was the measure for the multiple perspectives in solving a sustainable problem. Results show that the number of arguments did not change significantly, $F(2, 493) = 0.925, p = 0.402$, when participants learned with the empathic robotic tutor compared to the other study conditions ($M = 1.05, SD = 0.19$; $M = 1.00, SD = 0.20$, for pre- and post-tests, respectively).

6.4.3 Personal Values. We performed verbal behavior analysis across the 3 study conditions to measure personal values, considering *scores comment*, *in-depth discussions*, and *meaningful conversations* (see coding scheme in Section 5.3.3 for more details). When running a Chi-squared test, we can see that there is a statistically significant association between the conditions of the study and how children exchanged personal values about sustainability, $\chi^2(4) = 9.375, p = 0.05$, with the strength of the relationship with Cramér's V being $\phi_c = 0.109$ revealing a medium effect. We performed a post-hoc test to analyze contingency tables to understand in which study conditions personal values exchanges were statistically significant (Beasley and Schumacker 1995).

Scores Comment. We found that in the empathic robotic tutor condition participants commented on scores statistically less comparing to the other study conditions, $p = 0.01$. Thus, participants commented less on the scores when in the condition with the empathic robotic tutor (64.52%), compared to the condition with the non-empathic robotic tutor (77.97%) and with the no-robot

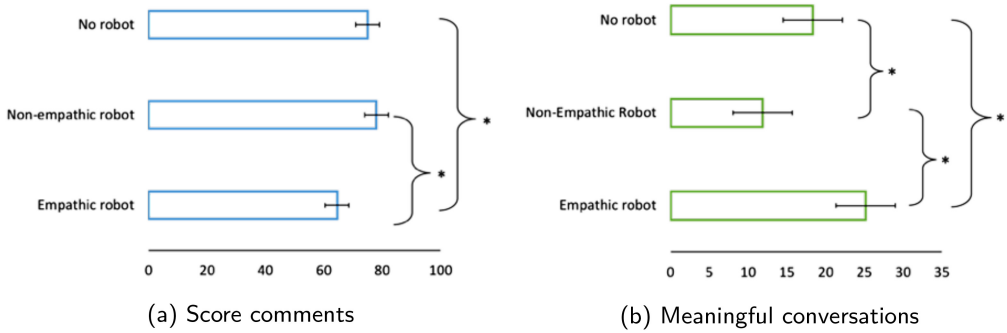


Fig. 5. Results for comments and meaningful conversations about sustainable development during the short-term learning activity across the three study conditions. Results are presented in frequencies and significant results are represented with the symbol *, $p < 0.05$.

condition (75.00%), as illustrated in Figure 5(a). The abovementioned results are significant for $p = 0.016$, according to the procedure of residual analysis.

In-depth Discussions. No significant result was found for in-depth discussions between study conditions, $p > 0.05$.

Meaningful Conversations. By using the adjusted standardized residuals method of analysis (Garcia-Perez and Nunez-Anton 2003), we discovered significant differences in meaningful conversations between study conditions. Therefore, when children learned with the empathic robotic tutor more meaningful conversations emerged (25.16%, $p = 0.013$), followed by the no-robot condition (18.33%, $p = 0.012$), and the least meaningful discussion occurred when children learned with the non-empathic robotic tutor (11.86%, $p = 0.016$) (see Figure 5(b)).

6.5 Discussion

This section discusses the results regarding the learning gains across the three study conditions for this short-term study in a school classroom.

Children learned to have meaningful conversations about sustainability and worried less about scores when learning with an empathic robot.

Because sustainable development education is about engaging in deep discussions that consider the existence of “*complexity, uncertainty, risk, and necessity*” to solve sustainability-related problems (Fior et al. 2010), the results showed that this was successfully accomplished when children learned with a robot with empathy competencies. However, when children interacted with a robot without empathy or without a robot at all, they seemed to be more concerned about passing levels (the traditional way of playing any game), instead of engaging in dialogue about sustainable education. We emphasize that the tutoring behaviors for the empathic robot vs. the non-empathic conditions were the same, i.e., the sustainable learning dialogue that the robots had was similar in both study conditions. This makes the result particularly important, since it shows that empathic competencies in a robot impacted the way children engaged in the learning process, partially supporting our first study hypothesis. In fact, engaging in meaningful conversations implies that children share their personal values by objecting or taking a stance that generates discussion about sustainability while playing the game. The empathic robotic tutor fostered and motivated the children to engage in this type of dialogue as a way to increase their knowledge about sustainable education. Future studies should include more in depth analysis of educational interactions,

specifically related with the emergence of educational dialogues (e.g., if it is facilitated by the robot or instigated by children themselves).

No other impact on sustainable development learning was found.

Results from the multiple-choice questionnaire on factual knowledge about sustainability did not present significant results. Additionally, the writing exercise in which children were invited to choose solutions to solve a problem related to sustainable development and to argue about their options for solving it, also did not show significant results. The results might be a reflection of the short-term interaction with the robot. In fact, learning takes time (Fisher et al. 1981), especially when we consider sustainable education, a challenging curricular topic to teach that is complex to learn (Moore 2005). The short-term interaction that students had with the M-Enercities game, which was their learning environment, can also help explain the lack of learning gains. Indeed, M-Enercities enables students to explore the virtual world of the game in an unrestricted way. By providing freedom to open game menus that seem interesting to them according to the game action they want to perform, this can result in students not opening all of the game menus; thus, they may be exposed to only part of the overall knowledge that the game can offer. Therefore, due to the short-term nature of the interaction with the learning environment and with the robotic tutor, children would benefit for more extended period of interaction to be exposed to more learning content. This aspect will be explored in the long-term study described in Section 7.

7 LONG-TERM STUDY

A descriptive long-term study was conducted to investigate the learning outcomes of students that learned in groups with an empathic robotic tutor over an extended period of time in school. To achieve our goal, we have deployed a robot with empathy capabilities for 2 consecutive months in a school setting (four sessions, one session every other week), to teach small groups of students about sustainable education using M-Enercities as the collaborative learning environment. To sustain the achievements of children within these weeks, the robotic tutor would recall leanings and difficulties of previous sessions upon starting each learning session, thus ensuring reflection over previous acquisitions. This study related wit hypothesis number 2, explained in Section 5.1.

7.1 Sample

A total of 20 children ($M = 13.78$, $SD = 0.65$, 9 female) participated in the study. Due to technical problems, one session was excluded and the final sample resulted in 18 children. The results presented for this study exclude the session with technical problems.

7.2 Procedure

Although this was a different study, the procedure is similar to the one present in Section 6.3. We present in this section the variations in the procedure. Thus, participants filled-in tests about sustainable education at three time periods: (1) *baseline*, to measure their initial knowledge on sustainable development before interacting with the robot and with the learning environment; (2) at the *end of the first collaborative session*, to measure learning achievements after one interaction with the robotic tutor, which is typical for many studies in the HRI field; (3) at the *end of the 2-month period*, to be able to compare learning achievements and understand the learning curve after a long-term interaction with the empathic robotic tutor. Learning sessions were performed once every other week, therefore, two sessions per month in a total of four sessions in two consecutive months. Each session lasted about 30min with the first and last sessions taking longer as the assessments of sustainable education were applied in these sessions. The session dynamics

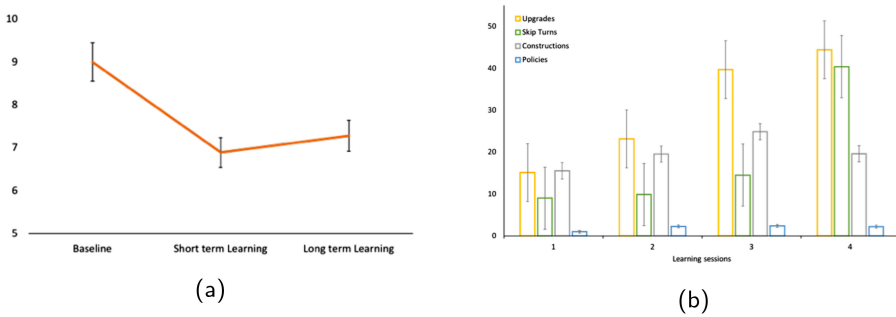


Fig. 6. Results of the long-term study: (a) factual knowledge achievements; (b) actions performed by students in the learning environment across the four learning sessions with the empathic robotic tutor. Results revealed to be significant for the upgrades and skip turns' actions.

were organized with the school teachers to minimize disturbances in the usual daily activities that children are involved in while in school.

7.3 Results

In this Section, we present the results for the long-term study with a robot with empathy in a classroom environment in school.

7.3.1 Factual Knowledge. The factual knowledge learning about sustainability was analyzed using the Friedman's test, and students' achievements in sustainability education showed a statistically significant difference over time, $\chi^2(2) = 15.464$, $p < 0.001$ with a Kendall W of .43 indicating a moderate effect (Tomczak and Tomczak 2014). Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, revealing a significance difference when comparing the baseline ($M = 9.00$, $SD = 1.41$) to the short-term learning results ($M = 6.89$, $SD = 1.28$), $Z = -3.344$, $p = 0.001$, $r = -0.56$; and when comparing the baseline with the long-term learning results ($M = 7.28$, $SD = 1.23$), $Z = -3.084$, $p = 0.002$, $r = -0.51$. No other statistical significant result was found, $p > 0.05$. From Figure 6(a), we can see that participants' knowledge about sustainability topics started high, and after interacting with the robot in the pedagogical activity it decreased, albeit with a slightly increase in the long-term, possibly showing a tendency to return to the baseline. This result possibly translates normative results in children's learning, in which they question previously accommodated knowledge about the topics.

7.3.2 Trade-offs and Multiple Perspectives. Similar to the analysis performed for the short-term evaluation, we evaluated the trade-offs and multiple perspectives about sustainability according to the *number of solutions* proposed to solve a given sustainable dilemma, and the *number of arguments* considered to justify their solutions. We explain the results below.

Trade-offs (or Number of Solutions). We analyzed the number of solutions that participants considered possible to solve the given environmental problem. Participants considered more options over time, however this increase was not significant, $\chi^2(2) = 1.125$, $p > 0.05$.

Multiple Perspectives (or Number of Arguments). Although there was a slight increase in the number of perspectives considered by participants, Friedman's test showed that this was not statistically significant, $\chi^2(2) = 3.375$, $p > 0.05$.

7.3.3 Personal Values. A Repeated Measures ANOVA with a Greenhouse-Geisser correction determined that mean personal values did not differ significantly between the first and the last

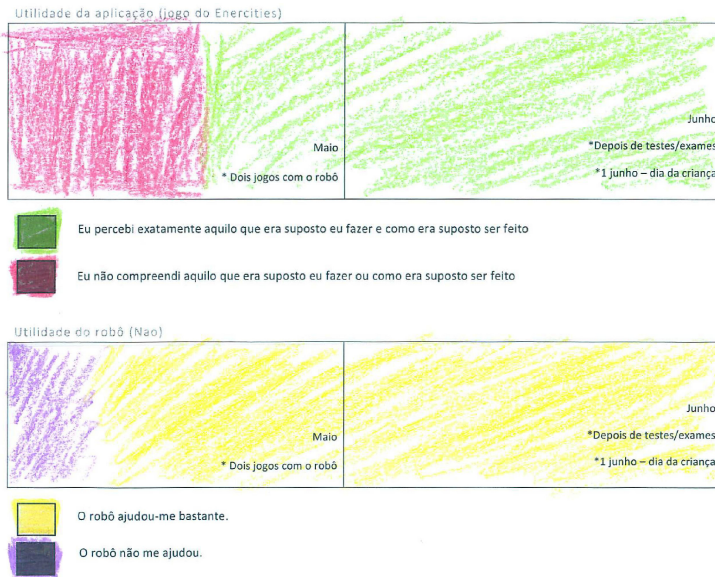


Fig. 7. Example of a MemoLine filled out by one of the children and adapted for the context of the present study. Children were asked to use pencils with different colors to express how hard/easy it was for them to (1) play the game (in which red signals “I did not understand what I should do in the game” and green signals “I understood what I should do in the game”), (2) and how much the robotic tutor was useful for them in understanding how to play the game (purple signals “The robotic tutor did not help me understand the game” and yellow signals “The robotic tutor helped me understand how to play the game”). The MemoLine should be read in timeline manner. Therefore, the line in the center signals the middle point of the sessions to situate children in their assessment. The left side of the line concerns Sessions 1 and 2 and the left side concerns the Sessions 3 and 4.

learning session with the empathic robotic tutor, $F(1, 1.756) = 3.530, p = 0.061$. Since personal values were measured using behavioral analysis of the 4 learning sessions, we have no baseline result (baseline considers assessments conducted prior to the starting of the intervention).

7.3.4 Game-play. The behavior (or actions) of the participants during game-play also served as a measure for analyzing sustainability learning. By using the game logs, we were able to extract what actions children performed during the game, representing where they invested and the dynamics of their game-play. The exact McNemar’s test showed that participants’ actions related with applying policies, $p = 1.00$, and performing constructions, $p = 0.125$, did not differ statistically between the first and the last session. However, participants performed statistically significantly more constructions’ upgrades from the first to the last session, $p < 0.001$, as they also skipped more turns (representing passage of time in the game as a way to enjoy from the city resources), $p < 0.001$ (see Figure 6(b)). This means that the proportion of upgrades and skipped turns was higher in the the last session compared with the first learning session.

To better understand the changes in game-play behavior, we analyzed how helpful the robot was toward a better understanding of how to play M-Enercities. For this, we have used MemoLine, a retrospect evaluation method dedicated to children that asks them to recall previous experiences with a given product or application (Vissers et al. 2013). See Figure 7 for an example of a MemoLine filled in by a participant. A MemoLine was given to all participants at the end of the long-term study. According to its developers, MemoLine should be evaluated by making comparisons of the

areas colored by children in a timeline perspective, to extract patterns of use. Our results indicated consistency in responses, as most MemoLines showed that children were slightly confused by the game and did not find the robotic tutor particularly helpful during the initial session. However, as time passed, they have perceived more help from the robotic tutor, and could better understand the game itself revealing a mastery in game-play. This result is representative of the overall sample of the study, since 18 of 20 children filled the MemoLine in the same way. Taken together, participants' change in the behaviors of the game seem to be in line with their better understanding of the game dynamics, an understanding that was guided by the interaction with the robotic tutor. More importantly, understanding the game requires mastering concepts of sustainable education, which can explain the changes in game-play.

7.4 Discussion

This Section presets the discussion of the learning gains across a long-term study of four sessions distributed over a period of 2 consecutive months in school.

No learning gains were found after children interacted with the empathic robotic tutor for long periods of time.

The overall results from the long-term study seems to show that although some variations in the learning outcomes of students occurred, those differences were not deemed significant after repeated educational interactions with the empathic robotic tutor. Although research about null results in learning gains with technology use are still scarce, a meta-analysis conducted by Wouters et al. (2013) studied factors that impacted learning gains when students used serious games for learning. The authors concluded that more learning occurred if: (1) learning sessions were complemented with other instruction methods, (2) when multiple training sessions were involved, (3) and if learned worked in groups (Wouters et al. 2013). Although our study involved multiple learning sessions and occurred in a group educational context, the learning sessions with the robotic tutor were not supplemented with other instruction methods. This deserves further investigation to understand the role of robots in learning contexts and their long-term effects on students' learning gains.

Similarly, Hew and Brush (2007) discussed factors related with successful technology integration in schools, such as the reconsideration of assessment metrics. The authors refer that *"because curriculum and assessment are closely intertwined, there is a need to either completely reconsider the assessment approaches when technology is integrated into the school curriculum, or consider more carefully how the use of technology can meet the demands of standards-based accountability."* This suggests that metrics used to measure students' learning gains that account for technology inclusion (e.g., robots) require further investigation to meet curricular goals. In the case of our study, despite the support provided to students, both from the robot and stimulated by the M-Enercities, it did not impacted in a measurable way students' learning in sustainability education present in formal school curriculum.

Another explanation for the lack of learning progress is provided by Gerber et al. (2001) and concerns the lack of clearly defined roles of educational aids that can hinder learning gains due to an undefined presence in school. A result presented in a previous publication related with our learning scenario, showed that students assigned the role of a classmate to the robot although being explicitly instructed they would be learning with a robotic tutor (Alves-Oliveira et al. 2016). Therefore, we can be a facing a need to develop precise design guidelines to develop specific roles for robots in the educational sector whose goal is to increase learning outcomes.

In addition to this, more research is needed when designing and including robots for education. For example, studies have been showing that the mere physical presence of a robot can lead

to learning gains (Kennedy et al. 2015b; Leyzberg et al. 2012); however, the variables that affect learning gains are not established yet. Furthermore, the design of robots for education should be tailored to the time required to learn certain curricular concepts, and we hypothesize that sustainable learning can be a case in which more learning time is required. The null results from our long-term study seem to indicate that long-term HRI installations for education require more investigation and may even require a change in the interaction design between the robot and children. Our work introduced this discussion highlighting the need for a better understanding of long-term deployment of social robots amongst the educational sector.

Game-play behavior changes over time due to a better understating of the game guided by the robotic tutor.

The way children played the game about sustainable education changed over time and this change seems to be related with the interaction with the robotic tutor. Indeed, there was a statistically significant result found in the game-play behavior of children during the long-term study. Children's actions in the game showed that, over time, they have performed more upgrades to their city they have performed more skip turns (representing the passage of time in the game, such as days passing by). By performing upgrades, the city can become more sustainable, and by skipping some turns the players allow the structures and upgrades they have chosen to implement, to get their full effect before advancing to the next level. This behaviour thus seems to indicate a more thoughtful design of the city, which matches with Antle's design principles for sustainability games (Antle et al. 2015). These changes in game play are not trivial, as children needed to move away from the traditional competitiveness of passing levels (a so-called traditional mind set of a game-play), to become concerned about spending less money and taking more advantages of the resources from previous constructions. This seems to show that the change in game-play behavior goes hand in hand with the perceived easiness to understand and play the the game, guided by the interactions with the robotic tutor.

8 GENERAL DISCUSSION AND CONCLUSIONS

In this article, we presented a novel educational scenario for social robots, in which a group of children interacts with an autonomous robot in a serious collaborative game. The goal of the interaction was to infer learning outcomes with regards to environmental sustainability and the associated trade-offs involved when designing a city. We conducted a short-term study that addressed the effects of the empathic robot and a long-term study that addressed the long-term deployment of the robot in school. The short-term study compared three conditions: empathic robot, non-empathic robot, and no robot. The results showed no significant results in the majority of learning outcomes; however, there was an increase in meaningful conversations with the empathic robotic tutor, which is a stated goal for collaborative learning scenarios targeting sustainability. During the long-term study, the empathic robot was deployed for 2 months in school. Results showed no significant change in learning gains over time. Additionally, a change in game-play behaviors related to the game was observed, in which children perform more game actions toward sustainability over the sessions. The lack in learning progress may be due to several aspects, such as the quality of the interaction, the role of robots in school, and group dynamics. This reflects the need to conduct additional research in group interactions between humans and robots for educational purposes.

Summarizing, the highlights of our research are:

- (1) We designed, developed, and evaluated a robot tutor for education that autonomously interacted with students in a real-world environment of a school for 2 months.

- (2) We contribute to the field of group interaction studies between humans and robots by framing the educational context as a collaborative group learning scenario.
- (3) We concluded that an autonomous robot with empathy (compared with a robot without empathy or learning without a robot) fosters meaningful discussions about sustainability between students, which is a learning outcome in sustainability education.
- (4) We concluded that long-term educational interactions with an empathic robot did not impacted in a significant way learning gains, although there was a change in game-actions to achieve more sustainability during game-play.

8.1 Limitations and Future Work

Empathy is a complex construct that is highly dependent on the content of what people say to each other. Our empathic robot was able to perform contingent behaviors that translate empathy by receiving limited input from the children but had no access to their verbal discussions. For effective empathic robotic tutors to operate in group learning environments, they should be able to understand what children say and accordingly personalize their empathic responses. To this end, developments in speech recognition for children are needed to build better empathic interactions.

We also did not observe as many learning gains as expected, highlighting the importance of conducting more research in collaborative group learning environments with robots. Aspects such as the time duration for a learning gain to occur need to be considered when deploying a robot in a school setting. Additional qualitative research is also needed to understand the factors that favor learning gains and the factors that can hinder it.

Regarding the wider use of robotic tutors in learning, there is already a large body of research investigating robots for second language acquisition, handwriting skills, and even the understanding of other complex curricular topics, such as chemistry and wind formations. Although this reflects positive and promising directions for HRI for education, more educational scenarios need to be considered to understand how robots can be used for best impact on learning. This work has provided a corpus of reflection for future research, leading to questions, such as “Which variables lead to learning gains when using a robot for collaborative group education?” and “What is the role that a robot can have in school that fosters learning gains?” With this work, we have begun to explore the potential for robots in group learning, bringing attention to empathy as an important competence for a robot to have when interacting with students.

ACKNOWLEDGMENTS

We show our gratitude to the teachers, students, and school-staff from Escola Quinta do Marquês (Oeiras, Portugal) for their involvement in the studies. We also thank Daniel Silva for collaborating in the coding of the verbal behavioral. The authors are solely responsible for the content of this publication. It does not represent the opinion of the European Commission (EC), and the EC is not responsible for any use that might be made of data appearing therein.

REFERENCES

- Patricia Alves-Oliveira, Srinivasan Janarthanam, Ana Candeias, Amol Deshmukh, Tiago Ribeiro, Helen Hastie, Ana Paiva, and Ruth Aylett. 2014. Toward dialogue dimensions for a robotic tutor in collaborative learning scenarios. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 862–867.
- Patricia Alves-Oliveira, Pedro Sequeira, and Ana Paiva. 2016. The role that an educational robot plays. In *Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 817–822. DOI:<https://doi.org/10.1109/roman.2016.7745213>
- Patricia Alves-Oliveira, Pedro Sequeira, Eugenio Di Tullio, Sofia Petisca, Carla Guerra, Francisco S. Melo, and Ana Paiva. 2015. “It’s amazing, we are all feeling it!”—Emotional climate as a group-level emotional expression in HRI. *Proceedings of the Artificial Intelligence and Human-Robot Interaction AAAI Fall Symposium Series*.

- Alissa N. Antle, Jillian L. Warren, Aaron May, Min Fan, and Alyssa F. Wise. 2015. Emergent dialogue: Eliciting values during children's collaboration with a tabletop game for change. In *Proceedings of the 2014 Conference on Interaction Design and Children*. ACM, 37–46. DOI: <https://doi.org/10.1145/2593968.2593971>
- Alissa N. Antle, Alyssa F. Wise, Amanda Hall, Saba Nowroozi, Perry Tan, Jillian Warren, Rachael Eckersley, and Michelle Fan. 2013. Youtopia: A collaborative, tangible, multi-touch, sustainability learning activity. In *Proceedings of the 12th International Conference on Interaction Design and Children*. ACM, 565–568. DOI: <https://doi.org/10.1145/2485760.2485866>
- Paul Baxter, Emily Ashurst, Robin Read, James Kennedy, and Tony Belpaeme. 2017. Robot education peers in a situated primary school study: Personalisation promotes child learning. *PLoS One* 12, 5 (2017), e0178126. DOI: <https://doi.org/10.1371/journal.pone.0178126>
- T. Mark Beasley and Randall E. Schumacker. 1995. Multiple regression approach to analyzing contingency tables: Post hoc and planned comparison procedures. *J. Exper. Edu.* 64, 1 (1995), 79–93. DOI: <https://doi.org/10.1080/00220973.1995.9943797>
- Tony Belpaeme, James Kennedy, Aditi Ramachandran, Brian Scassellati, and Fumihide Tanaka. 2018a. Social robots for education: A review. *Sci. Robot.* 3, 21 (2018), eaat5954. DOI: <https://doi.org/10.1126/scirobotics.aat5954>
- Tony Belpaeme, Paul Vogt, Rianne Van den Berghe, Kirsten Bergmann, Tilbe Gökşun, Mirjam De Haas, Junko Kanero, James Kennedy, Aylin C Küntay, Ora Oudgenoeg-Paz, Fotios Papadopoulos, Thorsten Schodde, Josje Verhagen, Christopher D. Wallbridge, Bram Willemsen, Jan de Wit, Vasfiye Geçkin, Laura Hoffmann, Stefan Kopp, Emiel Kraemer, Ezgi Mamus, Jean-Marc Montanier, Cansu Oranç, and Amit Kumar Pandey. 2018b. Guidelines for designing social robots as second language tutors. *Int. J. Soc. Robot.* (2018), 1–17. DOI: <https://doi.org/10.1007/s12369-018-0467-6>
- Phyllis C. Blumenfeld, Ronald W. Marx, Elliot Soloway, and Joseph Krajcik. 1996. Learning with peers: From small group cooperation to collaborative communities. *Edu. Res.* 25, 8 (1996), 37–39. DOI: <https://doi.org/10.3102/0013189X025008037>
- Cynthia Breazeal, Paul L. Harris, David DeSteno, Jacqueline M. Kory Westlund, Leah Dickens, and Sooyeon Jeong. 2016. Young children treat robots as informants. *Topics Cogn. Sci.* 8, 2 (2016), 481–491. DOI: <https://doi.org/10.1111/tops.12192>
- Elizabeth Broadbent, Danielle Alexis Feerst, Seung Ho Lee, Hayley Robinson, Jordi Albo-Canals, Ho Seok Ahn, and Bruce A. MacDonald. 2018. How could companion robots be useful in rural schools? *Int. J. Soc. Robot.* 10, 3 (2018), 295–307. DOI: <https://doi.org/doi.org/10.1007/s12369-017-0460-5>
- Shruti Chandra, Patrícia Alves-Oliveira, Séverin Lemaignan, Pedro Sequeira, Ana Paiva, and Pierre Dillenbourg. 2015. Can a child feel responsible for another in the presence of a robot in a collaborative learning activity? In *Proceedings of the 24th IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 167–172. DOI: <https://doi.org/10.1109/ROMAN.2015.7333678>
- Chih-Wei Chang, Jih-Hsien Lee, Chao Po-Yao, Wang Chin-Yeh, and Chen Gwo-Dong. 2010. Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *J. Edu. Technol. Soc.* 13–24, 2 (2010), 13.
- Catherine C. Chase, Doris B. Chin, Marilyn A. Oppezzo, and Daniel L. Schwartz. 2009. Teachable agents and the protégé effect: Increasing the effort toward learning. *J. Sci. Edu. Technol.* 18, 4 (2009), 334–352. DOI: <https://doi.org/10.1007/s10956-009-9180-4>
- Elisa Ciaramelli, Francesco Bernardi, and Morris Moscovitch. 2013. Individualized theory of mind (iToM): When memory modulates empathy. *Front. Psychol.* 4 (2013), 1–18. DOI: <https://doi.org/10.3389/fpsyg.2013.00004>
- Jeffrey Cornelius-White. 2007. Learner-centered teacher-student relationships are effective: A meta-analysis. *Rev. Edu. Res.* 77, 1 (2007), 113–143. DOI: <https://doi.org/10.3102/003465430298563>
- Filipa Correia, Sofia Petisca, Patrícia Alves-Oliveira, Tiago Ribeiro, Francisco S. Melo, and Ana Paiva. 2018. “I choose... YOU!” Membership preferences in human-robot teams. *Auton. Robots* (2018), 1–15. DOI: <https://doi.org/10.1007/s10514-018-9767-9>
- Henriette Cramer, Jorrit Goddijn, Bob Wielinga, and Vanessa Evers. 2010. In *Proceedings of the 5th ACM/IEEE International Conference on Human-Robot Interaction*. IEEE, 141–142. DOI: <https://doi.org/10.1109/HRI.2010.5453224>
- Mark H Davis. 2018. *Empathy: A Social Psychological Approach*. Routledge.
- Alessandro De Gloria, Francesco Bellotti, and Riccardo Berta. 2014. Serious games for education and training. *Int. J. Serious Games* 1, 1 (Feb 2014). DOI: <https://doi.org/10.17083/ijsg.v1i1.11>
- Pierre Dillenbourg. 1999. What do you mean by collaborative learning?. In *Collaborative-Learning: Cognitive and Computational Approaches*. Elsevier, Oxford, 1–19.
- Han Du and Lijuan Wang. 2016. The impact of the number of dyads on estimation of dyadic data analysis using multilevel modeling. *Methodology* 12 (2016), 21–31. DOI: <https://doi.org/10.1027/1614-2241/a000105>
- Nathan J. Emery. 2000. The eyes have it: The neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev.* 24, 6 (2000), 581–604. DOI: [https://doi.org/10.1016/S0149-7634\(00\)00025-7](https://doi.org/10.1016/S0149-7634(00)00025-7)
- Norma Deitch Feshbach and Seymour Feshbach. 2009. Empathy and education. In *The Social Neuroscience of Empathy*, J. Decety and W. Ickes (Eds.). MIT Press, Cambridge, 85–98.

- Meghann Fior, Alejandro Ramirez-Serrano, Tanya Beran, Sarah Nugent, and Roman Kuzyk. 2010. Children's relationships with robots: Robot is child's new friend. *J. Phys. Agents* 4, 3 (2010), 9–17. DOI : <https://doi.org/10.14198/JoPha.2010.4.3.02>
- Charles W. Fisher, David C. Berliner, Nikola N. Filby, Richard Marliave, Leonard S. Cahen, and Marilyn M. Dishaw. 1981. Teaching behaviors, academic learning time, and student achievement: An overview. *J. Class. Interact.* 17, 1 (1981), 2–15.
- Stanley Frager and Carolyn Stern. 1970. Learning by teaching. *Read. Teach.* 23, 5 (1970), 403–417.
- Marlena R. Fraune, Satoru Kawakami, Selma Sabanovic, P. Ravindra S. De Silva, and Michio Okada. 2015a. Three's company, or a crowd?: The effects of robot number and behavior on HRI in Japan and the USA. In *Proceedings of the International Conference on Robotics: Science and Systems*. DOI : <https://doi.org/10.15607/RSS.2015.XI.033>
- Marlena R. Fraune, Yusaku Nishiwaki, Selma Sabanović, Eliot R. Smith, and Michio Okada. 2017. Threatening flocks and mindful snowflakes: How group entitativity affects perceptions of robots. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 205–213. DOI : <https://doi.org/10.1145/2909824.3020248>
- Marlena R. Fraune, Steven Sherrin, Selma Sabanović, and Eliot R. Smith. 2015b. Rabble of robots effects: Number and type of robots modulates attitudes, emotions, and stereotypes. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 109–116. DOI : <https://doi.org/10.1145/2696454.2696483>
- Marina Fridin. 2014. Kindergarten social assistive robot: First meeting and ethical issues. *Comput. Hum. Behav.* 30 (2014), 262–272. DOI : <https://doi.org/10.1016/j.chb.2013.09.005>
- Miguel A. Garcia-Perez and Vicente Nunez-Anton. 2003. Cellwise residual analysis in two-way contingency tables. *Edu. Psychol. Measure.* 63, 5 (2003), 825–839. DOI : <https://doi.org/10.1177/0013164403251280>
- Susan B. Gerber, Jeremy D. Finn, Charles M. Achilles, and Jayne Boyd-Zaharias. 2001. Teacher aides and students' academic achievement. *Edu. Eval. Policy Anal.* 23, 2 (2001), 123–143. DOI : <https://doi.org/10.3102/01623737023002123>
- Jeong-Hye Han, Mi-Heon Jo, Vicki Jones, and Jun-H Jo. 2008. Comparative study on the educational use of home robots for children. *J. Info. Process. Syst.* 4, 4 (2008), 159–168. DOI : <https://doi.org/10.3745/JIPS.2008.4.4.159>
- James Hendler. 2000. *Robots for Kids: Exploring New Technologies for Learning*. Morgan Kaufmann.
- Khe Foon Hew and Thomas Brush. 2007. Integrating technology into K-12 teaching and learning: Current knowledge gaps and recommendations for future research. *Edu. Technol. Res. Dev.* 55, 3 (2007), 223–252. DOI : <https://doi.org/10.1007/s11423-006-9022-5>
- Martin L. Hoffman. 2001. *Empathy and Moral Development: Implications for Caring and Justice*. Cambridge University Press.
- Deanna Hood, Séverin Lemaignan, and Pierre Dillenbourg. 2015. When children teach a robot to write: An autonomous teachable humanoid which uses simulated handwriting. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 83–90. DOI : <https://doi.org/10.1145/2696454.2696479>
- Zac E. Imel, Jacqueline S. Barco, Halley J. Brown, Brian R. Baucom, John S. Baer, John C. Kircher, and David C. Atkins. 2014. The association of therapist empathy and synchrony in vocally encoded arousal. *J. Counsel. Psychol.* 61, 1 (2014), 146. DOI : <https://doi.org/10.1037/a0034943>
- Felix Jimenez, Masayoshi Kanoh, Tomohiro Yoshikawa, and Takeshi Furuhashi. 2014. Effect of collaborative learning with robot that prompts constructive interaction. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*. IEEE, 2983–2988. DOI : <https://doi.org/10.1109/SMC.2014.6974384>
- Aidan Jones, Susan Bull, and Ginevra Castellano. 2017. “I know that now, I'm going to learn this next”—Promoting self-regulated learning with a robotic tutor. *Int. J. Soc. Robot.* (2017), 1–16. DOI : <https://doi.org/10.1007/s12369-017-0430-y>
- Aidan Jones and Ginevra Castellano. 2018. Adaptive robotic tutors that support self-regulated learning: A longer-term investigation with primary school children. *Int. J. Soc. Robot.* (2018), 1–14. DOI : <https://doi.org/10.1007/s12369-017-0458-z>
- Aidan Jones, Dennis Küster, Christina Anne Basedow, Patricia Alves-Oliveira, Sofia Serholt, Helen Hastie, Lee J. Corrigan, Wolmet Barendregt, Arvid Kappas, Ana Paiva, and Ginevra Castellano. 2015. Empathic robotic tutors for personalised learning: A multidisciplinary approach. In *Proceedings of the International Conference on Social Robotics*. Springer, 285–295. DOI : https://doi.org/10.1007/978-3-319-25554-5_29
- Malte F. Jung, Selma Šabanović, Friederike Eysel, and Marlena Fraune. 2017. Robots in groups and teams. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 401–407. DOI : <https://doi.org/10.1145/3022198.3022659>
- Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. 2004. Interactive robots as social partners and peer tutors for children: A field trial. *Hum.-Comput. Interact.* 19, 1 (2004), 61–84. DOI : <https://doi.org/10.1080/07370024.2004.9667340>
- James Kennedy, Paul Baxter, and Tony Belpaeme. 2015a. Comparing robot embodiments in a guided discovery learning interaction with children. *Int. J. Soc. Robot.* 7, 2 (2015), 293–308. DOI : <https://doi.org/10.1007/s12369-014-0277-4>
- James Kennedy, Paul Baxter, and Tony Belpaeme. 2015b. The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction*, Vol. 15. 67–74. DOI : <https://doi.org/10.1145/2696454.2696457>
- James Kennedy, Paul Baxter, Emmanuel Senft, and Tony Belpaeme. 2016. Social robot tutoring for child second language learning. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction*. IEEE, 231–238. DOI : <https://doi.org/10.1109/HRI.2016.7451757>

- Myint Swe Khine. 2017. *Robotics in STEM Education: Redesigning the Learning Experience*. Springer.
- Erik Knol and Peter W. De Vries. 2011. EnerCities—A serious game to stimulate sustainability and energy conservation: Preliminary results. *eLearning Papers* 25 (2011).
- Jean Lave and Etienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge University Press.
- Iolanda Leite, Ginevra Castellano, André Pereira, Carlos Martinho, and Ana Paiva. 2014. Empathic robots for long-term interaction. *Int. J. Soc. Robot.* 6, 3 (2014), 329–341. DOI: <https://doi.org/10.1007/s12369-014-0227-1>
- Iolanda Leite, Carlos Martinho, and Ana Paiva. 2013a. Social robots for long-term interaction: A survey. *Int. J. Soc. Robot.* 5, 2 (2013), 291–308. DOI: <https://doi.org/10.1007/s12369-013-0178-y>
- Iolanda Leite, André Pereira, Samuel Mascarenhas, Carlos Martinho, Rui Prada, and Ana Paiva. 2013b. The influence of empathy in human-robot relations. *Int. J. Hum.-Comput. Studies* 71, 3 (2013), 250–260. DOI: <https://doi.org/10.1016/j.ijhcs.2012.09.005>
- Séverin Lemaignan, Alexis Jacq, Deanna Hood, Fernando Garcia, Ana Paiva, and Pierre Dillenbourg. 2016. Learning by teaching a robot: The case of handwriting. *IEEE Robot. Autom. Mag.* 23, 2 (2016), 56–66. DOI: <https://doi.org/10.1109/MRA.2016.2546700>
- Daniel Leyzberg, Samuel Spaulding, Mariya Toneva, and Brian Scassellati. 2012. The physical presence of a robot tutor increases cognitive learning gains. In *Proceedings of the Cognitive Science Society*, Vol. 34.
- Nan Li, Himanshu Verma, Afroditi Skevi, Guillaume Zufferey, Jan Blom, and Pierre Dillenbourg. 2014. Watching MOOCs together: Investigating co-located MOOC study groups. *Distance Edu.* 35, EPFL-ARTICLE-199425 (2014), 217–233. DOI: <https://doi.org/10.1080/01587919.2014.917708>
- Christoforos I. Mavrogiannis, Wil B. Thomason, and Ross A. Knepper. 2018. Social momentum: A framework for legible navigation in dynamic multi-agent environments. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 361–369. DOI: <https://doi.org/10.1145/3171221.3171255>
- Gretchen McAllister and Jacqueline Jordan Irvine. 2002. The role of empathy in teaching culturally diverse students: A qualitative study of teachers' beliefs. *J. Teach. Edu.* 53, 5 (2002), 433–443. DOI: <https://doi.org/10.1177/002248702237397>
- Marek P. Michalowski, Selma Sabanovic, and Hideki Kozima. 2007. A dancing robot for rhythmic social interaction. In *Proceedings of the 2nd ACM/IEEE International Conference on Human-Robot Interaction*. IEEE, 89–96. DOI: <https://doi.org/10.1145/1228716.1228729>
- N. Miyake and S. Y. Okita. 2012. Robot facilitation as dynamic support for collaborative learning. In *Proceedings of the International Conference of the Learning Sciences*. 57–63.
- Janet Moore. 2005. Barriers and pathways to creating sustainability education programs: Policy, rhetoric, and reality. *Environ. Edu. Res.* 11, 5 (2005), 537–555. DOI: <https://doi.org/10.1080/13504620500169692>
- Omar Mubin, Catherine J. Stevens, Suleman Shahid, Abdullah Al Mahmud, and Jian-Jie Dong. 2013. A review of the applicability of robots in education. *J. Technol. Edu. Learn.* 1 (2013), 209–0015. DOI: <https://doi.org/10.2316/Journal.209.2013.1.209-0015>
- H. Nwana. 1990. Intelligent tutoring systems: An overview. *Artific. Intell. Rev.* 4 (1990), 251–277. DOI: <https://doi.org/10.1007/BF00168958>
- Ana Paiva, Iolanda Leite, Hana Boukricha, and Ipke Wachsmuth. 2017. Empathy in virtual agents and robots: A survey. *ACM Trans. Interact. Intell. Syst.* 7, 3 (2017), 11. DOI: <https://doi.org/10.1145/2912150>
- Ana Paiva, Samuel Mascarenhas, Sofia Petisca, Filipa Correia, and Patrícia Alves-Oliveira. 2018. Toward more humane machines: Creating emotional social robots. In *New Interdisciplinary Landscapes in Morality and Emotion*. Routledge, 125–139.
- Seymour Papert. 1980. *Mindstorms: Children, Computers, and Powerful Ideas*. Basic Books, Inc.
- Hae Won Park, Mirko Gelsomini, Jin Joo Lee, and Cynthia Breazeal. 2017a. Telling stories to robots: The effect of backchanneling on a child's storytelling. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 100–108. DOI: <https://doi.org/10.1145/2909824.3020245>
- Hae Won Park, Rinat B. Rosenberg-Kima, Maor Rosenberg, Goren Gordon, and Cynthia Breazeal. 2017b. Growing growth mindset with a social robot peer. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. 137–145. DOI: <https://doi.org/10.1145/2909824.3020213>
- Jean Piaget. 2013. *Play, Dreams, and Imitation in Childhood*. Routledge.
- Stephanie D. Preston and Frans B. M. De Waal. 2002. Empathy: Its ultimate and proximate bases. *Behav. Brain Sci.* 25, 1 (2002), 1–20. DOI: <https://doi.org/10.1017/S0140525X02000018>
- Laurel D. Riek, Philip C. Paul, and Peter Robinson. 2010. When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *J. Multimod. User Interfaces* 3, 1-2 (2010), 99–108. DOI: <https://doi.org/10.1007/s12193-009-0028-2>
- U. Ritterfeld, M. Cody, and P. Vorderer (Eds.). 2009. *Serious Games: Mechanisms and Effects*. Routledge.
- Debora L. Roorda, Helma M. Y. Koomen, Jantine L. Spilt, and Frans J. Oort. 2011. The influence of affective teacher-student relationships on students' school engagement and achievement: A meta-analytic approach. *Rev. Edu. Res.* 81, 4 (2011), 493–529. DOI: <https://doi.org/10.3102/0034654311421793>

- Ernest Savage and Leonard Sterry. 1990. *A Conceptual Framework for Technology Education*. Technical Report. International Technology Education Association, Reston, VA.
- Catherine E. Sembroski, Marlena R. Fraune, and Selma Šabanović. 2017. He said, she said, it said: Effects of robot group membership and human authority on people's willingness to follow their instructions. In *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'17)*. IEEE, 56–61. DOI: <https://doi.org/10.1109/ROMAN.2017.8172280>
- Pedro Sequeira, Patrícia Alves-Oliveira, Tiago Ribeiro, Eugenio Di Tullio, Sofia Petisca, Francisco S. Melo, Ginevra Castellano, and Ana Paiva. 2016. Discovering social interaction strategies for robots from restricted-perception wizard-of-oz studies. In *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI'16)*. IEEE, 197–204. DOI: <https://doi.org/10.1109/HRI.2016.7451752>
- Pedro Sequeira and Cláudia Antunes. 2010. Real-time sensory pattern mining for autonomous agents. In *Proceedings of the 6th International Workshop on Agents and Data Mining Interaction, ADMI 2010 (ADMI'10)*, Longbing Cao, Ana L. C. Bazzan, Vladimir Gorodetsky, Pericles A. Mitkas, Gerhard Weiss, and Philip S. Yu (Eds.). Springer, Berlin, 71–83. DOI: https://doi.org/10.1007/978-3-642-15420-1_7
- Pedro Sequeira, Francisco S. Melo, and Ana Paiva. 2013. An associative state-space metric for learning in factored MDPs. In *Proceedings of the 16th Portuguese Conference on Artificial Intelligence (EPIA'13)*, Luís Correia, Luís Paulo Reis, and José Cascalho (Eds.). Springer, Berlin, 163–174. DOI: https://doi.org/10.1007/978-3-642-40669-0_15
- Pedro Sequeira, Francisco S. Melo, and Ana Paiva. 2015. "Let's save resources!": A dynamic, collaborative AI for a multi-player environmental awareness game. In *Proceedings of the IEEE Conference on Computational Intelligence and Games (IEEE CIG'15)*, Longbing Cao, Ana L. C. Bazzan, Vladimir Gorodetsky, Pericles A. Mitkas, Gerhard Weiss, and Philip S. Yu (Eds.). IEEE, 399–406. DOI: <https://doi.org/10.1109/CIG.2015.7317952>
- Sofia Serholt. 2018. Breakdowns in children's interactions with a robotic tutor: A longitudinal study. *Comput. Hum. Behav.* 81 (2018), 250–264. DOI: <https://doi.org/10.1016/j.chb.2017.12.030>
- Sofia Serholt, Wolmet Barendregt, Dennis Küster, Aidan Jones, Patrícia Alves-Oliveira, and Ana Paiva. 2016. Students' normative perspectives on classroom robots. *What Social Robots Can and Should Do: Proceedings of Robophilosophy (TRANSOR'16)* 290 (2016), 240. DOI: <https://doi.org/10.3233/978-1-61499-708-5-240>
- Sofia Serholt, Wolmet Barendregt, Asimina Vasalou, Patrícia Alves-Oliveira, Aidan Jones, Sofia Petisca, and Ana Paiva. 2017. The case of classroom robots: Teachers' deliberations on the ethical tensions. *AI Society* 32, 4 (2017), 613–631. DOI: <https://doi.org/10.1007/s00146-016-0667-2>
- Elaine Short and Maja J. Mataric. 2017. Robot moderation of a collaborative game: Toward socially assistive robotics in group interactions. In *Proceedings of the 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'17)*. IEEE, 385–390. DOI: <https://doi.org/10.1109/ROMAN.2017.8172331>
- Newton Spolaôr and Fabiane B. Vavassori Benitti. 2017. Robotics applications grounded in learning theories on tertiary education: A systematic review. *Comput. Edu.* 112 (2017), 97–107. DOI: <https://doi.org/10.1016/j.compedu.2017.05.001>
- Leslie P. Steffe and Jerry Edward Gale. 1995. *Constructivism in Education*. Lawrence Erlbaum Hillsdale, NJ.
- Sarah Strohkorb, Iolanda Leite, Natalie Warren, and Brian Scassellati. 2015. Classification of children's social dominance in group interactions with robots. In *Proceedings of the ACM on International Conference on Multimodal Interaction*. ACM, 227–234. DOI: <https://doi.org/10.1145/2818346.2820735>
- Sarah Strohkorb Sebo, Margaret Traeger, Malte Jung, and Brian Scassellati. 2018. The ripple effects of vulnerability: The effects of a robot's vulnerable behavior on trust in human-robot teams. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 178–186. DOI: <https://doi.org/10.1145/3171221.3171275>
- Fumihide Tanaka, Aaron Cicourel, and Javier R. Movellan. 2007. Socialization between toddlers and robots at an early childhood education center. *Proc. Natl. Acad. Sci. U.S.A.* 104, 46 (2007), 17954–17958. DOI: <https://doi.org/10.1073/pnas.0707769104>
- Fumihide Tanaka and Shizuko Matsuzoe. 2012. Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *J. Hum.-Robot Interact.* 1, 1 (2012), 78–95. DOI: <https://doi.org/10.5898/JHRI.1.1.Tanaka>
- Adriana Tapus and Maja J. Mataric. 2007. Emulating empathy in socially assistive robotics. In *Proceedings of the AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*. 93–96.
- Maciej Tomczak and Ewa Tomczak. 2014. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends Sport Sci.* 21, 1 (2014).
- Hans van der Meij, Eefje Albers, and Henny Leemkuil. 2011. Learning from games: Does collaboration help? *British J. Edu. Technol.* 42, 4 (2011), 655–664. DOI: <https://doi.org/10.1111/j.1467-8535.2010.01067.x>
- Marynel Vázquez, Elizabeth J. Carter, Braden McDorman, Jodi Forlizzi, Aaron Steinfeld, and Scott E. Hudson. 2017. Toward robot autonomy in group conversations: Understanding the effects of body orientation and gaze. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 42–52. DOI: <https://doi.org/10.1145/2909824.3020207>

- Jorick Visser, Lode De Bot, and Bieke Zaman. 2013. MemoLine: Evaluating long-term UX with children. In *Proceedings of the 12th International Conference on Interaction Design and Children*. ACM, 285–288. DOI : <https://doi.org/10.1145/2485760.2485836>
- Jacqueline M. Kory Westlund, Hae Won Park, Randi Williams, and Cynthia Breazeal. 2018. Measuring young children’s long-term relationships with social robots. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. ACM, 207–218. DOI : <https://doi.org/10.1145/3202185.3202732>
- P. Wittenburg, H. Brugman, A. Russel, A. Klassmann, and H. Sloetjes. 2006. ELAN: A professional framework for multi-modality research. In *Proceedings of the 5th International Conference on Language Resources and Evaluation*. 1556–1559.
- Pieter Wouters, Christof Van Nimwegen, Herre Van Oostendorp, and Erik D. Van Der Spek. 2013. A meta-analysis of the cognitive and motivational effects of serious games. *J. Edu. Psychol.* 105, 2 (2013), 249–265. DOI : <https://doi.org/10.1037/a0031311>
- Cristina Zaga, Manja Lohse, Khiet P. Truong, and Vanessa Evers. 2015. The effect of a robot’s social character on children’s task engagement: Peer versus tutor. In *Proceedings of the International Conference on Social Robotics*. Springer, Cham, 704–713. DOI : https://doi.org/10.1007/978-3-319-25554-5_70

Received November 2017; revised July 2018; accepted December 2018