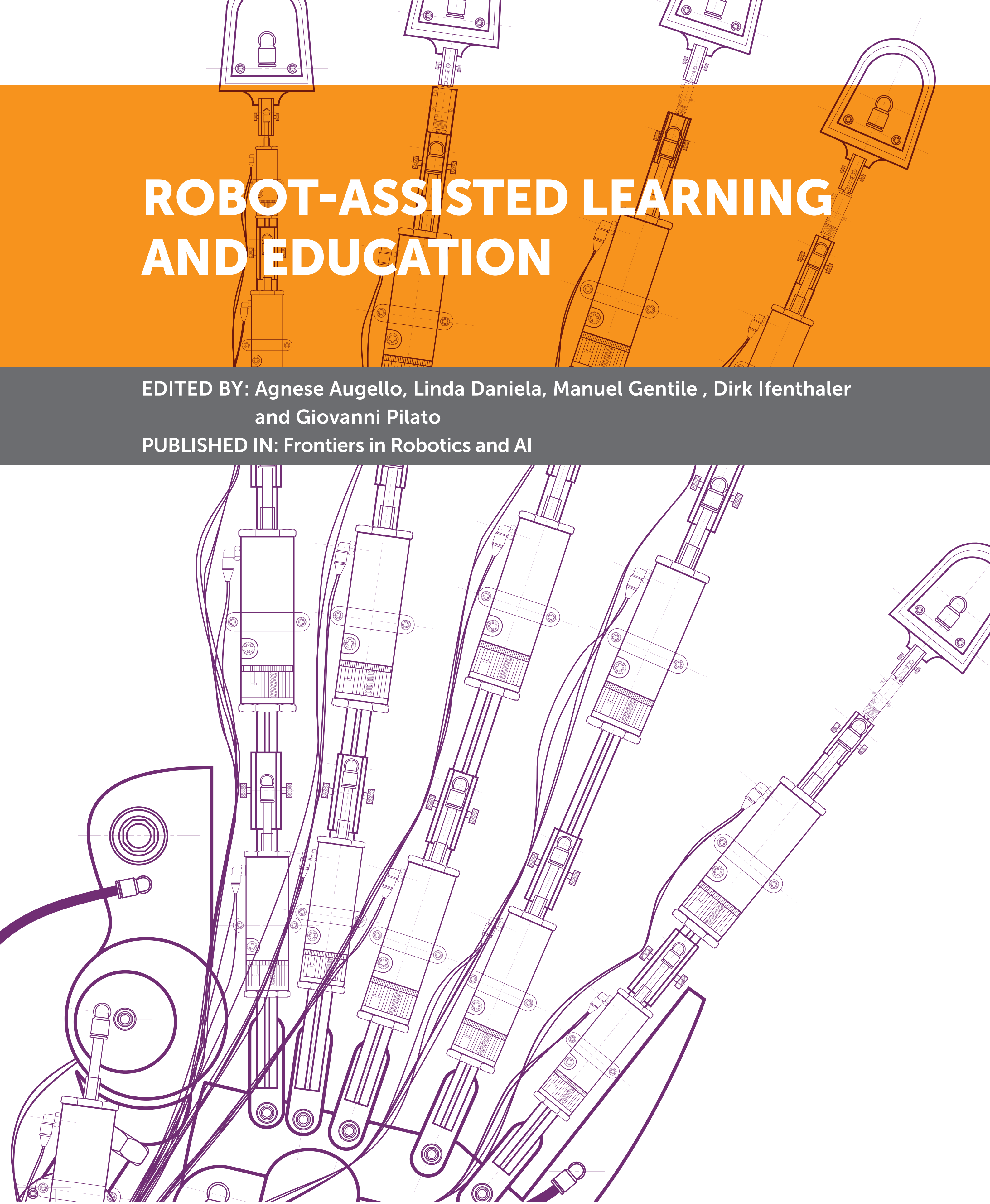


ROBOT-ASSISTED LEARNING AND EDUCATION

EDITED BY: Agnese Augello, Linda Daniela, Manuel Gentile , Dirk Ifenthaler and Giovanni Pilato

PUBLISHED IN: Frontiers in Robotics and AI

frontiers Research Topics

ISSN 1664-8714
ISBN 978-2-88966-325-5
DOI 10.3389/978-2-88966-325-5

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.
Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

ROBOT-ASSISTED LEARNING AND EDUCATION

Topic Editors:
Agnese Augello, Institute for High Performance Computing and Networking (ICAR), Italy
Linda Daniela, University of Latvia, Latvia
Manuel Gentile, Istituto per le Tecnologie Didattiche ITD - Consiglio Nazionale delle Ricerche, Italy
Dirk Ifenthaler, Universität Mannheim, Germany
Giovanni Pilato, Institute for High Performance Computing and Networking (ICAR), Italy

Citation: Augello, A., Daniela, L., Gentile, M., Ifenthaler, D., Pilato, G., eds. (2021). Robot-Assisted Learning and Education. Lausanne: Frontiers Media SA.
doi: 10.3389/978-2-88966-325-5

Table of Contents

frontiers in Robotics and AI

EDITORIAL
published: 09 November 2020
doi: 10.3389/frobt.2020.591319

Editorial: Robot-Assisted Learning and Education

Agnese Augello[1]*, *Linda Daniela*[2], *Manuel Gentile*[3], *Dirk Ifenthaler*[4,5] *and Giovanni Pilato*[1]

[1] Institute for High Performance Computing and Networking (ICAR), National Research Council (CNR), Palermo, Italy, [2] University of Latvia, Riga, Latvia, [3] Institute for Educational Technology (ITD), National Research Council (CNR), Palermo, Italy, [4] Learning, Design and Technology, University of Mannheim, Mannheim, Germany, [5] UNESCO Deputy-Chair of Data Analytics in Higher Education Learning and Teaching, Curtin University Perth, Bentley, WA, Australia

Keywords: HRI (human robot interaction), technology-enhanced learning, educational robotics, social robotics, robotics for collaborative learning

Editorial on the Research Topic

Robot-Assisted Learning and Education

Robots are increasingly being introduced in social environments to support the process of learning (e.g., Atmatzidou and Demetriadis, 2016; El Hamamsy et al., 2019; Kory-Westlund and Breazeal, 2019; Vogt et al., 2019) with different roles, such as smart teaching platforms, assistants, and in some cases also as companions and co-learners (Brown and Howard, 2014; Gordon et al., 2015; Pandey and Gelin, 2016; Belpaeme et al., 2018). Empirical research in educational robotics (ER) focuses on the adaptation of the robot behavior to specific learning needs and assessment of student learning and understanding. It is common to use robots to foster STEM and STEAM curricula (Brown and Howard, 2014; Shiomi et al., 2015; Città et al., 2017) with positive outcomes (Benitti, 2012). Research in ER have documented a greater involvement of students in learning activities, a support for critical thinking and complex problem solving as well as an increased comprehension of complex concepts and procedures, especially if the robots are endowed with a human-like appearance and social abilities (Leyzberg et al., 2012; Li, 2015). Some studies focused on the perceptions of robots and their social behavior and the consequent effects on learning (e.g., Mutlu et al., 2006; Kory and Breazeal, 2014; Michaelis and Mutlu, 2019), to support the process of understanding and memorization of concepts and the interpretation of emotional contents and social dynamics (Leite et al., 2017; Park et al., 2019; Bono et al., 2020; Conti et al., 2020).

The contributions in the Research Topic focuses on robotics approaches and architectures supporting human learning. Scaradozzi et al. apply machine learning techniques for the identification of different problem-solving pathways. Authors came to the conclusion that a "steadier incremental steps" strategy of programming correlated to a better performance in the resolution of the exercise. This supports the idea that a step by step knowledge building process is more effective than a big changes approach.

D'Amico et al. prove that introducing a robot leads to a better understanding of STEM concepts and higher participation in the activities. According to the authors, ER combines physical and mental experiences, which allow students to learn by doing, to manipulate concepts, and to embody cognition. During ER sessions, students had the opportunity to approach an idea from both an abstract and a concrete point of view. This leads to the creation of different forms of memory (semantic and procedural) and accurate episodic learning. The authors also conclude that robotics may increase motivation for learning in situations that are generally seen by children as passive and not very stimulating.

The development of cognitive strategies for the transition from exploratory actions toward intentional problem-solving in children is in the center of the development of human cognition.

OPEN ACCESS

Edited and reviewed by:
Bilge Mutlu,
University of Wisconsin-Madison,
United States

Correspondence:
Agnese Augello
agnese.augello@icar.cnr.it

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *05 August 2020*
Accepted: *22 September 2020*
Published: *09 November 2020*

Citation:
Augello A, Daniela L, Gentile M, Ifenthaler D and Pilato G (2020) Editorial: Robot-Assisted Learning and Education. Front. Robot. AI 7:591319. doi: 10.3389/frobt.2020.591319

Charisi et al. illustrate an exploratory behavioral study to show the relationship between child-robot voluntary interaction and both the problem-solving process and performance of a child. In their study, the authors pay particular attention to the importance of exploration. Twenty children took part in the study, including 72 sessions with 113 Tower of Hanoi tasks. The platform used was a tabletop robot. The findings indicate that the children who participated in the voluntary interaction setting showed a better performance in the problem-solving activity. Implications are considered for the development of intelligent robotic systems that allow child-initiated interaction as well as targeted and not constant robot interventions.

De Haas et al. investigate how feedback from a robot can influence children's engagement and support second language learning; 72 children (5 years old) learned animal names from a humanoid in three different sessions, receiving varying types of feedback from a robot. The findings indicate that children tended to be more engaged with the robot and task when the robot used a preferred type of feedback. Implications suggest the use of robots and varying feedback in long-term interactions where engagement of children often drops.

Zhexenova et al. verify the effect of using a robot to help primary school children learning a newly-adopted script and its handwriting system. The differences between using the robot with a tablet, a tablet-only, and a teacher were not significant, revealing a similar learning effect in the three conditions. An important outcome is that children's mood improved when interacting with the robot compared to other learning aids considered in the study.

Guneysu Ozgur et al. analyze the possible role of haptic-enabled tangible robots in training visual-spatial skills. They designed an educational path to support children in learning to write cursive letters proposing tasks based on playful and collaborative activities. Starting from previous experience and applying an iterative approach, the authors adapted the activities for children with attention and visuomotor coordination issues. The experimental results gathered within occupational therapy sessions provide exciting insights (children having writing problems can improve in letter writing after the use of the system for just one session) and open up further research perspectives.

The work of Kostrubiec and Kruck belongs to the growing field of robotics for therapeutic support of children with autism syndrome. Compared to the literature, this work is characterized by the goal of collecting pieces of evidence and suggestions that can guide the realization of new robotic tools. Experimental activities have been carried out according to the ABA approach by using a spherical robotic tool not yet available on the market. The results, while showing a good acceptance by educators about the adoption of the robot, confirm some undesirable effects typical of the use of robots in these contexts, such as the difficulty of these tools to be efficient social mediators. The work highlights the need to look beyond the purely technological aspect, and to analyze in more detail how technologies integrate and interact with the adopted therapeutic approach and with the physical and social environment in which the therapies are conducted.

ER convey other important aspects. As an example, instead of focusing on the personal side of the learning process, several works investigated the use of robots in collaborative learning activities (e.g., Jung et al., 2015; Alves-Oliveira et al., 2019; Oliveira et al., 2019), to foster mutual cooperation between students strengthening the formation of social links and to support inclusive education (Catlin and Blamires, 2019; Daniela and Lytras, 2019).

Rosenberg-Kima et al. compare the effectiveness of a social-robot and a human instructor in facilitating groups in the classroom. A tablet application mediated the students' interactions to overcome the limitation of the robot in managing verbal communication. The study highlights that the physical presence of the robot and factors such as perceived intelligence, anthropomorphism, likeability, significantly influence the efficacy of the facilitator role played by the robot. Improving communication skills and providing the robot with the ability to solve situations typical of collaborative works could increase the effectiveness of these interventions.

Ponticorvo et al. show that ER can be more effective in promoting positive social ties and connections between students than other tasks when it is proposed as a group activity. A study on secondary school students (in an area strongly affected by school dropout) compares the outcomes obtained by three situations: (1) a laboratory with robots, (2) a laboratory with Scratch used for coding, and (3) a control group. The results confirm that the involvement of students in a robotics lab can effectively encourage the rise of ties among students. Furthermore, the ER, together with sociometric tools, can be used to evaluate group dynamics in a synthetic and manageable manner.

The work presented by Serholt et al. focuses the attention on troublesome situations that can occur during interaction with a robot in a classroom setting. Video analysis of children's group interactions with a robot tutee within the context of a mathematics game was conducted by examining the nature of the troubles and the strategies employed both by individual children and other involved actors to address with them. The results show as troubles mainly related to the robot's social norm violations that, although it could be traced back to technical limitations, are considered from a social interaction perspective (e.g., irrelevant comments and interruptions of the robots).

Other works shift the attention on the robot's learning side, taking inspiration from well-known learning techniques such as learning by imitation, to design algorithms enabling robots to learn procedures through observations and interaction with a human being (Tai et al., 2016; Hussein et al., 2017; Zhu and Hu, 2018).

Focusing on the recognition process of Contact States (CSs) during an assembly task Al-Yacoub et al. consider an imitation Learning approach, observing that humans can effectively manage assembly tasks by using haptic Force/Torque feedback. Collected F/T data were pre-processed and segmented. A robot learned the extracted features by temporal knowledge modeling in the symbolic domain. This makes it possible to catch complex human behaviors with models that are simpler, more compact, and with better computational performances with regards to non-symbolic models. The features are used to train a probabilistic model. Experimental trials show the effectiveness of the approach, whose main advantages are its simplicity and the minimal a priori knowledge on the geometrical characteristics on the mating parts.

AUTHOR CONTRIBUTIONS

AA outlined the draft by collecting the editors comments. AA, LD, MG, DI, and GP have made a substantial, intellectual contribution to the work, and approved it for publication.

REFERENCES

Alves-Oliveira, P., Sequeira, P., Melo, F. S., Castellano, G., and Paiva, A. (2019). Empathic robot for group learning: a field study. *ACM Trans. Hum. Robot Interact.* (Daegu) 8, 1–34. doi: 10.1145/3300188

Atmatzidou, S., and Demetriadis, S. (2016). Advancing students' computational thinking skills through educational robotics: a study on age and gender relevant differences. *Robot. Auton. Syst.* 75, 661–670. doi: 10.1016/j.robot.2015.10.008

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: a review. *Sci. Robot.* 3:eaat5954. doi: 10.1126/scirobotics.aat5954

Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: a systematic review. *Comput. Educ.* 58, 978–988. doi: 10.1016/j.compedu.2011.10.006

Bono, A., Augello, A., Gentile, M., and Gaglio, S. (2020). "Social practices based characters in a robotic storytelling system," in *RO-MAN 2020 - The 29th IEEE International Symposium on Robot and Human Interactive Communication* (Naples: IEEE).

Brown, L. N., and Howard, A. M. (2014). "The positive effects of verbal encouragement in mathematics education using a social robot," in *2014 IEEE Integrated STEM Education Conference* (Princeton, NJ: IEEE), 1–5.

Catlin, D., and Blamires, M. (2019). Designing robots for special needs education. *Technol. Knowl. Learn.* 24, 291–313. doi: 10.1007/s10758-018-9378-8

Città, G., Arnab, S., Augello, A., Gentile, M., Zielonka, S. I., Ifenthaler, D., et al. (2017). "Move your mind: creative dancing humanoids as support to steam activities," in *International Conference on Intelligent Interactive Multimedia Systems and Services* (Vilamoura: Springer), 190–199.

Conti, D., Cirasa, C., Di Nuovo, S., and Di Nuovo, A. (2020). "Robot, tell me a tale!": a social robot as tool for teachers in kindergarten. *Interact. Stud.* 21, 220–242. doi: 10.1075/is.18024.con

Daniela, L., and Lytras, M. D. (2019). Educational robotics for inclusive education. *Technol. Knowl. Learn.* 24, 219–225. doi: 10.1007/s10758-018-9397-5

El Hamamsy, L., Johal, W., Asselborn, T., Nasir, J., and Dillenbourg, P. (2019). "Learning by collaborative teaching: an engaging multi-party cowriter activity," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (New Delhi: IEEE), 1–8.

Gordon, G., Breazeal, C., and Engel, S. (2015). "Can children catch curiosity from a social robot?," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, OR), 91–98.

Hussein, A., Gaber, M. M., Elyan, E., and Jayne, C. (2017). Imitation learning: a survey of learning methods. *ACM Comput. Surv.* 50, 1–35. doi: 10.1145/3054912

Jung, M. F., Martelaro, N., and Hinds, P. J. (2015). "Using robots to moderate team conflict: the case of repairing violations," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, OR), 229–236.

Kory, J., and Breazeal, C. (2014). "Storytelling with robots: learning companions for preschool children's language development," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (Edinburgh: IEEE), 643–648.

Kory-Westlund, J. M., and Breazeal, C. (2019). A long-term study of young children's rapport, social emulation, and language learning with a peer-like robot playmate in preschool. *Front. Robot. AI* 6:81. doi: 10.3389/frobt.2019.00081

Leite, I., McCoy, M., Lohani, M., Ullman, D., Salomons, N., Stokes, C., et al. (2017). Narratives with robots: The impact of interaction context and individual differences on story recall and emotional understanding. *Front. Roboti. AI* 4:29. doi: 10.3389/frobt.2017.00029

Leyzberg, D., Spaulding, S., Toneva, M., and Scassellati, B. (2012). "The physical presence of a robot tutor increases cognitive learning gains," in *Proceedings of the Annual Meeting of the Cognitive Science Society* (Sapporo), 34.

Li, J. (2015). The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *Int. J. Hum. Comput. Stud.* 77, 23–37. doi: 10.1016/j.ijhcs.2015.01.001

Michaelis, J. E., and Mutlu, B. (2019). "Supporting interest in science learning with a social robot," in *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (Boise, ID), 71–82.

Mutlu, B., Forlizzi, J., and Hodgins, J. (2006). "A storytelling robot: modeling and evaluation of human-like gaze behavior," in *2006 6th IEEE-RAS International Conference on Humanoid Robots* (Genova: IEEE), 518–523.

Oliveira, R., Arriaga, P., Correia, F., and Paiva, A. (2019). "The stereotype content model applied to human-robot interactions in groups," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (IEEE), 123–132.

Pandey, A. K., and Gelin, R. (2016). "Humanoid robots in education: a short review," in Humanoid Robotics: A Reference, eds A. Goswami and P. Vadakkepat, 1–16.

Park, H. W., Grover, I., Spaulding, S., Gomez, L., and Breazeal, C. (2019). "A model-free affective reinforcement learning approach to personalization of an autonomous social robot companion for early literacy education," in *Proceedings of the AAAI Conference on Artificial Intelligence* (Honolulu, HI), Vol. 33, 687–694.

Shiomi, M., Kanda, T., Howley, I., Hayashi, K., and Hagita, N. (2015). Can a social robot stimulate science curiosity in classrooms? *Int. J. Soc. Robot.* 7, 641–652. doi: 10.1007/s12369-015-0303-1

Tai, L., Zhang, J., Liu, M., Boedecker, J., and Burgard, W. (2016). A survey of deep network solutions for learning control in robotics: From reinforcement to imitation. *arXiv preprint* arXiv:1612.07139.

Vogt, P., van den Berghe, R., de Haas, M., Hoffman, L., Kanero, J., Mamus, E., et al. (2019). "Second language tutoring using social robots: a large-scale study," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Daegu: IEEE), 497–505.

Zhu, Z., and Hu, H. (2018). Robot learning from demonstration in robotic assembly: a survey. *Robotics* 7:17. doi: 10.3390/robotics7020017

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

METHODS
published: 17 October 2019
doi: 10.3389/frobt.2019.00099

Symbolic-Based Recognition of Contact States for Learning Assembly Skills

Ali Al-Yacoub[1], Yuchen Zhao[2], Niels Lohse[1], Mey Goh[1], Peter Kinnell[1], Pedro Ferreira[1] and Ella-Mae Hubbard[1]*

[1] Intelligent Automation Centre, Loughborough University, Loughborough, United Kingdom, [2] Beijing Ewaybot Technology LLC, Beijing, China

Imitation learning is gaining more attention because it enables robots to learn skills from human demonstrations. One of the major industrial activities that can benefit from imitation learning is the learning of new assembly processes. An essential characteristic of an assembly skill is its different contact states (CS). They determine how to adjust movements in order to perform the assembly task successfully. Humans can recognize CSs through haptic feedback. They execute complex assembly tasks accordingly. Hence, CSs are generally recognized using force and torque information. This process is not straightforward due to the variations in assembly tasks, signal noise and ambiguity in interpreting force/torque (F/T) information. In this research, an investigation has been conducted to recognize the CSs during an assembly process with a geometrical variation on the mating parts. The F/T data collected from several human trials were pre-processed, segmented and represented as symbols. Those symbols were used to train a probabilistic model. Then, the trained model was validated using unseen datasets. The primary goal of the proposed approach aims to improve recognition accuracy and reduce the computational effort by employing symbolic and probabilistic approaches. The model successfully recognized CS based only on force information. This shows that such models can assist in imitation learning.

Keywords: symbolic representation, imitation learning, feature transformation, Piecewise Aggregate Approximation (PAA), K-means, Hidden Markov Model (HMM)

Edited by:
Agnese Augello, Institute for High Performance Computing and Networking (ICAR), Italy

Reviewed by:
Soheil Keshmiri, Advanced Telecommunications Research Institute International (ATR), Japan
Jose James, Amrita Vishwa Vidyapeetham, India

***Correspondence:**
Ali Al-Yacoub
a.al-yacoub@lboro.ac.uk

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *15 March 2019*
Accepted: *26 September 2019*
Published: *17 October 2019*

Citation:
Al-Yacoub A, Zhao Y, Lohse N, Goh M, Kinnell P, Ferreira P and Hubbard E-M (2019) Symbolic-Based Recognition of Contact States for Learning Assembly Skills. Front. Robot. AI 6:99. doi: 10.3389/frobt.2019.00099

1. INTRODUCTION

Industrial robots can efficiently manipulate and assemble objects in a controlled environment with minimum variations. However, they have limitations in assembling parts with geometrical variations and tighter tolerances. In such applications, force signals play a crucial role especially when the robots have to interact with the surrounding environment. Nevertheless, the force signals are noisy and ambiguous to interpret and use (Wen et al., 2014). Humans, on the other hand, can robustly perform assembly tasks with tight tolerances (Park et al., 2008) because they are very efficient at using haptic (F/T) information, especially when vision cannot provide the required information. Consequently, robots can benefit from understanding how humans use such haptic feedback information during an assembly process. This can empower robots to use force and torque with human-like capabilities allowing them to learn and adapt according to the variations in the environment and adjust movement for tight tolerance assembly.

Most of the research work reported in the area of imitation learning is based on visual perception. This is mainly because humans mostly rely on vision to gain adequate information about objects' relative positions and their geometrical properties (Ernst and Banks, 2002; Rozo et al., 2013). In assembly applications, perception importance can vary with motion, where gross motion relies on vision while fine motion requires haptic information, especially in contact situations. The focus of the work reported in this paper is on the use of haptic information to learn an assembly task. Capturing human skills is particularly complicated for assembly processes which often involve an understanding of hidden process features and tacit knowledge. For example, for a successful assembly task, an understanding of various types of contacts between objects and their corresponding forces is required. Another important aspect of an assembly process is the sequential relations between different CSs during the assembly. Henceforth, different skilled operators can perform the stages of the same task with different temporal properties (transition between states and durations). In order to capture, understand and interpret human skills from a number of trials, those trials must be aligned (in terms of duration). Also, the underlying pattern of the haptic information must be extracted to reveal the sequential (temporal) knowledge (i.e., human skill). Hence, those skills must be modeled so that they can adapt to task variations for robotic assembly.

A great deal of research has been conducted on the recognition of the CS. The approaches for CS recognition can be arranged into two groups, i.e., analytical approaches and learning-based approaches. Essentially, the analytical model of the mating system has no single structure. The general model is composed of a set of analytical equations (sub-models), where each equation describes a particular contact state based on a physical analysis of the state. Furthermore, these sub-models usually rely on a set of approximation and assumptions to simplify the given problem. Hence, current analytical approaches to recognize CS is limited in terms of robustness and speed (Jakovljevic et al., 2012). The main limitation of analytical approaches is latency since it relies on a very complex computation (Nuttin, 1995). Learning approaches, on the other hand, appear to be a better alternative when taking the recognition of the CSs into consideration.

Various learning-based approaches to recognize CSs have been presented in the literature. For example, the Hidden Markov Model (HMM) has been implemented to recognize CS based on F/T information in tele-manipulation and result were presented in Hannaford and Lee (1990). However, the proposed models rely on extensive training and are only applicable to large clearance between the assembled parts. In Dong and Naghdy (2007) an HMM was used to recognize the CS of a Peg-in-Hole (PiH) assembly in a virtual environment, and to recognize the CS during the on-line PiH process. However, the accuracy of the trained HMM depended on the accuracy of the virtual world model which generally has nominal behavior. Lau (2003) proposed a framework of CS recognition in industrial robot assembly platform using HMM and F/T information, where it was experimentally proven that HMM-based with F/T is superior compared to the conventional CS recognition (CAD-based and kinematic-based).

Jasim et al. (2017) have developed a method that combines the Expectation Maximization and Gaussian Mixture Model (EM-GMM) to recognize the CS of PiH insertion during an automated process. In Jasim et al. (2017) the number of Gaussian were determined using Distribution Similarity Measure-based (DSM). In this research, the trained GMM models were evaluated using a rubber PiH insertions with two different parts elasticity. Yet, the work reported in Jasim et al. (2017) did not employ feature selection or transformation algorithms in order to reduce the computational effort. A Piecewise Affine Autoregressive Exogenous (PWARX) method has been presented in Okuda et al. (2008) to recognize the CS during the PiH assembly process. The core idea of the PWARX was used to control a robot during the PiH process based on a set of mathematical models (PWARX sets). In this case, the control was achieved by switching between the PWARX models using a Support Vector Machine (SVM). The SVM functionality was to recognize CS and accordingly switch over controllers to select the suitable models for the given CS. The computational power required for this method is quite high (Mikami et al., 2010), and the PWARX model is a complicated model (Nakabayashi et al., 2013). In Jakovljevic et al. (2012) a SVM has been employed to classify two successive states based on pre-designed features sequentially. The selected features were designed based on the quasi-static insertion force model (Whitney, 2004). This method relies on pre-defined features and a complex hierarchical classification algorithm since SVM is only a binary classification approach. This work also relied on designed features which were pre-selected by designers thus making the method less autonomous. Hertkorn et al. (2012) generated a wrench matrix based on the CAD models of the assembly parts with a particle filter to recognize the CS based on the F/T measurements. This method was implemented to resolve the ambiguity of the force measurements and recognize the contact formation of a rectangular workpiece on a flat surface. The drawback of this work was the simplicity of the part's geometries used to validate the proposed approach.

Jamali et al. (2014) presented a CS learning algorithm based on a symbolic representation of temporal behavior during robot valve opening process where force signals were clustered using the Minimum Message Length (MML) (Wallace and Dowe, 2000). The labeled symbolic data were used to train an HMM to recognize the CS. The overall accuracy achieved by this method was 81% about x-axis and 85% for rotation about the y-axis. Nevertheless, the convergence time of the GMM/MML might delay the recognition of the CS. Also, it relies on exploration movements in order to recognize the CS.

Most of the aforementioned research follow pattern recognition in the extracted/selected features by temporal knowledge modeling (capturing). This can be captured in the symbolic or non-symbolic domain. The main advantages of the non-symbolic models are their parametric nature and their capability to capture variations in human skills (Nejati and Könik, 2009). On the other hand, the symbolic approaches are well-known for capturing complex human behavior with simpler and shortened models that have better computational performance. For instance, symbolic approaches can capture the assembly sequence at different hierarchical levels (granularity), which is difficult using probabilistic approaches. Even though symbolic models have traditionally been considered unsuitable for controlling real-world systems (Calinon and Billard, 2008), researchers are now making effective use of these models for skills representation, evaluation, generalization and robot control (Mohammad and Nishida, 2014). These models are computationally efficient, simple, and capable of capturing complex human skills. Therefore, the research work reported in this paper explores the use of symbolic models to capture human assembly skills.

Despite significant progress in the field, researchers have been relying on algorithms which have significant latency. Furthermore, symbolic-based recognition of CS for imitation learning of PiH problems has not been sufficiently explored in the presenters of geometrical variation, in analogy to the material property (elasticity) variation presented in Jasim et al. (2017). In fact, probabilistic models trained based on symbolic representation converges faster than probabilistic models trained based on numeric representation (Kwiatkowska et al., 2004). Thus, it is believed that combining symbolic representation based on a simple segmentation approach [i.e., Piecewise Aggregate Approximation (PAA) or K-means] will result in more computationally efficient CS recognition with comparable robustness and accuracy.

This paper investigates a symbolic-based CS recognition approach which combines feature transformation methods, i.e., Principal Component Analysis (PCA), time-series segmentation, symbolic assignment, data labeling and HMM training, in order to reduce the computational effort required for CS recognition. As a validation example, the PiH assembly was adopted to demonstrate the efficiency of the proposed approach. Despite the apparent simplicity of the PiH assembly, it belongs to the group of parts mating problems that are highly non-linear and difficult process (Chen, 2011; Kronander et al., 2014). The main contribution of this paper is to develop a method that can identify contact states in an actual assembly process, i.e., PiH assembly. The development of this method involves the identification of CS during the PiH process based on symbolic representations of the force/torque signals (non-vision information). In addition to that, the relation between the probabilistic model and how robustly it responds to part variations (clearances) has been explored in this research.

The remainder of this paper is organized as follows; the problem description is introduced in section 2. Section 3 introduces the research methodology. The experimental setup is presented in section 4. The results are described in section 5 and a set of conclusions are drawn in section 6.

2. PROBLEM STATEMENT

The assembly process is generally split into two sub-tasks: gross motion and fine motion. In general, a gross motion is subject to no constraints in the environment, while during fine motion, the parts' movements are tightly constrained by the assembled parts' geometry. In this motion, a small error in a movement might cause an extensive force interaction leading to a failure of the assembly process. Hence, a force-based control is required to identify the CS and control the robot accordingly. In this context, the problem of CS recognition can be described as a classification problem, in which the F/T components are the raw data input $\mathbf{F} \in \mathbb{R}^{N\times 6}$ (three forces and three torques components in $x - y$ and z directions) (Equation 1), where N is the number of samples, and the output is $\mathbf{Y} \in \mathbb{R}^{N\times 1}$, where $\mathbf{Y}$ is a pre-defined CS. Accordingly, the goal of the CS model is to identify the contact state of a PiH assembly process.

$$\begin{gathered} \mathbf{F} = \left[f_0, \ldots, f_N\right] \\ \mathbf{f}_i = \left[F_x, F_y, F_z, T_x, T_y, T_z\right] \\ i = 0, \ldots, N \end{gathered} \tag{1}$$

Accordingly, the classification problem can be described as identifying a mapping function $h(\mathbf{F}, \mathbf{Y})$ that maps the given force measurements into a CS ($\mathbf{F} \overset{h(\mathbf{F},\mathbf{Y})}{\rightarrow} \mathbf{Y}$).

3. METHODOLOGY

The methodology adopted in this research relies on dimensionality reduction and symbolic representation of multi-dimensional F/T signals, which aims to recognize the CSs of an assembly process. In order to capture the CSs of a PiH insertion, the force/torque time-series data is recorded, filtered, normalized, its dimensionality reduced and the resulting time-series is represented in a string of symbols. The mapping of these time-series data can be performed under the assumption that the normalized time-series is Gaussian (Lin et al., 2003). Each symbol in the resultant string is labeled to match a member from a pre-defined CS set. The resultant strings and their associated labels set are used to train an HMM to capture the assembly process sequence.

The training approach adopted for this research is shown in **Figure 1**. The first step involves filtering and scaling F/T features using a low-pass filter and magnitude normalization, respectively. The data is projected into a new sub-space which maximizes the data variation and reduces dimensionality and noise using PCA. After that, the time-series is transformed to their symbolic representations. The symbolic representation is being assigned in two steps. Firstly, the time-series is segmented using Piecewise Aggregate Approximation (PAA) or K-means. Secondly, each segment from the previous step is being represented by a symbol based on its location in a normal distribution.

To verify the resulting models, unseen test sets were used. The accuracy of the trained models was measured based on a confusion matrix[1]. The pre-processing, feature transformation and symbolic representation stages of the research methodology are explained in more detail in the following sub-sections.

3.1. Pre-processing

The pre-processing consists of two stages, i.e., filtration and normalization, which are explained as described below.

[1] Performance of classification is commonly evaluated using the data in confusion matrix.

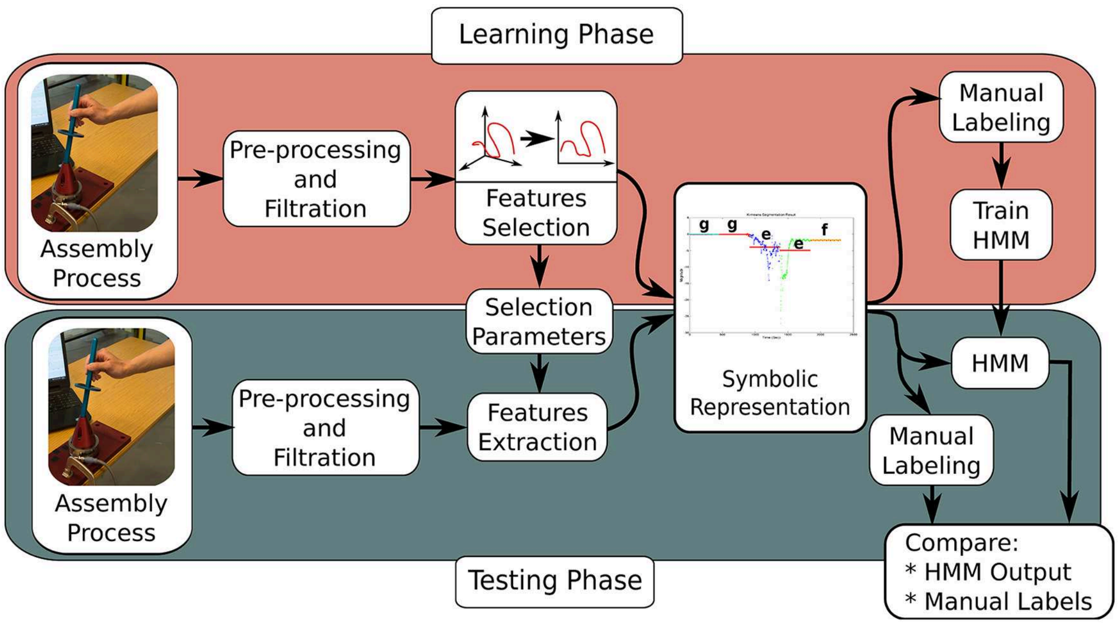


FIGURE 1 | Overview of the proposed approach (training and testing phases).

- Filtration: The F/T signals are subject to electromagnetic noise which severely affects the F/T signal. It is noticeable that the raw data from the F/T sensor contains random fluctuations, burrs and spikes. Shielding of sensors and their wiring can partially solve this problem. However, this is not always practical. In Wu et al. (2014), a comparison amongst different filters to alleviate the noise effects on F/T signals is presented. A performance measure, called stability index, was used to evaluate those different filters. In conclusion, it was recommended to use an FIR filter together with a Double-Threshold Filter (DTF). Hence, in this work, a finite impulse response (FIR) low-pass filter with DTF was adopted for the data pre-processing step of the F/T signal. The F/T signal was sampled at 500 Hz and filtered using a low-pass filter with 35 Hz cutoff frequency and DTF.
- Normalization: In order to capture and compare features that occur at different force levels on different trials, the force information during different trials needs to be normalized. Normalization is a powerful feature scaling method especially when the extreme values (minimum and maximum) of given features are unknown (Han, 2005; Jamali et al., 2015). On the other hand, the test data must be normalized based on the normalization coefficient of the training data.

3.2. Feature Transformation

Transformation can be perceived as a search algorithm that attempts to find a new set of features to make the machine learning problem easier (Liu and Motoda, 1998). PCA is one of the most common feature transformation tools that rely on allocating the directions that maximize the variation in the features' space (Sophian et al., 2003). The PCA is a mathematical tool used to analyse data sets based on their variations. One main characteristic of PCA is a reduction in dimensionality which often results from this tool. This dimensionality reduction involves the selection of features with maximum variation based on the accumulative-variance and a user-defined threshold (Calinon and Billard, 2008). The PCA threshold defined the amount of data which can be returned from the PCA after feature extraction.

3.3. Symbolic Representation

For the symbolic representation, the Symbolic Aggregate Approximation tool (SAX) was modified and employed in this research due to its simplicity. The SAX tool is a symbolic representation tool of time-series data that assigns the representation of numeric values based on Euclidean distance and discretization process (Lin et al., 2007). It also allows us to represent different time-series (various lengths) with the same number of symbols (Keogh et al., 2005). This property is of great importance in time-series alignments. The symbolic representation is achieved in two steps: time-series segmentation and segments mapping into symbols.

3.3.1. Time-Series Segmentations

Time-series segmentation can be achieved using PAA or K-means segmentation. In this paper, a brief comparison between the PAA segmentation and the well-known K-means time-series segmentation is presented.

3.3.2. Piecewise Aggregate Approximation (PAA)

The PAA splits time-series data with length N into M segments. This is very useful, especially in encoding temporal data during human demonstrations, where each trial has its different temporal properties (e.g., duration of each state). The PAA approximates a single time-series $S(n)$ into a vector of segments' averages; ($\bar{S} = (\bar{s}_1, \ldots, \bar{s}_M)$) for any random length ($M \leq N$), where each $\bar{s}_i$ is calculated as shown in Equation (2).

$$\bar{s}_i = \frac{M}{N} \sum_{t=\frac{N}{M}(i-1)+1}^{\frac{N.i}{M}} S(n) \tag{2}$$

Accordingly, the resulting time-series $\bar{S}(n)$ is shown in Equation (3).

$$\bar{S}(t) = \begin{cases} \bar{s}_1 & 0 \leq t < \frac{N}{M} \\ \vdots & \vdots \\ \bar{s}_i & \frac{N}{M}(i-1)+1 \leq n < \frac{N}{M}i \\ \vdots & \vdots \\ \bar{s}_M & \frac{N}{M}(M-1)+1 \leq n < N \end{cases} \tag{3}$$

The PAA represents a single time-series (1D) data into a sequence of averages $\bar{S}$. However, applying the PAA on multi-dimensional time-series will result in a sequence of vectors ($\bar{\mathbf{S}}$) where each element in the vector is a D dimension (selected features) corresponding to each time-series from the PCA. In this research, it is required to represent the multi-variable time-series with a single sequence of symbols. Hence, the PAA needed to be modified for the multi-variable time-series to be represented using a one-dimensional sequence of averages. Accordingly, a further dimensionality reduction is needed on the PCA result. This reduction can be performed using the average of the multi-dimensional data over different sectors of the PAA. Another alternative is to employ the norm of the multi-dimension data. In this paper, the norm method was used since it can be physically interpreted as the magnitude of the feature vector. Equation (4) represents the modified PAA using norm, where $\bar{S}(n)$ is a vector of data at time n.

$$\bar{s}_i = \frac{M}{N} \sum_{n=\frac{N}{M}(i-1)+1}^{\frac{N.i}{M}} ||\mathbf{S}(n)||_2^2 \tag{4}$$

The result from the PCA and its corresponding PAA results are shown in **Figure 11**. Then, each segment is mapped into a symbol as illustrated in the next section.

3.3.3. K-means Time-Series Segmentation

One of the simplest and most popular methods to overcome the clustering problem is the K-means algorithm (De la Torre and Kanade, 2006). K-means clustering splits a set of N samples (e.g., time-series) into M groups by maximizing the ratio amongst different clusters and the variation of each cluster. A K-segmentation of a time-series S is a sequence of mean values $\bar{S}$. Under consideration of the given context, the K-means problem can be described as the problem of allocating a segments boundaries (temporal information) (Vlachos et al., 2003). Equation (5) depicts the interval definition over all segments. The input for the K-means algorithm is the norm value of the multi-dimensional data from the PCA and the temporal information. The output is a time-series $\bar{S}(n)$, where each data point is represented by the centroid of the ith cluster/segment. The drawback of using K-means is its dependence on the initial estimation of the centroid and the number of clusters, which means that K-means might have different segmentation results for different initialization.

$$\begin{aligned} \bar{s}_i &\in \{s(t_a), \ldots, s(t_b)\} \\ (t_a)_i &= \min_t \bar{s}_i \\ (t_b)_i &= \max_t \bar{s}_i \end{aligned} \tag{5}$$

Where $\bar{s}_i$ is the *ith* segment that starts at $(t_a)_i$ and ends at $(t_b)_i$. The accuracy of the K-means was tested under a different number of clusters (as explained in section 5), and the number of segments with the best accuracy was selected. Based on PAA and K-means, the different time-series (trials) with different length N were represented using the same number of segments. The resulting segments have a unity magnitude. After that, each segment is represented by a single symbol based on its location in the normal distribution. It is worth mentioning that the number of K-means centroids and segments in the PAA were determined based on the elbow method, where classifier accuracy was tested with a different number of centroids and segments.

3.3.4. Segmentation Mapping

Having transferred the time-series data into segments (PAA or K-means), a further transformation must be applied to achieve the symbolic representation. Under the assumption that the normalized time-series is Gaussian as highlighted in section 3, the mapping of segments into symbols adopted in this paper has been introduced in Lin et al. (2007). In which, the outputs from the PAA and K-means are mapped into a series of symbols using predetermined "breakpoints" that produce equal-sized areas under a Gaussian curve with ($\mathcal{N}(0,1)$). The maximum number of breakpoints supported by the tool developed in Lin et al. (2007) is 12, these were adopted in this research to reduce the effect of the discretization error.

Figure 2 shows how a segmented signal based on subsection 3.3.1 mapped into symbols based on their location with respect to the predetermined breakpoints. Then, the force-time-series for the different trials are represented in a single sequence of symbols; e.g., ($Symbols := \{jjjiihcbaafff\}$), where a sequence of symbols encodes the CSs (hidden). From **Figure 2**, any segment that appears lower than the break line at -0.84 will have the symbol a throughout the trials, the force/torque time-series were represented using the same number of segments, even though the insertion process durations were different for each trial. Similar stages were represented using similar symbols using a normal distribution. For example, J and I are representing no-contact stage and H, and C are representing Chamfer-Crossing stage. Accordingly, different trials can be aligned using their corresponding symbols. The goal is to capture the relation between the recognized pattern (symbols) and the CSs. One possible solution for such a problem is to use an HMM.

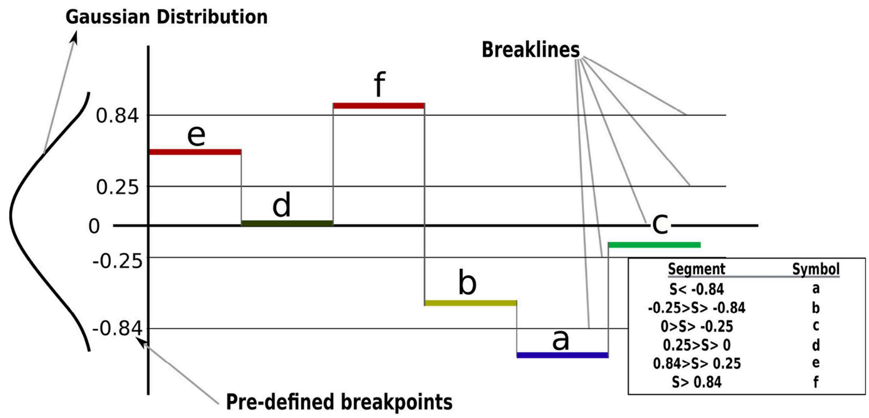


FIGURE 2 | Segments mapping into a symbolic representation of time-series data. For example, the time-series is segmented into six segments and each segment mapped into a symbol based on its location with respect to the Gaussian distribution (breaklines).

3.3.5. Manual Labeling of PiH Insertion

The resulting sequence of symbols introduced in the previous section is vague, and an expert should manually label it. A manual labeling process was performed based on analysing the F_z component of the data sets, because the F/T sensor is stationary and most of the force variation occurs on z direction. **Figure 3** illustrates the F_z component and the corresponding process stage based upon specific features of the F_z trend. The red circles indicate the start of a new stage and the end of the previous stage. The first circle highlights the force trend as the first contact occurs and the Chamfer-Crossing starts, as shown in stage 1 of **Figure 4**. After this, the operator starts correcting the angular error (the angle between the hole axis and insertion force direction). Once the angular error approaches zero (approximately), the friction force reaches its maximum due to further contact, which causes an overshoot in the force trend. This overshoot is highlighted in the second circle in **Figure 3**. Stage 1 of **Figure 5** shows the force analysis when the first contact point occurs and Equation (6) explores the force analysis at this stage. Stage 2 of **Figure 4** outlines the initial alignment, where the friction force F_{fr} is doubled whilst the insertion force F_{In} stays relatively constant as shown in Equation (7). This alignment explains the spike at the end of the Chamfer-Crossing (**Figure 3**, second circle). The insertion process then commences, and the peg is pushed fully into the hole. Once the peg is fully inserted in the hole, the operators release the peg causing a relaxation in the insertion force. This results in the small spike in the third circle in **Figure 3** which indicates the end of the insertion process. It is worth mentioning that these characterizations were observed in all PiH insertion trials. Therefore, the CS set (**Y**) is defined based on the PiH assembly stages as follow: ($\mathbf{Y} = \{NC, CC, I, FI\}$); where NC is No Contact State, CC is Chamfer-Crossing, I is Insertion, and FI is Full Insertion.

$$F_z = (F_{fr} + F_{No} - F_{In})\cos(\phi) \tag{6}$$

$$F_z = 2\,F_{fr} - F_{In} \tag{7}$$

The manual labeling of the symbolic representation was applied to enhance the process of obtaining human skills and to highlight the physical meaning of the discovered patterns. Also, the labeled data is only used for training and testing purposes and is not required for later interpretation of new PiH processes once the model has been verified.

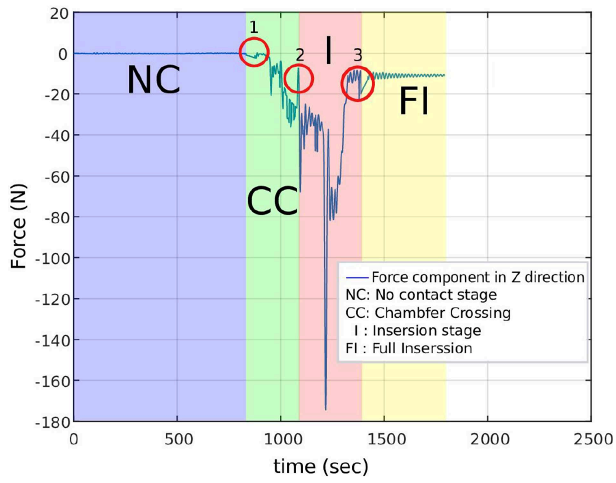


FIGURE 3 | Manual labeling of the PiH insertion process based on F_z force component.

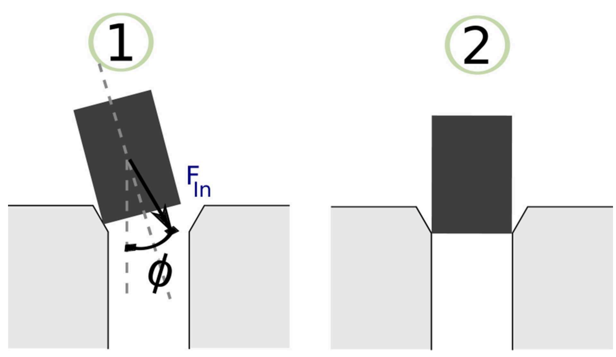


FIGURE 4 | Chamfer-Crossing stage.

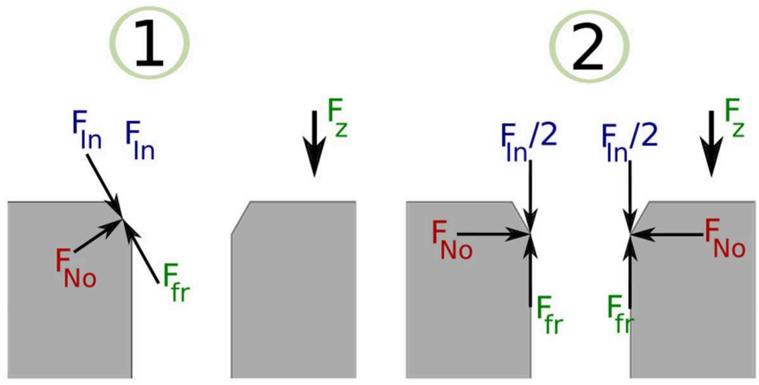


FIGURE 5 | Chamfer-Crossing PiH stage: force analysis.

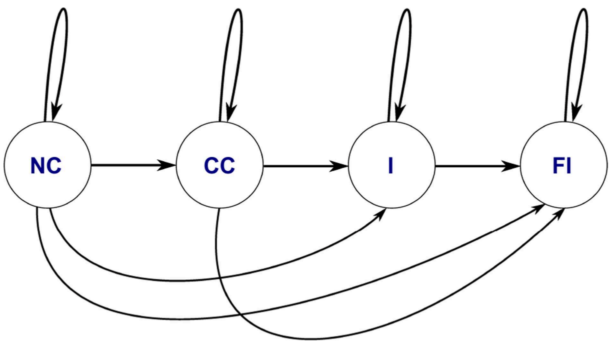


FIGURE 6 | HMM topology of the PiH insertion process based on the given dataset in Equation (8).

3.3.6. Hidden Markov Model (HMM)

Once the F/T information is transformed into strings of symbols which represent the temporal information of the PiH assembly process, an HMM was used to encode the temporal information and detect the pattern of each CS. Accordingly, each assembly trial (human demonstration) was represented in a string of symbols. The resulting strings were manually labeled by combining each symbol in the strings with one element from the CS set as explained in subsection 3.3.5. This resulting dataset can be represented as shown in Equation (8), which is used to train the HMM models. The same data set were used to initialize emission and transition matrix of the HMM using the Baum-Welch (BM) algorithm (Hochberg et al., 1991).

$$X_{Training\ Data} = ...(d, CC)(e, CC)(b, I)(c, I)(a, I)... \tag{8}$$

Figure 6 depicts the typology of the HMM used for the symbolic representation of each trial. This HMM encoded the PiH assembly skills, which was represented in a sequence of symbols. The HMM was trained using the string of symbols (as the observation) and CSs (as hidden states) to predict the new cases.

To summarize, the proposed approach is composed of three main stages. The first stage is segmentation which discovers the spatial structure within the data. Secondly, the symbolic representation reduces the high dimensional time-series data into one-dimensional data. The third stage captures the temporal knowledge embedded in the symbolic representation. For testing purposes, the labels for the randomly chosen test data sets were generated based on the trained model without using manual labeling. The results were then compared with the manual labels to evaluate the accuracy of the trained model.

4. EXPERIMENTATION SETUP AND DATA ACQUISITION

The experimental setup shown in **Figure 7** was used to collect data from different human operators performing a PiH assembly process. This setup was composed of a six-axis F/T sensor, a hole with a diameter D of 16.20 mm, and two round mating parts with different diameters (*Peg 1* and *Peg 2*). Where, the diameter of *Peg 1* is 15.98 mm and the diameter of *Peg 2* is 15.87 mm. **Figure 8** depicts one trial of the insertion process. The F/T data has been recorded while the human operators performed the assembly task.

A total number of 60 experiments were carried out with three different operators. Each operator performed 20 trials, to capture a wide range of human skills and variation in the initial position of the peg. Each trial contains on average 1,500 data points of F/T signals. The collected data were randomly split into training data (80% $\approx$ 48 trials) and test data (20% $\approx$ 12 trials). The six-dimensional time-series data (features) recorded by the F/T sensor was reduced to two-dimensional data using PCA. Then, the two-dimensional time-series data were reduced to 1D data by taking their norm value in the modified form of the PAA or K-means. After that, the segmented data were represented by a string of symbols. Those strings were labeled and used to train an HMM to discover the temporal aspects of the assembly process.

The quality of the classifier based on the HMM was evaluated using an unseen test set. This process was repeated four times to get an average performance of the classifier based on the proposed approach (see section 3). **Figure 1** depicts the evaluation process using the test set. It is worth mentioning here

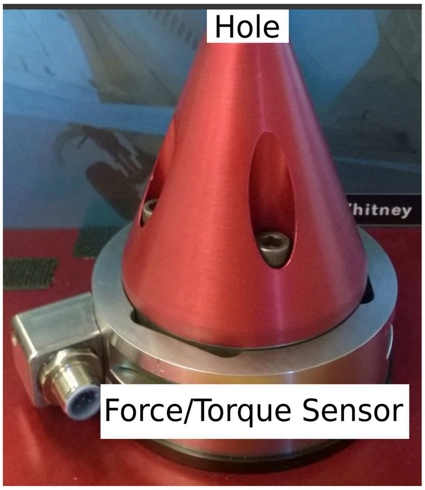


FIGURE 7 | Hole and F/T sensor.

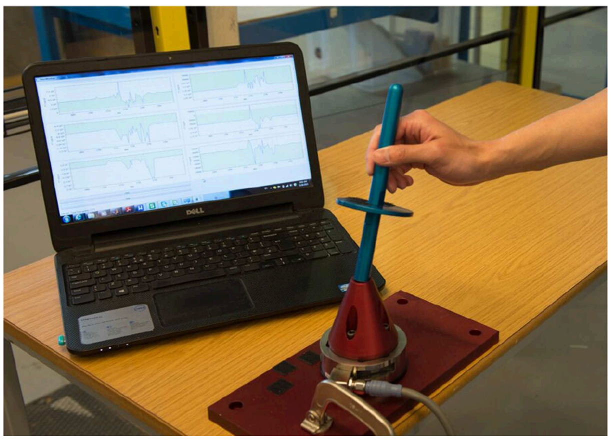


FIGURE 8 | The experiment setup during PiH insertion.

that the same mixing matrix ζ and the normalization coefficients of the training stage were used to pre-process test data; under the assumption that statistical properties of the test data are unknown, during the evaluation process. Then, the accuracy of the HMM model calculated with respect to the label data.

5. RESULTS AND DISCUSSION

The proposed approach was designed to recognize CSs during the assembly task efficiently, and then it was evaluated using PiH insertion problem as discussed before. Next, fitted models were evaluated as described in section 4. At the beginning, the collected during PiH insertion were six-dimensional ($\mathbf{X} \in \mathbb{R}^6$) as shown in **Figure 9**, while the transformed data is two-dimensional ($\mathbf{X}_{red} \in \mathbb{R}^2$) as illustrated in **Figure 10**, which indicates the PCA selected features from the raw data in **Figure 9**. The resulting PCA components were signals that have an accumulative variance that is higher than 90% of the total variance. The selected features were segmented using the modified PAA and K-means. The modified PAA and symbolic representations of the time-series data are shown in **Figure 11**. **Figure 12** depicts the symbolic representation results based on the PAA segmentations. **Figure 13** illustrates the K-means segmentations and the corresponding symbolic representation, where each color represents a segment.

In order to compare the segmentation approaches (PAA and K-means) and to determine the suitable number of segments for each segmentation approach, the symbolic representation was carried out based on PAA and K-means separately with a different number of segments. A critical difference between the PAA and the K-means segmentation is that the temporal and spatial features are crucial for the K-means segmentation. In contrast, PAA splits data into segments of equal length (temporal length) without taking spatial data into account. After that, temporal knowledge can be captured using HMM.

Figure 14 shows the accuracy of the HMM model based on PAA segmentation. The highest accuracy is 94% using 30 segments with 0.88 s computational time. In comparison **Figure 15** illustrates the accuracy of the HMM model based on K-means segmentation. The highest accuracy is 95% using 10 segments with 11.86 s. Those results indicate that models generated based on K-means segmentation do not require a large number of segments to achieve high accuracy. The models created using PAA require a large number of segments to improve the accuracy of the model. The model based on K-means segmentation achieved higher accuracy with a lower number of segments. This requires an extensive search until it converges to the optimal segmentation with resulting segmentation depending on the initial estimation of the segments' centroids. Surprisingly, the accuracy decreased dramatically with an increased number of segments. This shows there is no linear relationship between the number of segments and the accuracy. Therefore, an optimal number of segments needs to be identified requiring an additional iterative process. Conversely, the models generated using the PAA are more robust and do not request an iterative search. Also, the PAA segmentation returns the same segments for the same trial repeatedly. The results presented so far correspond to the data collected during the insertion of *Peg 1* without considering the variation in clearance.

Another important aspect in the PiH assembly process is the clearance, where assembly of tight clearance parts is more difficult than loose clearance parts. In order to test the models for different clearances, two models; *model 1* and *model 2*, were trained separately using the sequences captured during the assembly of *Peg 1* (tight) and *Peg 2* (loose), respectively (see section 4). Both the models were tested to explore the relationship between the accuracy of CS recognition and the clearances.

To evaluate the classification accuracy of the two models both models were tested with unseen labeled data (for assembling *Peg 1* and *Peg 2*). The resulting accuracy is shown in the confusion matrices in **Tables 1**, **2**. **Table 1** shows the confusion matrix of the HMM trained using the PAA with 30 segments (*model 1*). It can be observed from the table that the *CC* stage is being the least accurately classified.

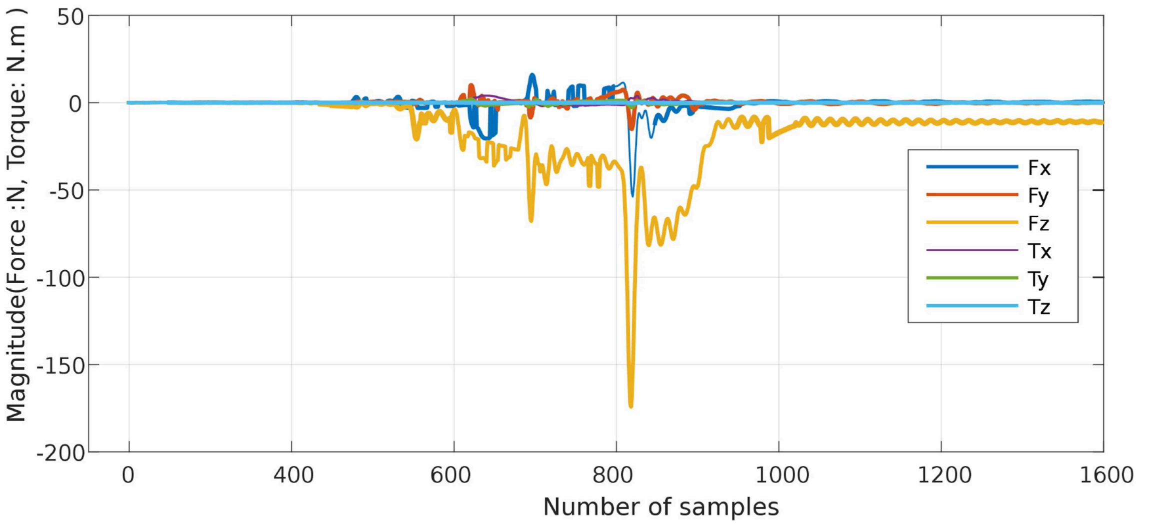


FIGURE 9 | Six-dimensional F/T signal during PiH assembly (original input data $\mathbb{R}^6$).

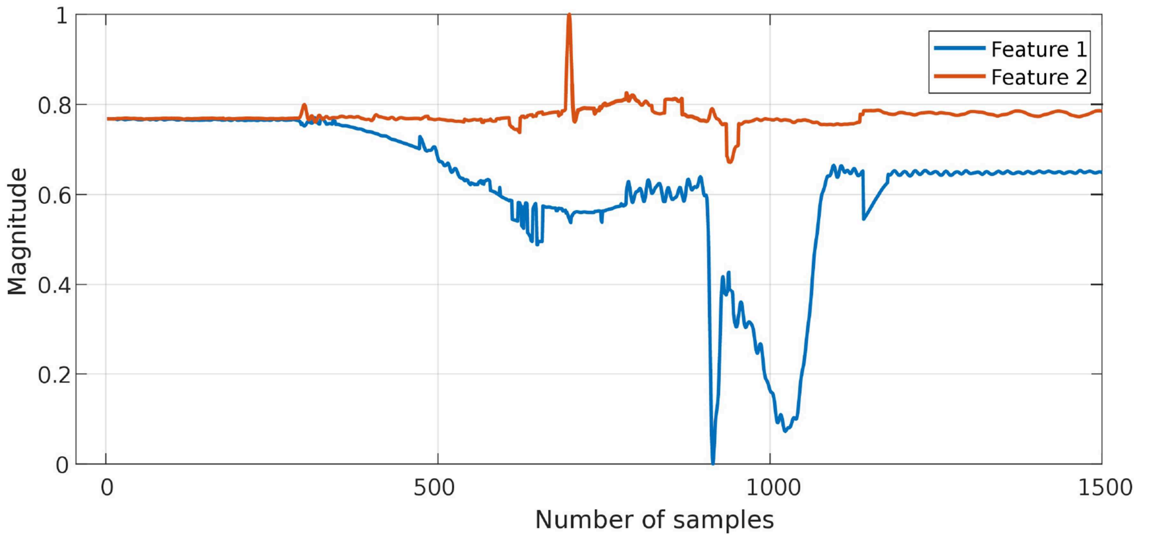


FIGURE 10 | The transformed F/T data ($\mathbb{R}^2$) after features transformation in the latent space using PCA. The accumulative-variance threshold was 90% of the total variance of all signals.

Table 2 shows the confusion matrix of the HMM trained using the PAA with 30 segments (*model 2*). An analysis of the result reveals that the misclassification of the *CC* stage that happens due to the static friction that occurs directly after the first contact. Also, the force level during this stage overlaps with the force level at the full-insertion stage which means that the mapping process will assign the same symbols for both stages (*CC* and *FI*).

The overall accuracy of *model 1* and *model 2* are 94 and 64%, respectively. Therefore, the trained models derived from the insertion of the larger clearance peg have a lower accuracy than the model based on the tighter clearance peg. The reason behind this is that the tighter clearance creates a stronger boundary amongst the CSs. Nevertheless, parts with larger clearances can partially change their contact state without causing distinguishable variation in the F/T signal which makes the recognition of distinct CS more difficult.

Additionally, the model with higher accuracy *(model 1)* was used to recognize the assembly CSs of *Peg 2* to examine the robustness against clearance variation. The performance of CS recognition based on *model 1* is illustrated in the confusion matrix as shown in **Table 3**. The overall accuracy reduced from 96 to 82.4%. Though, the accuracy of *model 1* on *Peg 2* is still better than the accuracy of *model 2* on *Peg 2*, this shows that *model 1* is quite robust against clearance variation.

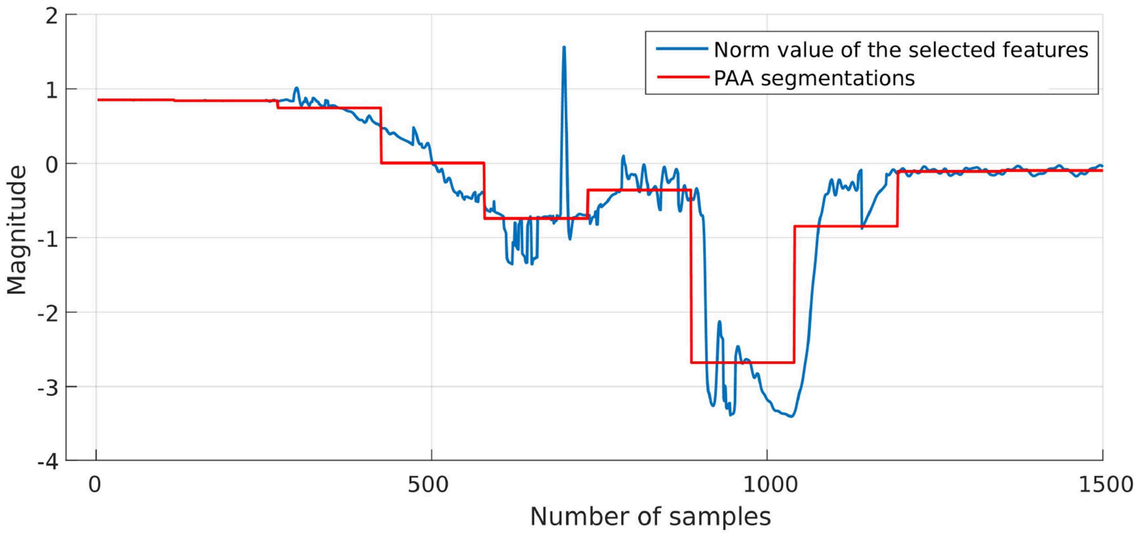


FIGURE 11 | PCA and corresponding PAA result.

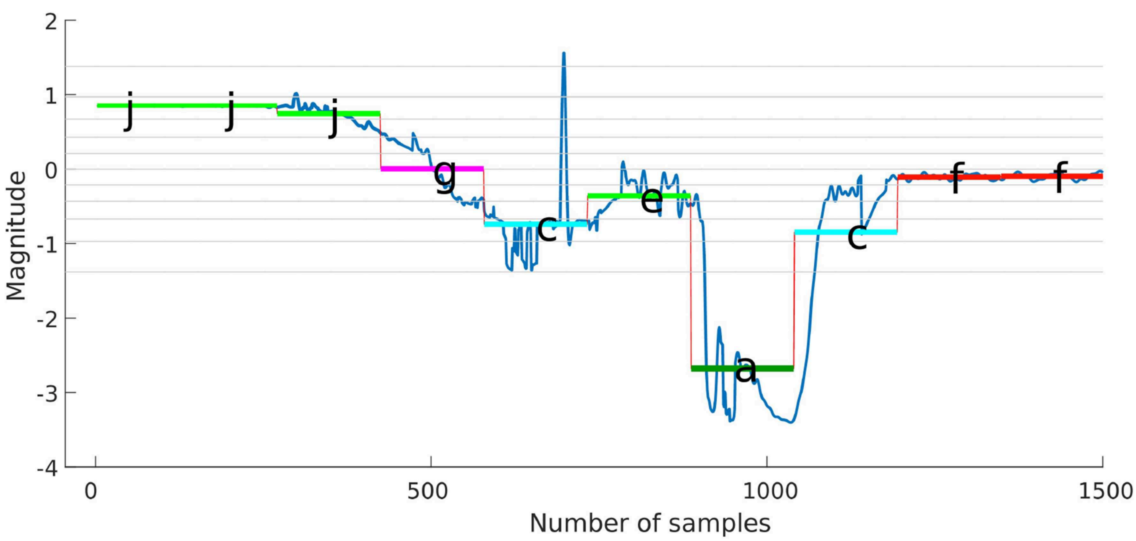


FIGURE 12 | Symbolic representation with five segments using PAA.

The results generated were compared with the most relevant work from the literature. In this regard, the method introduced by Jamali et al. (2014) achieved an overall accuracy of 81%, and 85% for rotation about the x-axis and the y-axis, respectively. The HMM-PAA models proposed in this paper has an accuracy of 94% and is, therefore, an improvement. However, to ensure that the accuracy is not due to chance, the datasets from all users for *Peg 1* and *Peg 2* have been combined and then randomly split 100 times into train and test data. The confusion matrices of the 100 times split using HMM-PAA and HMM-K-means are shown in **Tables 4**, **5**, respectively. The average accuracy of the HMM-PAA model is (90 $\pm$ 1.38)%, while it was only (76 $\pm$ 1.45)% for the HMM-K-means model. **Table 6** illustrates the overall accuracy, precision, and F-score of both HMM-PAA and HMM-K-means models. These numbers show better accuracy and robustness (precision) of the HMM-PAA in comparison with the HMM-K-means. The overall accuracy of the HMM-PAA was 90% with σ equals to 8.4%, while HMM-K-means has an accuracy of 76% with σ equals to 8.2%. This shows that the accuracy of both approaches has similar standard variation with different overall accuracy.

The proposed approach greatly reduces the required computation time, although it relies on multi-stage processes. **Table 7** shows the computational complexity of the proposed approach in comparison with three similar research approaches, namely Jamali et al. (2014), Jasim et al. (2017), and Hannaford

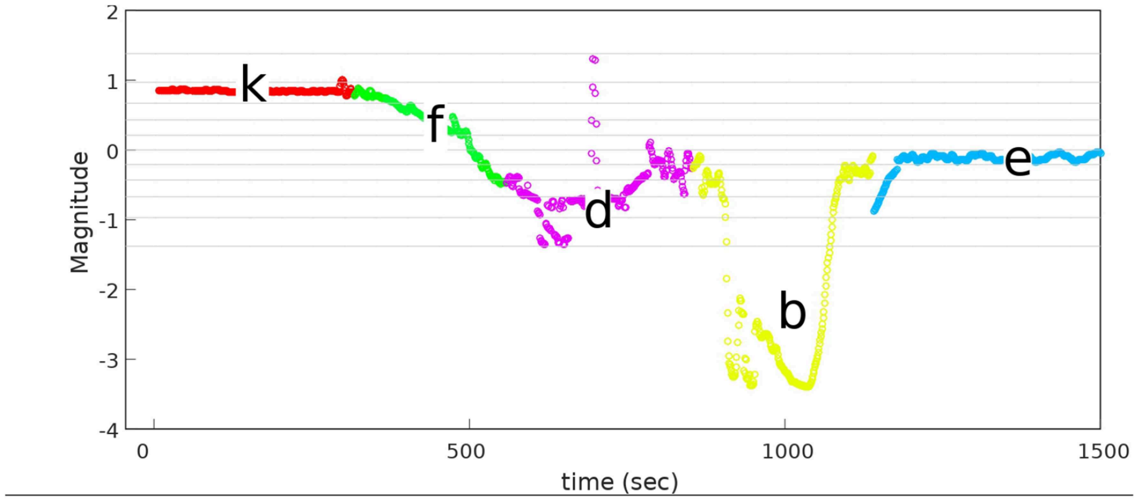


FIGURE 13 | Symbolic representation with five segments using K-means.

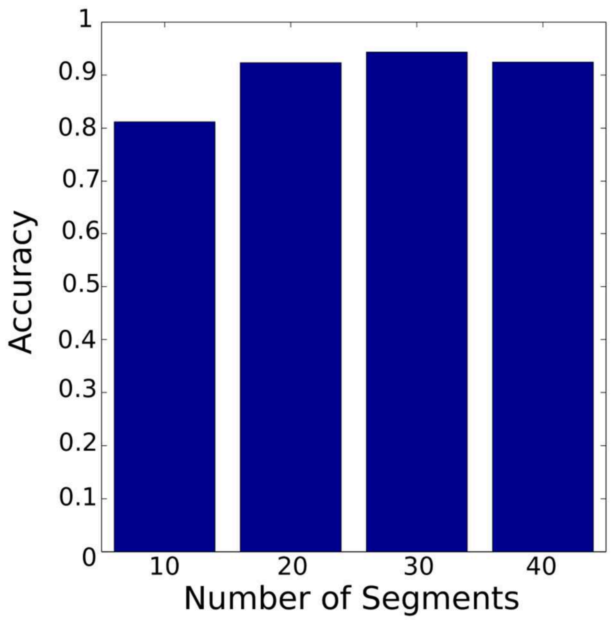


FIGURE 14 | Classifier accuracy with PAA segmentation using different number of segments. The best accuracy was achieved with 30 segments.

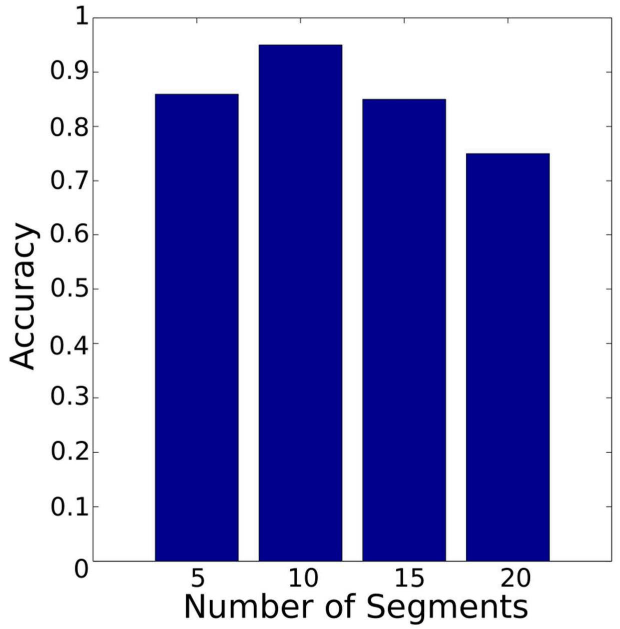


FIGURE 15 | Classifier accuracy with K-means segmentation using different number of segments. The best accuracy was achieved with 10 segments.

and Lee (1991), where $N_{symbols}$ is the number of symbols, K is the number of original dimensions before the PCA, M is the number of segments, N_{sample} is the number of samples within the time-series and D is the number of selected features (selected dimensions based on the PCA). For the proposed approach with PAA, the worst case scenario occurs when the $N_{symbols}$ is 12, and M is 30. In this case, the complexity of the HMM is the bottleneck; hence, the total complexity is $\mathcal{O}(2\,K\,N_{samples}\,D)$. On the other hand, the worst case for the proposed approach with K-means occurs when the $N_{symbols}$ is 12, and M is 10; however, the time complexity of the K-means is quadratic of the $N_{samples}$, which was on average 1500 samples. Henceforth, the K-means is the bottleneck for this case, which explains the long execution time to recognize the CS in comparison with PAA. In comparison with the method introduced in Jamali et al. (2014), the complexity of MML-GMM, that was used to cluster the Force/Torque data), was $\mathcal{O}(M\,N_{samples}\,D)$. Nevertheless, the complexity of

TABLE 1 | Confusion matrix of *model 1* for *Peg 1* clearance $c = 0.11\,mm$, where NC, no contact; CC, Chamfer-Crossing; I, insertion; FI, full insertion.

	NC	CC	I	FI
NC	83	7	0	0
CC	0	14	1	0
I	0	0	50	0
FI	0	0	0	78

TABLE 2 | Confusion matrix of *model 2* for *Peg 2* with clearance $c = 0.17\,mm$.

	NC	CC	I	FI
NC	82	5	0	12
CC	0	10	0	36
I	0	21	31	0
FI	0	0	12	31

TABLE 3 | Confusion matrix of *model 1* validated using observation of *Peg 2*.

	NC	CC	I	FI
NC	82	12	0	0
CC	0	9	1	0
I	0	21	31	0
FI	0	0	11	71

TABLE 4 | Confusion matrix of HMM trained with PAA 30 segments.

	NC	CC	I	FI
NC	1,115	90	0	0
CC	65	180	25	0
I	0	20	500	130
FI	0	0	15	1,160

the proposed approach was $\mathcal{O}(M)$, that only depends on M, while the complexity of MML-GMM depends on M multiplied by $N_{samples}$ and D. The overall performance appears similar in both methods. However, the method proposed in Jamali et al. (2014) requires additional exploration stage (set of random movements on x and y direction) before starting the recognition stage. Henceforth, it might require a longer time until it converges. In Jasim et al. (2017) the EM-GMM were utilized without dimensionality reduction, which means that the complexity is $\mathcal{O}(M\,N_{samples}\,D)$. While in the proposed approach the dimensionality reduction greatly reduced the number of features and the samples. Also, as shown in **Table 7** the total complexity of the proposed approach is $\mathcal{O}(2\,K\,N_{samples}\,D)$ which is less than the complexity of the EM-GMM utilized in Jasim et al. (2017) as long as $2\,K < M$. Finally, the computational complexity of the HMM presented by Hannaford and Lee (1990) was $\mathcal{O}(N^2_{samples}\,D)$, which is higher than the total complexity of the proposed approach.

TABLE 5 | Confusion matrix of HMM trained with K-means 10 segments.

	NC	CC	I	FI
NC	290	30	5	0
CC	90	60	10	0
I	0	55	135	65
FI	0	0	10	345

TABLE 6 | Overall accuracy (100 times split) of the HMM models with PAA and K-means.

Method		Accuracy (%)	Precision (%)	F-score (%)
HMM-PAA	μ	90	85	84
	σ	8.4	7.5	7.5
HMM-K-means	μ	76	73	72
	σ	8.2	7.5	7.3

6. CONCLUSIONS

This paper proposed a method to capture human skills during the PiH assembly process utilizing a learning algorithm to encode the assembly process. The proposed algorithm was based on a symbolic representation of F/T signals in the presence of geometrical variation of the assembled parts. This approach is capable of recognizing the CSs of PiH assembly process based on a symbolic representation of force and torque information. It can accommodate variations in the insertion force levels and compensate for process noise. The main benefits of this method are its simplicity and minimal pre-knowledge requirements about the geometrical information of the mating parts.

During the symbolic representation, two segmentation approaches, i.e., the K-means and the PAA, were investigated for their effectiveness. It was found that a higher accuracy of CS recognition can be achieved with a small number of segments when using K-means to segment the F/T time-series whereas the models trained based on the PAA segmentation require a higher number of segments. The model which was trained based on the K-means resulted in an accuracy of 70% with 10 segments with an 12 s computational time. The model generated based on the PAA resulted in an accuracy of 90% accuracy with 30 segments with 0.95 s computational time. The K-means requires more computational effort due to its iterative nature, whereas the PAA is a simpler and faster segmentation procedure. The use of the PAA in the symbolic representation reduces the required computational effort and increases the robustness of the model against process noise.

In this research, the robustness of the trained models was examined by varying part mating clearances. The results showed that the CS recognition is more accurate for tight clearance mating. This observation implies that there is an inverse relationship between the clearance and the accuracy of the CS recognition. This is due to the higher physical constraints

TABLE 7 | Computational complexity comparison.

	Proposed approach (PAA)	Proposed approach (K-means)	MML-GMM (Jamali et al., 2014)	EM-GMM (Jasim et al., 2017)	HMM (Hannaford and Lee, 1990)
PCA (after training)	$\mathcal{O}(2\,K\,N_{samples}\,D)$	$\mathcal{O}(2\,K\,N_{samples}\,D)$	$\mathcal{O}(2\,K\,N_{samples}\,D)$	–	–
GMM	–	–	$\mathcal{O}(M\,N_{samples}\,K)$	$\mathcal{O}(M\,N_{samples}\,D)$	–
PAA	$\mathcal{O}(N_{samples})$	–	–	–	–
K-means	–	$\mathcal{O}(N^2_{samples})$	–	–	–
HMM	$\mathcal{O}(M\,N^2_{symbols})$	$\mathcal{O}(M\,N^2_{symbols})$	$\mathcal{O}(M\,N^2_{symbols})$	–	$\mathcal{O}(D\,N^2_{samples})$
Total	$\mathcal{O}(2\,K\,N_{samples}\,D)$	$\mathcal{O}(N^2_{samples})$	$\mathcal{O}(2\,K\,N_{samples}\,D)$	$\mathcal{O}(M\,N_{samples}\,D)$	$\mathcal{O}(D\,N^2_{samples})$

in a tight clearance insertion process, providing a better-defined boundary that separates the consecutive CSs. The model trained based on tight clearances peg is more robust against geometrical variation.

The availability of robust and computational efficient representations is an essential precursor for imitation learning. The proposed approach achieves those two goals. However, it heavily relies on approximation and dimensionality reduction that might remove essential features from the force trend. Accordingly, the proposed approach might be not suitable for applications that require high accuracy, such as textile recognitions. Future work will consider the transformation of the trained models to an industrial robot by extending the proposed approach to a complete imitation learning framework. It is believed that humans often rely on visual perception to perform handling task. Hence, the proposed methods can be extended to include visual features that might improve the models' accuracy.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the manuscript/supplementary files.

AUTHOR CONTRIBUTIONS

AA-Y and YZ conducted the design, the adopted methodology, and the experiment. All authors conducted the data collection, data analysis, and results interpretation.

ACKNOWLEDGMENTS

The authors acknowledge support from the Intelligent Automation Centre/Loughborough University (previously: the EPSRC Centre for Innovative Manufacturing in Intelligent Automation), in undertaking this research work under grant reference number EP/IO33467/1.

REFERENCES

Calinon, S., and Billard, A. (2008). "A probabilistic programming by demonstration framework handling constraints in joint space and task space," in *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2008. IROS 2008* (Nice), 367–372. doi: 10.1109/IROS.2008.4650593

Chen, C.-W. (2011). Stability analysis and robustness design of nonlinear systems: an nn-based approach. *Appl. Soft Comput.* 11, 2735–2742. doi: 10.1016/j.asoc.2010.11.004

De la Torre, F., and Kanade, T. (2006). "Discriminative cluster analysis," in *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06 (New York, NY: ACM), 241–248.

Dong, S., and Naghdy, F. (2007). Application of hidden markov model to acquisition of manipulation skills from haptic rendered virtual environment. *Robot. Comput. Integr. Manuf.* 23, 351–360. doi: 10.1016/j.rcim.2006.02.010

Ernst, M. O., and Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415:429. doi: 10.1038/415429a

Han, J. (2005). *Data Mining: Concepts and Techniques*. San Francisco, CA: Morgan Kaufmann Publishers Inc.

Hannaford, B., and Lee, P. (1990). "Hidden markov model analysis of force/torque information in telemanipulation," in *Experimental Robotics I, Volume 139 of Lecture Notes in Control and Information Sciences*, eds V. Hayward, and O. Khatib (Berlin; Heidelberg: Springer Berlin Heidelberg), 135–149. doi: 10.1007/BFb0042517

Hannaford, B., and Lee, P. (1991). Hidden Markov model analysis of force/torque information in telemanipulation. *Int. J. Robot. Res.* 10, 528–539.

Hertkorn, K., Roa, M., Preusche, C., Borst, C., and Hirzinger, G. (2012). "Identification of contact formations: resolving ambiguous force torque information," in *2012 IEEE International Conference on Robotics and Automation (ICRA)* (Saint Paul, MN), 3278–3284. doi: 10.1109/ICRA.2012.6225148

Hochberg, M., Niles, L., Foote, J., and Silverman, H. (1991). "Hidden markov model/neural network training techniques for connected alphadigit speech recognition," in *[Proceedings] ICASSP 91: 1991 International Conference on Acoustics, Speech, and Signal Processing* (Toronto: IEEE), 109–112.

Jakovljevic, Z., Petrovic, P. B., and Hodolic, J. (2012). Contact states recognition in robotic part mating based on support vector machines. *Int. J. Adv. Manuf. Technol.* 59, 377–395. doi: 10.1007/s00170-011-3501-5

Jamali, N., Kormushev, P., and Caldwell, D. (2014). "Robot-object contact perception using symbolic temporal pattern learning," in *2014 IEEE International Conference on Robotics and Automation (ICRA)* (Hong Kong), 6542–6548.

Jamali, N., Kormushev, P., Vinas, A. C., Carreras, M., and Caldwell, D. G. (2015). "Underwater robot-object contact perception using machine learning on force/torque sensor feedback," in *2015 IEEE International Conference on Robotics and Automation (ICRA)* (Seattle, WA: IEEE), 3915–3920.

Jasim, I. F., Plapper, P. W., and Voos, H. (2017). Contact-state modelling in force-controlled robotic peg-in-hole assembly processes of flexible objects using optimised Gaussian mixtures. *Proc. Inst. Mech. Eng. Part B: J. Eng. Manufacture* 231, 1448–1463. doi: 10.1177/0954405415598945

Keogh, E., Lin, J., and Fu, A. (2005). "Hot sax: efficiently finding the most unusual time series subsequence," in *Fifth IEEE International Conference on Data Mining* (Houston, TX), 8. doi: 10.1109/ICDM.2005.79

Kronander, K., Burdet, E., and Billard, A. (2014). "Task transfer via collaborative manipulation for insertion assembly," in *Workshop on Human-Robot Interaction for Industrial Manufacturing, Robotics, Science and Systems*. Berkeley.

Kwiatkowska, M., Norman, G., and Parker, D. (2004). Probabilistic symbolic model checking with prism: a hybrid approach. *Int. J. Softw. Tools Technol. Transf.* 6, 128–142. doi: 10.1007/s10009-004-0140-2

Lau, H. (2003). A hidden markov model-based assembly contact recognition system. *Mechatronics* 13, 1001–1023. doi: 10.1016/S0957-4158(03)00012-6

Lin, J., Keogh, E., Lonardi, S., and Chiu, B. (2003). "A symbolic representation of time series, with implications for streaming algorithms," in *Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery* (San Diego, CA: ACM), 2–11.

Lin, J., Keogh, E., Wei, L., and Lonardi, S. (2007). Experiencing sax: a novel symbolic representation of time series. *Data Mining Knowl. Discov.* 15, 107–144. doi: 10.1007/s10618-007-0064-z

Liu, H., and Motoda, H. (1998). *Feature Extraction, Construction and Selection: A Data Mining Perspective*, Vol. 453. New York, NY: Springer Science & Business Media.

Mikami, K., Okuda, H., Taguchi, S., Tazaki, Y., and Suzuki, T. (2010). "Model predictive assisting control of vehicle following task based on driver model," in *2010 IEEE International Conference on Control Applications* (Yokohama), 890–895.

Mohammad, Y., and Nishida, T. (2014). "Robust learning from demonstrations using multidimensional sax," in *2014 14th International Conference on Control, Automation and Systems (ICCAS)* (Seoul), 64–71. doi: 10.1109/ICCAS.2014.6987960

Nakabayashi, A., Ukai, S., Wada, H., and Ohtani, T. (2013). "A bayesian robust identification method for piecewise affine autoregressive exogenous model from outlier-contaminated data," in *2013 Proceedings of SICE Annual Conference (SICE)* (Nagoya: IEEE), 511–518.

Nejati, N., and Könik, T. (2009). "Probabilistic relational learning of human behavior models," in *Human Behavior Modeling, Papers from the 2009 AAAI Spring Symposium, Technical Report SS-09-04* (Stanford, CA), March 23–25, 2009, 62–67.

Nuttin, M. (1995). "Learning approaches to contact estimation in assembly tasks with robots," in *Proc. 3rd Eur. Workshop Learning Robots* (Heraklion).

Okuda, H., Takeuchi, H., Inagaki, S., and Suzuki, T. (2008). "Modeling and analysis of peg-in-hole task based on mode segmentation," in *SICE Annual Conference, 2008* (Tokyo), 1595–1600. doi: 10.1109/SICE.2008.4654916

Park, K.-T., Shin, Y.-J., Park, C.-H., Mo, Y.-S., Jeong, D.-C., and Shin, Y. S. (2008). "Robot application for assembly process of engine part," in *International Conference on Control, Automation and Systems, 2008. ICCAS 2008* (Seoul), 2107–2110. doi: 10.1109/ICCAS.2008.4694443

Rozo, L., Jiménez, P., and Torras, C. (2013). A robot learning from demonstration framework to perform force-based manipulation tasks. *Intell. Serv. Robot.* 6, 33–51. doi: 10.1007/s11370-012-0128-9

Sophian, A., Tian, G. Y., Taylor, D., and Rudlin, J. (2003). A feature extraction technique based on principal component analysis for pulsed eddy current {NDT}. *{NDT} & E Int.* 36, 37–41. doi: 10.1016/S0963-8695(02)00069-5

Vlachos, M., Lin, J., Keogh, E., and Gunopulos, D. (2003). "A wavelet-based anytime algorithm for k-means clustering of time series," in *Workshop on Clustering High Dimensionality Data and Its Applications* (San Francisco, CA: Citeseer).

Wallace, C., and Dowe, D. (2000). Mml clustering of multi-state, poisson, von mises circular and gaussian distributions. *Stat. Comput.* 10, 73–83. doi: 10.1023/A:1008992619036

Wen, S., Zhu, J., Li, X., and Chen, S. (2014). A double-loop structure in the adaptive generalized predictive control algorithm for control of robot end-point contact force. *{ISA} Trans.* 53, 1603–1608. doi: 10.1016/j.isatra.2014.05.024

Whitney, D. (2004). *Mechanical Assembly: Their Design, Manufacture and the Role in Product Development*. New York, NY: Oxford University Press.

Wu, B., Yan, Q., Luo, J., and Wu, Z. (2014). Signal processing and application of six-axis force/torque sensor integrated in humanoid robot foot. *J. Signal Process. Syst.* 74, 263–271. doi: 10.1007/s11265-013-0796-5

Conflict of Interest: YZ was employed by Beijing Ewaybot Technology LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 10 January 2020
doi: 10.3389/frobt.2019.00148

Check for updates

Robot-Supported Collaborative Learning (RSCL): Social Robots as Teaching Assistants for Higher Education Small Group Facilitation

Rinat B. Rosenberg-Kima*[1], *Yaacov Koren*[2] *and Goren Gordon*[3,4]

[1] Faculty of Education in Science and Technology, Technion—Israel Institute of Technology, Haifa, Israel, [2] Department of Sociology and Anthropology, Tel-Aviv University, Tel-Aviv, Israel, [3] Curiosity Lab, Department of Industrial Engineering, Tel-Aviv University, Tel-Aviv, Israel, [4] Sagol School of Neuroscience, Tel-Aviv University, Tel-Aviv, Israel

OPEN ACCESS

Edited by:
*Agnese Augello,
Institute for High Performance Computing and Networking (ICAR), Italy*

Reviewed by:
*Sofia Serholt,
University of Gothenburg, Sweden
Birgit Lugrin,
Julius Maximilian University of Würzburg, Germany*

***Correspondence:**
*Goren Gordon
goren@gorengordon.com*

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *29 September 2019*
Accepted: *13 December 2019*
Published: *10 January 2020*

Citation:
Rosenberg-Kima RB, Koren Y and Gordon G (2020) Robot-Supported Collaborative Learning (RSCL): Social Robots as Teaching Assistants for Higher Education Small Group Facilitation. Front. Robot. AI 6:148. doi: 10.3389/frobt.2019.00148

Acknowledging the benefits of active learning and the importance of collaboration skills, the higher education system has started to transform toward utilization of group activities into lecture hall culture. In this study, a novel interaction has been introduced, wherein a social robot facilitated a small collaborative group activity of students in higher education. Thirty-six students completed a 3 h activity that covered the main content of a course in Human Computer Interaction. In this within-subject study, the students worked in groups of four on three activities, moving between three conditions: instructor facilitation of several groups using pen and paper for the activity; tablets facilitation, also used for the activity; and robot facilitation, using tablets for the activity. The robot facilitated the activity by introducing the different tasks, ensuring proper time management, and encouraging discussion among the students. This study examined the effects of facilitation type on attitudes toward the activity facilitation, the group activity, and the robot, using quantitative, and qualitative measures. Overall students perceived the robot positively, as friendly and responsive, even though the robot did not directly respond to the students' verbal communications. While most survey items did not convey significant differences between the robot, tablet, or instructor, we found significant correlations between perceptions of the robot, and attitudes toward the activity facilitation, and the group activity. Qualitative data revealed the drawbacks and benefits of the robot, as well as its relative perceived advantages over a human facilitator, such as better time management, objectivity, and efficiency. These results suggest that the robot's complementary characteristics enable a higher quality learning environment, that corresponds with students' requirements and that a Robot Supportive Collaborative Learning (RSCL) is a promising novel paradigm for higher education.

Keywords: social-robots, human-robot interaction, collaborative learning, active learning, educational technology, higher education

INTRODUCTION

Classrooms in the twenty-first century are slowly being transformed from frontal lectures halls filled with passive students, to collaborative small groups actively participating in project based learning (Helle et al., 2006; Kokotsaki et al., 2016). Studies have shown that such active participatory learning is more effective in content retention (Al-Balushi and Al-Aamri, 2014) and engagement (Fernandes et al., 2014). Thus, emphasis has been diverted to so-called twenty-first century skills (Crane, 2003; Saavedra and Opfer, 2012; Trilling and Fadel, 2012), with focus on the 4C super-skills, i.e., communication, collaboration, creativity, and critical thinking (Shulman, 1986; Kivunja, 2015). This focus has created several new pedagogies, such as to provide students with the opportunity, within the classroom, to observe, imitate, and practice *critical agency*, and reflect upon it (ten Dam and Volman, 2004); *collaborate* by learning to share tasks and resources and be responsible for their tasks (Lai et al., 2017); engage in inter-, trans-, and cross-disciplinary approaches to promote *creativity* (Harris and de Bruin, 2018); and use project-based learning as the basis for improving *communication* skills (Saenab et al., 2018).

In higher education, the proliferation of massive online open courses (MOOCs) (Bozkurt et al., 2017) has not lived up to its initial expectation (Khalil and Ebner, 2014; Thomas and Thorpe, 2019). However, the emergence of the "flipped classroom" paradigm (Gilboy et al., 2015; Schmidt and Ralph, 2016), in which students learn the material at home via on-line learning platforms and then discuss and practice it in small groups in the classroom, has been shown to be highly effective (Chen and Chen, 2015; Thomas and Thorpe, 2019).

These paradigms have started to reshape the role of the lecturer in higher education, wherein the role of group facilitator has become an important aspect of teaching in such scenarios (Franco and Nielsen, 2018). Group facilitation involves the mediation of the material via encouragement of communication, active participation, and discussion of all the group members (Phillips and Phillips, 1993). Best practices involve promotion of reflection and action (Franco and Nielsen, 2018) and maintaining engagement density (Matsuyama et al., 2015).

These changes to classic teaching methods have also introduced new challenges as large classrooms, restructured as several small discussion groups, demand the attention of the lecturer, and her TAs (Moust and Schmidt, 1994). While on-line discussion forums have prospered in recent years (Pendry and Salvatore, 2015; Yang et al., 2015; Chiu and Hew, 2018), with AI assisting in managing such forums (Goel and Joyner, 2017), studies have shown that personal face-to-face interactions and discussions in small groups have their advantages (Chen and Chen, 2015; Thomas and Thorpe, 2019). The question of scaling-up group facilitation is thus of prominent importance.

Concurrently, social robots have progressed drastically in the last decade, especially in the field of education (Mubin et al., 2013; Brown and Howard, 2014; Gordon et al., 2015; Belpaeme et al., 2018b). Compared to tablets and screens, social robots have been shown to convey more learning gains (Wainer et al., 2006; Leyzberg et al., 2012; Li, 2015; Luria et al., 2017) and evoke more emotional expressions (Spaulding et al., 2016). They have been used to teach science (Shiomi et al., 2015), math (Brown and Howard, 2014), languages (Kory and Breazeal, 2014; Belpaeme et al., 2015; Hein and Nathan-Roberts, 2018), and even nutrition (Short et al., 2014). Moreover, they have been used to promote meta-cognitive skills such as curiosity (Gordon et al., 2015; Ceha et al., 2019) and growth mindset (Park et al., 2017). Social robots in education have taken different roles. They have been used as peers or companions in learning with the students (Okita et al., 2009), or tutors in which the robot teaches students (Belpaeme et al., 2018b). Moreover, social robots have been used as teachers using frontal lecture mode (Sisman et al., 2018), one-on-one interaction (Short et al., 2014; Gordon et al., 2015) and even in two-person dialogues (Tahir et al., 2014). Several studies have addressed how a single robot can interact with small groups of children (Leite et al., 2015; Strohkorb et al., 2015), elderly (Matsuyama et al., 2008), and adults (Matsuyama et al., 2015). More specifically, several studies examined possible roles of social robots in group interaction (Jung et al., 2015; Shen et al., 2018; Alves-Oliveira et al., 2019; Correia et al., 2019; Oliveira et al., 2019).

These advances in social robots resulted in their slow introduction into the educational system (Belpaeme et al., 2018a; Kory-Westlund and Breazeal, 2019) and into homes (Scassellati et al., 2018). Many studies have focused on young children, from preschoolers (Kory and Breazeal, 2014), through elementary school (Leite et al., 2015), and adolescents (Björling et al., 2019), with special interest in children with Autism (Scassellati et al., 2018). In recent years, several applications of social robots in higher education have started to emerge (Brown and Howard, 2014; Edwards et al., 2016; Deublein et al., 2018). Pfeifer and Lugrin (2018) showed that a female robot can lead to better learning in female students while breaking stereotypical beliefs. Rosenberg-Kima et al. (2019) showed that social robots can serve as teaching assistants by answering simple questions of students working in small groups.

In this contribution we report on a higher education application of social robots as small group facilitators. Our goal was to compare the current state, in which an instructor attempts to facilitate several groups in the classroom, to a robot facilitator that is more limited in terms of emotional and cognitive capabilities yet remains with the group for the entire activity to facilitate it. An undergraduate course group activity that summarizes the material taught during a full semester has been converted into an interaction facilitated by a social robot, Nao, and mediated by tablets. Groups of four students performed the group activity, followed the instructions of the robot facilitator, discussed the material, and then answered questionnaires about the interaction. The same groups performed similar activities with tablets alone and with pen-and-paper, facilitated by the instructor of the course (within-subject design). Their impressions of the different activities' modalities are reported.

TABLE 1 | Description of the goals of each task in the Human-Computer Interaction course activities, where the overall goal was to develop a family-oriented App.

Activity	Overall goals of each sub-task
Activity 1: target audience	1.1. Defining the target audience of the application. The students were instructed to first work individually for 2 min and then combine the group members' lists into one list of two target audiences. 1.2. Building a survey: (a) given questions, identify different type of questions (e.g., questions appropriate for online survey, questions appropriate for focus groups, questions that might evoke confirmation bias etc). (b) Select five questions that fit the target audience of the application.
Activity 2: metaphors	2.1. Defining metaphors for the application. Again, the students were instructed to first work individually for 2 min and then combine the group members' lists into one list of two metaphors. 2.2. Screens: given a screenshot of an application, identify design features (e.g., centered, direct instruction, etc.)
Activity 3: interfaces	3.1. Defining interfaces for the application. Again, the students were instructed to first work individually for 2 min and then combine the group members' lists into one list of two interfaces. 3.2. Evaluating screens: rate two screens on a 1–5 rating scale on four given heuristics.

METHOD

Participants

Thirty-six students (age $M = 28$ years, SD = 0.3, 58.3% females) who participated in the course *Human Computer Interaction* completed a three-parts activity that covered the main content of the course and served as preparation for the final exam. The students consented to include their participation data in the study. The study was approved by the IRB.

Materials

All the participants completed three group-linked activities, each covering different content of the course, and serving as training for the final exam. The students worked collaboratively in groups of four students (nine groups in total). The overall goal of the activities was to design a family App that aims to provide all the needed information and tools to support family communication and planning (e.g., weekly schedule, messages, budget planning, etc.), while enabling each member of the family to be an active participant. Each activity lasted about 30 min. The specific goals of each activity are described in **Table 1**.

Conditions

The study had a within-subject design, wherein the students worked in groups of four, each group going through three conditions (**Figure 1**):

1. Robot condition: In this condition, each group of students performed the task using tablets and were facilitated by a social robot. Each group was located in a separate room.
2. Tablet condition: In this condition, the groups of students performed the task and were facilitated by tablets, located in a large lecture hall.
3. Instructor condition: In this condition, the groups of students performed the task using pen and paper. All the groups were located in a large lecture hall and were facilitated by a single instructor.

The order of the conditions differed between the groups, but the order of the activities with respect to the task itself was the same, as each activity was building on the previous one.

Each of the nine groups completed the three activities and experiences all three conditions. Thus, for example, the first group completed the first activity with a robot-facilitation (robot + tablets), then moved to a different room where it completed the second activity with a tablet-facilitation (tablets only), and lastly moved to a different room where it completed the third activity with paper based instruction, and an instructor-facilitation. The sequence of conditions varied between the groups to control for activity and order effect (see **Table 2** for a complete sequence of all the groups).

Setup Architecture

The setup architecture of the social-robot facilitator (robotator) condition included communication between four students, four android tablets (one per student), and a NAO robot (see **Figure 1A**). Unfortunately, state-of-the-art Natural Language Processing (NLP) could not support verbal communication facilitation of a group at this level. Hence, the robot spoke to the students utilizing pre-recorded sentences, but in order to establish bidirectional communication, the tablets served as additional sources of input and output between the robot and students. To implement this architecture, we used Python and Kivy to develop the tablet application, and ROS (Robot Operating Systems) and Python to control and manage the communication between the Robot and the tablets. The robotator facilitated this interaction by introducing the different tasks, managing the time (e.g., the robot said in relation to the design App task: "take 2 min to list different target audiences for the App, and then create a combined list with two target audiences," after which a timer of 2 min started followed by the next set of instructions), and encouraging discussion between the students (e.g., if two students answered the same question differently, the robot would say "I see that your answers are different, would you like to discuss that?").

The setup architecture of the tablets-only condition included communication between four students, each with one tablet (see **Figure 1B**). Python and Kivy were used to develop the tablet application, that included presentation of the tasks, guidelines, and time management via a presented timer.

The setup for the instructor condition included exchange of ideas between four students who worked with paper-based instructions that included the exact same instructions as in the tablet and robot-tablet conditions, but did not include support such as a timer. A human instructor was present in the classroom to answer questions of all the groups in this condition.

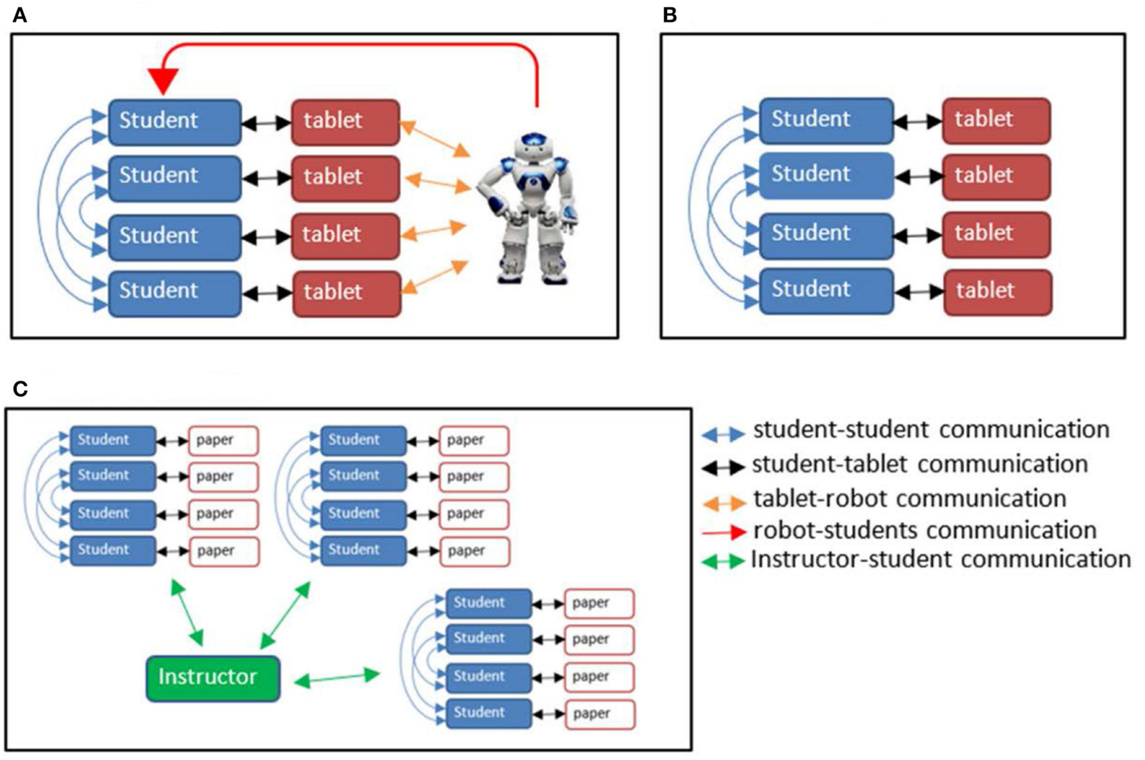


FIGURE 1 | Setup architecture of the three conditions used in the study. **(A)** Robot condition. **(B)** Tablet condition. **(C)** Instructor condition.

TABLE 2 | Sequence of activity and conditions for each group.

Groups	Activity 1	Activity 2	Activity 3
1–3	Robot	Tablets	Instructor
4–6	Instructor	Robot	Tablets
7–9	Tablets	Instructor	Robot

Measures

Attitude Toward the Robot\Tablet\Instructor Questionnaire

After each activity, students completed a 13-items questionnaire to evaluate their attitudes toward the robot\tablet\instructor depending on the condition. Students responded to a series of statements on a 5-point Likert-type scale (from 1 = Strongly disagree to 5 = Strongly agree) (e.g., "I trusted the information given by the robot\tablet\instructor"; see **Table 3** for the complete list). The questionnaire items were combined to form the attitudes toward the facilitation scale (Cronbach's alpha = 0.881).

Attitude Toward the Group Questionnaire

After each activity, students completed a 14-items questionnaire to evaluate their attitudes toward the group. Students responded to a series of statements on a 5-point Likert-type scale (from 1 = Strongly disagree to 5 = Strongly agree). The questionnaire resulted in two subscales: attitudes toward the specific group activity scale (e.g., "The group work contributed to understanding the content"; see **Table 3** for the complete list), which included items 1–10 excluding item 8 (Cronbach's alpha = 0.816), and attitudes toward group activities scale (e.g., "Group activities like this, are superior to individual activities"), which included items 11–14 (Cronbach's alpha = 0.743).

Godspeed Questionnaires

After the robot-facilitated activity, the students completed the 24-items Godspeed questionnaire, in which students responded to pairs of words and rated the robot on a 5-point semantic differential scale (e.g., Unfriendly-Friendly, Ignorant-Knowledgeable), resulting in 5 subscales: (I) Anthropomorphism consisting of 5 items (in this study Cronbach's alpha = 0.686), (II) Animacy consisting of six items (in this study Cronbach's alpha = 0.728), (III) Likable consisting of five items (in this study Cronbach's alpha = 0.867), (IV) Perceived Intelligence consisting of five items (in this study Cronbach's alpha = 0.845), and (V) Perceived Emotional safety (e.g., anxious vs. relaxed) consisting of three items (in this study Cronbach's alpha = 0.786) (Bartneck et al., 2009).

Qualitative Data

A semi open-ended questionnaire was used to collect qualitative data. The participants were asked to specify, in writing, three advantages, and three disadvantages the robot had as a facilitator

TABLE 3 | Results for attitudes toward the activity facilitation questionnaire (A), attitudes toward the group activity questionnaire (B), and Godspeed questionnaires (C).

	Statement	Robot		Tablet		Instructor		*p**	η^2
		M	SD	M	SD	M	SD		
(A) Attitudes toward the facilitation questionnaire	1. I understood the robot\app\Instructor	3.62	1.15	4.14	0.91	3.71	1.01	0.010	0.163
	2. The facilitation of the r\a\t was of high quality	3.47	1.08	3.59	0.94	3.42	0.85	0.083	0.221
	3. I trusted the information given by the robot\app\Instr.	3.71	1.06	4.18	0.87	3.93	0.87	0.023	0.145
	4. I felt comfortable with the robot\app\Instr. presence	4.09	0.83	3.97	0.72	3.78	0.81	0.586	0.040
	5. I felt comfortable with the behavior of the robot\app\Instr.	3.79	0.98	4.00	0.89	3.93	0.99	0.767	0.013
	6. The robot\app\Instr. adjusted to the class	3.62	0.89	3.69	0.96	3.40	1.13	0.204	0.062
	7. I would like more activities with the robot\app\Instr.	3.32	1.30	3.42	1.09	3.29	1.19	0.475	0.028
	8. The robot\app\Instr. responded to the group	2.94	1.18	3.11	1.32	3.47	1.15	0.168	0.069
	9. The robot\app\Instr. was friendly	3.79	1.01	3.69	0.87	3.84	1.04	0.750	0.011
	10. The robot\app\Instr. behaved human-like	2.71[a]	0.87	2.37[a]	0.81	3.84[b]	1.18	**0.000***	0.544
	11. I liked the robot\app\Instr. facilitator	3.26	0.83	2.83	0.92	3.27	1.10	0.08	0.045
	12. The activity with the robot\app\Instr. was pleasant	3.91	0.83	3.49	0.85	3.65	0.91	0.177	0.064
	13. The activity with the robot\app\Instr. was interesting	3.94[a]	0.92	3.03[b]	0.98	3.23[b]	1.11	**0.001***	0.226
	Attitudes toward the facilitation scale (items 1–13)	3.55	0.62	3.51	0.65	3.66	0.63	0.685	0.046
(B) Attitudes toward the group activity questionnaire	1. The group work contributed to understanding of the content	3.86	1.06	3.97	0.76	4.15	0.67	0.637	0.30
	2. I felt like I expressed myself during the discussions.	3.86	0.73	3.88	0.68	4.20	0.52	0.752	0.013
	3. All group members equally contributed to the discussion	3.74	1.01	3.53	0.96	3.90	1.02	0.636	0.030
	4. The work instructions were clear	3.00	1.21	3.53	1.08	3.65	0.99	0.519	0.043
	5. The contribution of the robot\tablet\Instr. was big	2.94	1.28	3.27	1.28	3.18	0.63	0.362	0.076
	6. I felt that the group members considered my opinions	4.17	0.75	4.12	0.48	4.35	0.62	0.597	0.034
	7. The sequence of tasks was logic and clear	3.60	1.09	3.91	0.93	3.85	0.93	0.712	0.022
	8. One group member managed most of the discussion	2.14	0.77	2.12	0.77	2.10	0.91	0.609	0.032
	9. I enjoyed working with my group members	4.29	0.57	4.03	0.83	4.40	0.60	0.554	0.039
	10. The group members felt free to express different opinions	4.37	0.73	4.32	0.73	4.50	0.60	0.944	0.004
	11. Group activities like this contribute to meaningful learning	3.77	0.94	3.65	0.84	3.89	0.74	0.552	0.042
	12. Group activities like this are a waste of time	2.26	0.98	2.50	0.75	2.25	0.85	0.895	0.007
	13. Group activities like this are superior to individual activities	3.57	0.95	3.62	0.74	3.70	1.08	0.523	0.042
	14. Groups activities contributes more than frontal lectures	3.51	0.82	3.67	0.96	3.70	0.92	0.139	0.123
	Attitude toward the group activity scale (items 1–7 and 9,10)	3.76	0.61	3.82	0.52	4.06	0.39	0.638	0.020
	General attitudes toward group activities scale (items 11–14)	3.65	0.69	3.62	0.56	3.75	0.64	0.599	0.036
(C) Godspeed questionnaires	Godspeed I: anthropomorphism	2.51	0.66	–	–	–	–	–	
	Godspeed II: animacy	2.66	0.65	–	–	–	–	–	
	Godspeed III: likable	3.64	0.73	–	–	–	–	–	
	Godspeed IV: perceived intelligence	3.15	0.72	–	–	–	–	–	
	Godspeed V: perceived safety	4.12	0.78	–	–	–	–	–	

**Bonferroni adjusted alpha value of 0.002 (0.05/14) was used for the single items.*
Bold value of p indicates a significant difference (given the Bonferroni correction) between [a] and [b].

of student groups. The open-ended questionnaire served as a means to get the perspective of students in their own words to provide "depth, detail, and meaning at a very personal level of experience" (Patton, 2014, p. 24). Nevertheless, given the limitations of an open-ended questionnaire in writing (e.g., dependent on writing skills of respondents or the impossibility of extending responses), observational data, based on video recording of activity, and a video sample analysis was used as a supportive tool to capture the context (Bauer and Gaskell, 2000).

Procedure

After signing a consent form, the students were placed in groups of four students. The groups were then guided to the location of their first activity settings according to their conditions as described in **Table 2**. Thus, groups 1–3 were guided to three different rooms in which the robots and tablets setting was located, groups 4–6 were placed in groups in one big room, where each student received a tablet, and groups 7–9 were placed in groups in one big room, where each student received paper-based instructions and a human instructor was present to answer questions. After completing the first activity, which took about 30 min, the students completed the questionnaires for about 15 min and were then guided to the location of the second activity according to the conditions, completed the second activity, filled again the questionnaires and were guided in the same way to the third activity. Overall completing the three activities, filling the

FIGURE 2 | Footage from the study of the robot condition settings.

questionnaires after each activity, and changing locations took 3 h (see **Figure 2** for footage of the robot condition).

RESULTS

This study examined the effects of facilitation type (robot facilitation, tablet facilitation, and instructor facilitation) on attitudes toward the activity and attitudes toward the group activity using one-way within-subject ANCOVA with group order as a covariant to control for order and activity (in groups 1–3 the robot facilitated the first activity, in groups 4–6 the robot facilitated the second activity, and in groups 7–9 the robot facilitated the third activity). Overall, we did not find an effect for the group order. In addition, attitudes toward the robot were measured using the Godspeed questionnaires and were correlated to the attitudes toward the robot facilitation and the attitudes toward the robot group activity.

Preliminary data analysis included examination of missing data and outliers, verification of the equivalence of treatment groups, and tests for assumptions of the parametric statistics. Some of the students missed some of the items in which case they were omitted from the analysis in the relevant places. Shapiro–Wilk normality test was used to detect violation of the normal distribution assumption. Results indicated that several dependent measures were not normally distributed. Nevertheless, it was suggested that ANOVA is robust enough to moderate violations of this assumption (Blanca et al., 2017). The overall scales were normally distributed. In addition, Bonferroni correction was applied to adjust the alpha values: Bonferroni adjusted alpha value of 0.002 (0.05/14) was used for the single items.

Attitudes Toward the Activity Facilitation

Overall the students reported that the activity with the robot was pleasant and interesting and the overall mean for the attitudes toward the robot facilitation scale was 3.55 (±0.62) (see **Table 3A** and **Figure 3**). Results of the within-subject ANCOVA for item 10 ("the robot\tablet\instructor behaved human-like") indicated a significant within-subject effect [$F_{(2,52)} = 30.982$, $p < 0.001$]. *Post-hoc* Bonferroni tests revealed a significant difference between the instructor and the robot conditions ($p < 0.001$) and between the instructor and the tablet condition ($p < 0.001$). As expected, the participants rated the instructor as significantly more human-like than the robot and the tablet. There was no significant difference between the robot and the tablet condition. In addition, results of the ANCOVA for item 13 ("the activity with the robot\tablet\instructor was interesting") indicated a significant within-subject effect [$F_{(2,52)} = 7.576$, $p = 0.001$]. *Post-hoc* Bonferroni tests revealed a significant difference between the robot and the tablet conditions ($p = 0.004$) and between the robot and the instructor condition ($p = 0.023$). The participants rated the activity with the robot as significantly more interesting than the tablet and the instructor conditions. There was no significant difference between the instructor and the tablet condition. Nevertheless, for the rest of the items there was no significance difference between the robot, the tablet, and the human instructor.

Attitudes Toward the Collaborative Group Activity

Overall students rated the group activity positively (see **Table 3B**). Results of the within-subject ANCOVA yielded no significance effects. Attitudes toward the robot facilitator had an overall mean of 3.76 (±0.61) for the attitudes toward the current group activity scale and an overall mean of 3.65 (±0.69) for the attitudes toward general group activities scale.

Attitudes Toward the Robot

Godspeed questionnaire, consisting of five subscales, was used to measure the participants' attitudes toward the robot used in the study. On a 1–5 scale, overall the participants rated

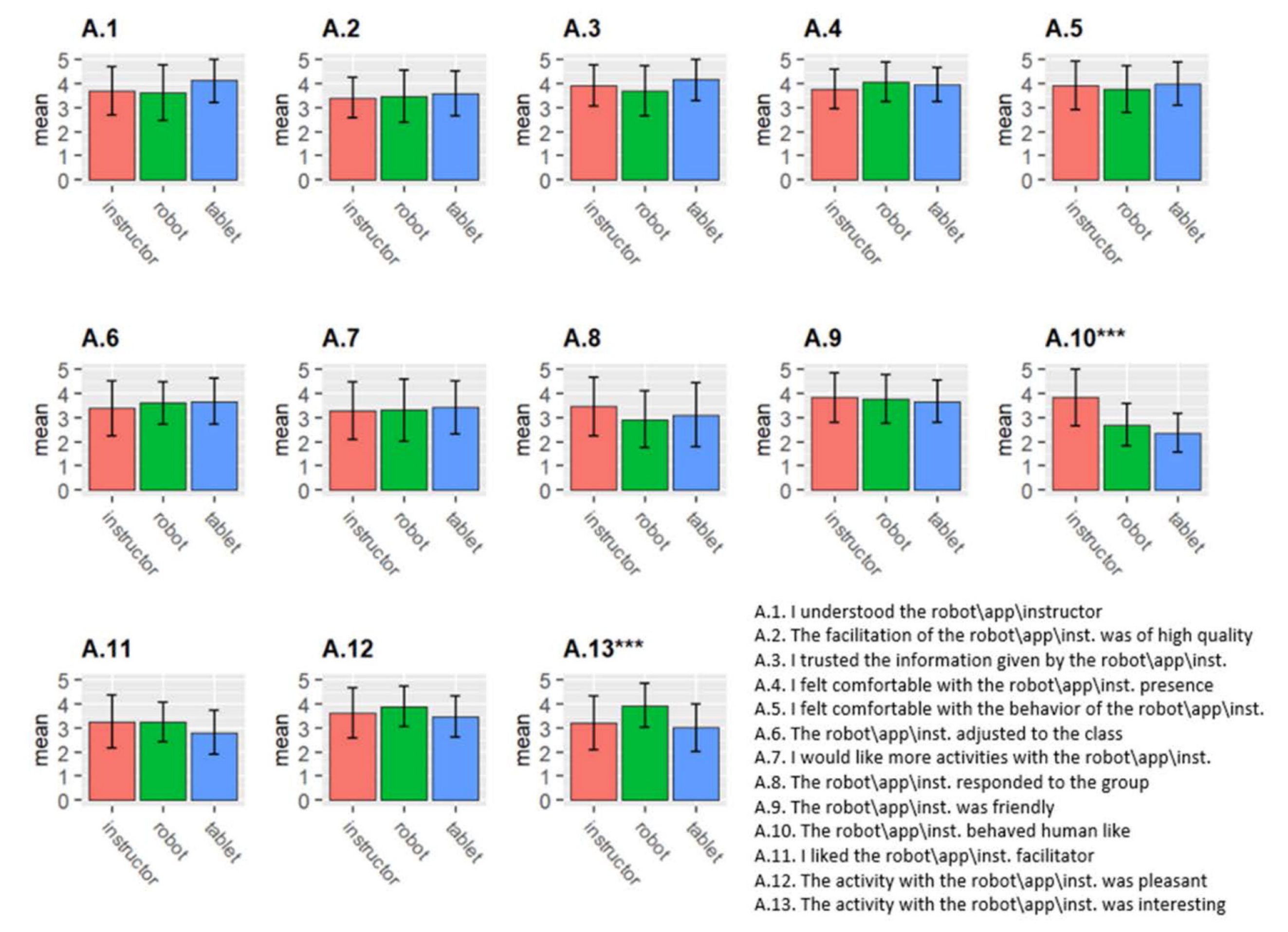


FIGURE 3 | Attitudes toward the facilitation (see **Table 3** for the complete list of statements). ***$p < 0.001$.

the robot 2.51 (±0.66) on anthropomorphism, 2.66 (±0.65) on animacy, 3.64 (±0.73) on likable, 3.15 (±0.72) on perceived intelligence, and 4.12 (±0.78) on perceived safety (see **Table 3C**). We were interested in finding what were the correlations between the Godspeed subscales and the three attitudes scales (attitudes toward the facilitation scale, the group activity scale, and group activities scale). With regard to the attitudes toward the facilitation scale, Pearson correlation tests indicated strong correlations between the scale and Anthropomorphism ($r = 0.632$, $p < 0.01$), Animacy ($r = 0.559$, $p < 0.01$), Likable ($r = 0.634$, $p < 0.01$), and Perceived Intelligence ($r = 0.595$, $p < 0.01$), but not with Perceived Safety ($r = -0.080$, $p = 0.655$). With regard to the attitudes toward the current group activity scale, only Perceived Intelligence of the robot was significantly correlated to the scale ($r = 0.429$, $p < 0.05$). With regard to attitudes toward general groups activities scale, none of Godspeed subscales was correlated to this scale (See **Figure 4**).

Qualitative Results

Thematic analysis method (Boyatzis, 1998) was used for analyzing and reporting themes within the data. The method is applicable to the research objective to report the ways individuals make meaning of their experience, on the one hand, and is not wedded to any pre-existing theoretical framework, on the other (Braun and Clarke, 2006). Following the template approach (Crabtree and Miller, 1998), and based on a preliminary scanning of the data, 157 students' statements were classified by two research team members to four principle categories: (1) Technical Functionality Benefits. (2) Social and Psychological Benefits. (3) Technical Functionality Drawbacks. (4) Social and Psychological Drawbacks. Within each category, statements were re-reviewed, collating statements into relevant themes. The analysis according to the aforementioned coding resulted in a total of 48% statements expressing benefits the robot had as a facilitator of student groups vs. 52% statements expressing drawbacks. Results analysis to the semi-open question indicated that students were well-attentive to the interaction with the robot (supported by video recording analysis), to its benefits as well as to the drawbacks of using a robot as a group facilitator. Excluding apparent novelty effect statements revealed that they were more concerned with technical functionalities issues, but also attentive to social, and psychological aspects.

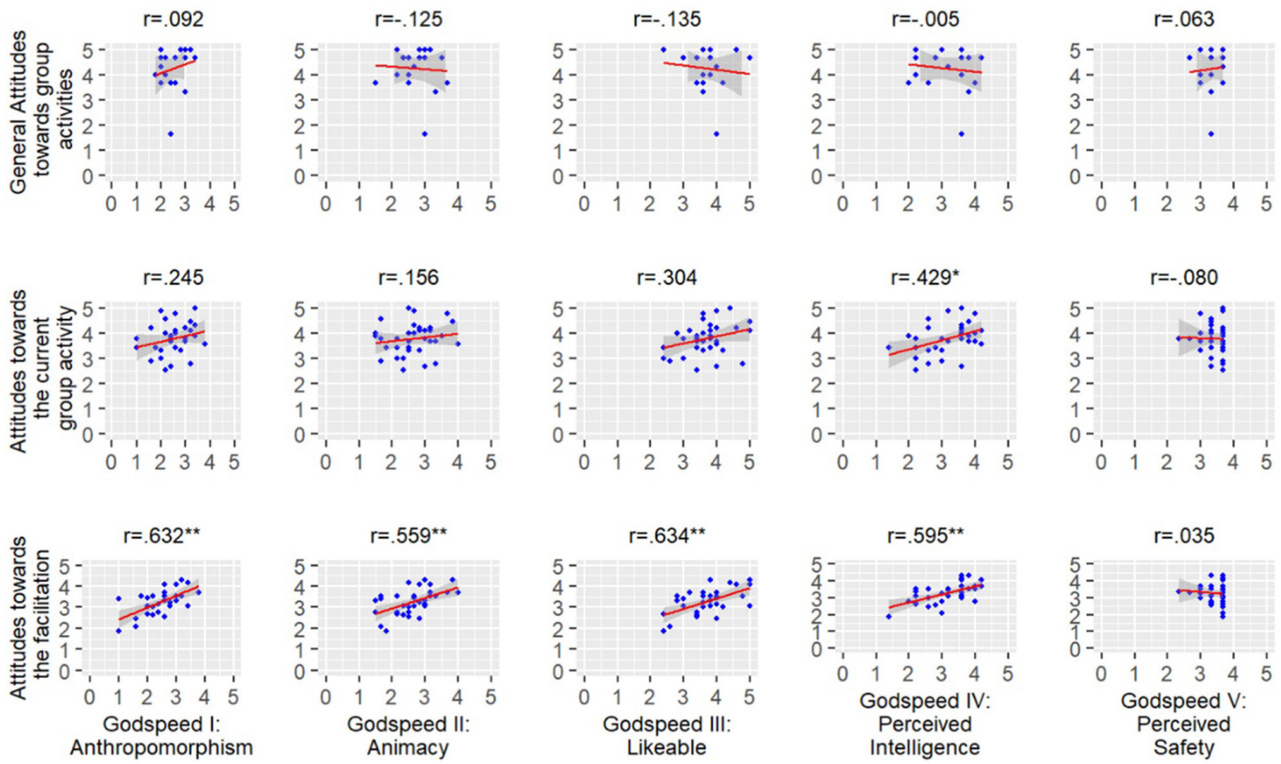


FIGURE 4 | Pearson correlations between Godspeed questionnaires and the attitudes scales. $^{**}p < 0.01$, $^{*}p < 0.05$.

Themes that emerged in the category *technical functionality benefits* include efficiency (e.g., "saves manpower," "time efficient," "put the activity to order")[1], focus (e.g., "its mind is not distracted," "focused on the tasks," "concentrated only in issues relevant to the task"), accurate (e.g., "accurate instructions," "don't forget anything"), and responsive (e.g., "responsive to topics addressed by the students via tablets"). Themes that emerged in the category *technical functionality drawbacks* include limited communication skills (e.g., "its voice was not clear enough," "did not respond to oral questions," "one-shot answer, cannot repeat it"), limited pedagogical skills (e.g., "you cannot ask it follow-on questions," "instructions were not always clear"), and technical problems ("there were some bugs," "its voice was not load enough," "slow boot").

Themes that emerged in the category *psychological drawbacks* include being inhuman (e.g., "not human," "mechanic," "frigid"), awkward (e.g., "strange eye contact," "caused strange feelings," "strange head movements"), limited communication skills ("did not interact enough with the group," "did not adjust itself to the group," "behaved in a not socially acceptable manner"), and impersonal ("no personal relationship"). Interestingly, for many students the fact that the robot was not human was an advantage. Thus, themes that emerged in the category *social and psychological benefits* include objective, not judgmental (e.g., "the robot have no personal bias against one of the students," "the robot does not have a favorite student"), friendly (e.g., "the robot was cute and friendly"), pleasant (e.g., "the robot was polite," "the robot was nice"). The themes *break routine* and *innovative* also emerged but were removed as they were related to the novelty effect.

In addition, analysis of the video recordings revealed that the robot served as a focal point and was very effective in facilitating the activity in terms of time management and group interaction. For example, when the robot gave the students 2 min for individual thinking, the students worked individually, and when it asked to regroup the students immediately regrouped and started to work together. In the tablet and instructor conditions, there was less of a clear distinction between individual and group activity. Thus, for example, when the students read the instruction to work individually for 2 min, in many cases they did not do that but rather worked in a group or pairs.

DISCUSSION

A novel interaction has been introduced, wherein a robot facilitated a small group activity of students in higher education. While we have not explicitly implemented a "flipped-classroom" paradigm (Gilboy et al., 2015; Thomas and Thorpe, 2019), since the students learned the material in a frontal lecture mode, we have applied principles of group facilitation to robot-directed interaction (Chen and Chen, 2015).

[1]Translated from Hebrew by team members.

The post-interaction questionnaire and its quantitative analysis revealed interesting insights into the interaction. Most items did not reveal significant differences between instructor and robot. The only highly significant differences were the expected questions of perception of the robot/tablet/ instructor as human-like, and the perception of the activity as interesting. In the human-like perception question, students rated the instructor as obviously more human-like, but the difference between tablet and robot, while not small, was not significant. This may represent the perception of the students that the robot was "a machine," much like a tablet, and not strictly "a social agent," like a human (Kahn et al., 2011). The perception of the activity as "interesting" was rated significantly higher for the robot condition, but this may be due to the novelty effect: this was the first interaction of the students with a social robot.

Moreover, even though the robot did not directly respond to the students' verbal communications, they still perceived it as friendly and responsive. However, these results should be taken in view of the similar ratings the tablet-condition received. It is unsure how students interpreted "the app was friendly," whereas "the robot was friendly" had a much more direct social interpretation.

The Godspeed questionnaire produced several important insights. The student's perception of the robots correlated with how they perceived the activities. However, the strongest correlations were between the perceived intelligence, anthropomorphism, animacy, likeability, and the facilitation itself. Hence, students who perceived the robot as more animate and likable, rated the facilitation higher. This conforms to previous studies with human facilitators that stressed the importance of the social presence of the facilitator on the activity (Franco and Nielsen, 2018). The rating of the current group activity was only correlated to the perceived intelligence of the robot, emphasizing the difference between activity, which relates to intelligence, and facilitation of the group, which relates also to animacy and likeability. In contrast, the perception of the robot was uncorrelated to the students' attitudes toward group activities in general. The robot's safety, while rated very high, did not correlate to any other scale. This may be due to the physical distance of the robot from the students, its more childlike appearance or lack of possibly threatening actions.

The qualitative analysis of the students' answers gave insights into the benefits of the current setup and raised issues that can be addressed in future applications. First, there were many benefits to the setup, e.g., time management which is an important concern in effective group activities (Gresalfi et al., 2012), accuracy and focus, which can add another layer of efficiency to repeated activities. Second, the fact that the robot was non-judgmental, as opposed to a human facilitator, raises the interesting topic of the benefits of social robots over humans in roles that involve possible judgments (Kidd and Breazeal, 2007). These results also support the media-equation according to which people relate to computers and other technologies, and in this case to robots, in the same way they relate to other human beings (Reeves and Nass, 1996).

However, many drawbacks shed light on possible improvements for future applications. The most obvious ones are technical, e.g., improved quality assurance tests on a larger scale setup are required. The biggest drawback that the students' commented on was the lack of communication skills and responsiveness. Due to technological challenges of natural language processing in a group scenario, especially in the students' native language, these lacks in the setup will not be overcome easily in the near future. However, improved perceptions, such as speaker recognition and engagement via facial expressions (Bhattacharya et al., 2018) can be implemented in such a setup and supply better social and emotional management for the group activity (Matsuyama et al., 2015). Overall, the students commented on the potential of this setup in terms of saving manpower and scalability, non-judgmental and objective facilitation, and increased focus and efficiency of activity management.

Considering the relative acceptance of the students of a robot facilitator puts the role of the future instructor in a new light (Franco and Nielsen, 2018). In our envisioned future "robot facilitated flipped classroom" paradigm, the group facilitation will be conducted by social robots. However, due to formidable technological challenges, the robot cannot understand the discussion's verbal content, nor deal with delicate emotional and social scenarios. Hence, the role of the robot could include for example time management and role assignment whereas the role of future instructors may focus more on answering complex questions, managing divergence from proper discussion content and dealing with emotional and social aspects of the task.

LIMITATIONS

Several limitations in this study should be noted. First, the interaction with a social robot facilitator was novel for all the students and a novelty effect was evident especially with respect to some benefits noted by the students. In order to get a deeper understanding of the long-lasting potential of a social robot facilitator longer interventions (e.g., lasting over a semester) should be examined. In addition, this study was holistic. We were interested in comparing the current state of an instructor facilitating several groups in parallel to the scenario where several robots assist the instructor in facilitating the groups. Nevertheless, this holistic comparison comes with a price tag of control. Thus, there were several differences between the conditions: students in the paper-based condition sat in a lecture hall with all the other groups, whereas in the robot condition they were alone with the robot in a separate room. This makes it more difficult to claim that the effect was of a robot vs. human, or the fact that it was a private facilitator (the robot) vs. a shared facilitator (the instructor). Yet another limitation was the lack of pre-post exams of the content that was due to the fact the HCI content involves skills that are hard to measure. Future studies should conduct a research in a content area that is easier to assess for learning.

CONCLUSIONS

We have introduced a novel educational paradigm in which a social robot facilitated a small group activity in higher educational settings. We have conducted a first study that compared the robot-facilitated setup to human facilitation and activities with tablets only. We have shown that while human facilitation is still considered better in most aspects, students could tap into the benefits of a robot facilitator, such as better time management, objectivity, and efficiency. Nevertheless, in terms of the quantitative data we did not find significant differences that cannot be attributed to a novelty effect (e.g., the robot was significantly more interesting).

Future work will include upgrading the setup to include augmented perception via a larger sensor suite composed of directional microphones and cameras. This will enable real-time speaker recognition and engagement detection to facilitate also the social and emotional sides of the group activity. Furthermore, applying the setup in primary and secondary educational settings raises new challenges, and new opportunities.

Furthermore, while the current study did not asses the students' communication and collaboration skills, future studies will examine the possible positive influence of repeated robot facilitation, using state-of-the-art pedagogy, such as time-management, and maintaining engagement density (Matsuyama et al., 2015), on students' 4C's super skills (Shulman, 1986; Kivunja, 2015).

To conclude, this contribution offers a new and exciting venue for using social robots for robot supported collaborative learning (RSCL) in education, as efficient, objective, and social facilitators for small group discussions.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Tel Aviv University Internal Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

RR-K contributed to the code of the setup, ran the study, and contributed to the data analyzes and writing of the paper. YK assisted in the execution of the research and the analysis of the qualitative findings. GG contributed to the code of the setup, helped in running the study, and contributed to the writing of the paper.

FUNDING

This research was funded by the Minducate Innovation Center of the Sagol School of Neuroscience to RR-K and by the generosity of Eric and Wendy Schmidt by recommendation of the Schmidt Futures program.

REFERENCES

Al-Balushi, S. M., and Al-Aamri, S. S. (2014). The effect of environmental science projects on students' environmental knowledge and science attitudes. *Int. Res. Geogr. Environ. Educ.* 23, 213–227. doi: 10.1080/10382046.2014.927167

Alves-Oliveira, P., Melo, F. S., Castellano, G., and Paiva, A. (2019). Empathic robot for group learning: a field study. *ACM Trans. Hum. Robot Interact.* 8:3. doi: 10.1145/3300188

Bartneck, C., Kulić, D., Croft, E., and Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *Int. J. Soc. Robot.* 1, 71–81. doi: 10.1007/s12369-008-0001-3

Bauer, M. W., and Gaskell, G. (2000). *Qualitative Researching With Text, Image and Sound: A Practical Handbook for Social Research*. London, UK: Sage. doi: 10.4135/9781849209731

Belpaeme, T., Kennedy, J., Baxter, P., Vogt, P., Krahmer, E. E., Kopp, S., et al. (2015). "L2TOR-second language tutoring using social robots," in *Proceedings of the ICSR 2015 WONDER Workshop* (Paris).

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018a). Social robots for education: a review. *Sci. Robot.* 3:eaat5954. doi: 10.1126/scirobotics.aat5954

Belpaeme, T., Vogt, P., Van den Berghe, R., Bergmann, K., Göksun, T., De Haas, M., et al. (2018b). Guidelines for designing social robots as second language tutors. *Int. J. Soc. Robot.* 10, 325–341. doi: 10.1007/s12369-018-0467-6

Bhattacharya, I., Foley, M., Zhang, N., Zhang, T., Ku, C., Mine, C., et al. (2018). "A multimodal-sensor-enabled room for unobtrusive group meeting analysis," in *Proceedings of the 20th ACM International Conference on Multimodal Interaction* (New York, NY), 347–355. doi: 10.1145/3242969.3243022

Björling, E. A., Rose, E., Davidson, A., Ren, R., and Wong, D. (2019). Can we keep him forever? teens' engagement and desire for emotional connection with a social robot. *Int. J. Soc. Robot*. doi: 10.1007/s12369-019-00539-6. [Epub ahead of print].

Blanca, M., Alarcon, R., Arnau, J., Bono, R., and Bendayan, R. (2017). Non-normal data: is ANOVA still a valid option? *Psicothema* 29, 552–557. doi: 10.7334/psicothema2016.383

Boyatzis, R. E. (1998). *Transforming Qualitative Information: Thematic Analysis and Code Development*. Sage.

Bozkurt, A., Akgün-Özbek, E., and Zawacki-Richter, O. (2017). Trends and patterns in massive open online courses: review and content analysis of research on MOOCs (2008–2015). *Int. Rev. Res. Open Distr. Learn.* 18, 118–147. doi: 10.19173/irrodl.v18i5.3080

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp063oa

Brown, L. N., and Howard, A. M. (2014). "The positive effects of verbal encouragement in mathematics education using a social robot," in *2014 IEEE Integrated STEM Education Conference (ISEC)* (Princeton, NJ), 1–5. doi: 10.1109/ISECon.2014.6891009

Ceha, J., Chhibber, N., Goh, J., McDonald, C., Oudeyer, P.-Y., Kulić, D., et al. (2019). "Expression of curiosity in social robots: design, perception, and effects on behaviour," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, UK), 1–406. doi: 10.1145/3290605.3300636

Chen, Y.-H., and Chen, P.-J. (2015). MOOC study group: facilitation strategies, influential factors, and student perceived gains. *Comput. Educ.* 86, 55–70. doi: 10.1016/j.compedu.2015.03.008

Chiu, T. K. F., and Hew, T. K. F. (2018). Factors influencing peer learning and performance in MOOC asynchronous online discussion forum. *Australas. J. Educ. Technol.* 34, 16–28. doi: 10.14742/ajet.3240

Correia, F., Melo, F. S., and Paiva, A. (2019). "Group intelligence on social robots," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Daegu), 703–705. doi: 10.1109/HRI.2019.8673089

Crabtree, B., and Miller, W. (1998). "A template approach to text anyalysis: developing and using codebooks," in *Doing Qualitative Research in Primary Care: Multiple Strategies,* eds B. F. Crabtree and W. L. Miller (Newbury Park, CA: Sage Publications), 93–109.

Crane, T. (2003). *Learning for the 21st Century: A Report and Mile Guide for 21st Century Skills*. Partnership for 21st Century Skills.

Deublein, A., Pfeifer, A., Merbach, K., Bruckner, K., Mengelkamp, C., and Lugrin, B. (2018). Scaffolding of motivation in learning using a social robot. *Comp. Educ.* 125, 182–190. doi: 10.1016/j.compedu.2018.06.015

Edwards, A., Edwards, C., Spence, P. R., Harris, C., and Gambino, A. (2016). Robots in the classroom: differences in students' perceptions of credibility and learning between "teacher as robot" and "robot as teacher." *Comp. Hum. Behav.* 65, 627–634. doi: 10.1016/j.chb.2016.06.005

Fernandes, S., Mesquita, D., Flores, M. A., and Lima, R. M. (2014). Engaging students in learning: findings from a study of project-led education. *Eur. J. Eng. Educ.* 39, 55–67. doi: 10.1080/03043797.2013.833170

Franco, L. A., and Nielsen, M. F. (2018). Examining group facilitation *in situ*: the use of formulations in facilitation practice. *Group Decis. Negot.* 27, 735–756. doi: 10.1007/s10726-018-9577-7

Gilboy, M. B., Heinerichs, S., and Pazzaglia, G. (2015). Enhancing student engagement using the flipped classroom. *J. Nutr. Educ. Behav.* 47, 109–114. doi: 10.1016/j.jneb.2014.08.008

Goel, A. K., and Joyner, D. A. (2017). Using AI to teach AI: lessons from an online AI class. *AI Magazine* 38, 48–59. doi: 10.1609/aimag.v38i2.2732

Gordon, G., Breazeal, C., and Engel, S. (2015). "Can children catch curiosity from a social robot?," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY), 91–98. doi: 10.1145/2696454.2696469

Gresalfi, M. S., Barnes, J., and Cross, D. (2012). When does an opportunity become an opportunity? Unpacking classroom practice through the lens of ecological psychology. *Educ. Stud. Math.* 80, 249–267. doi: 10.1007/s,10649-011-9367-5

Harris, A., and de Bruin, L. R. (2018). Secondary school creativity, teacher practice and STEAM education: an international study. *J. Educ. Change* 19, 153–179. doi: 10.1007/s10833-017-9311-2

Hein, M., and Nathan-Roberts, D. (2018). "Socially interactive robots can teach young students language skills; a systematic review," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Philadelphia, PA), 1083–1087. doi: 10.1177/1541931218621249

Helle, L., Tynjälä, P., and Olkinuora, E. (2006). Project-based learning in post-secondary education—theory, practice and rubber sling shots. *High. Educ.* 51, 287–314. doi: 10.1007/s10734-004-6386-5

Jung, M. F., Martelaro, N., and Hinds, P. J. (2015). "Using robots to moderate team conflict: the case of repairing violations," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY), 229–236. doi: 10.1145/2696454.2696460

Kahn, P. H., Reichert, A. L., Gary, H. E., Kanda, T., Ishiguro, H., Shen, S., et al. (2011). "The new ontological category hypothesis in human-robot interaction," in *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Lausanne), 159–160. doi: 10.1145/1957656.1957710

Khalil, H., and Ebner, M. (2014). *MOOCs Completion Rates and Possible Methods to Improve Retention—A Literature Review*, 1305–1313. Retrieved from: https://www.learntechlib.org/primary/p/147656/

Kidd, C. D., and Breazeal, C. (2007). "A robotic weight loss coach," in *Proceedings of the National Conference on Artificial Intelligence*, *22* (Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press).

Kivunja, C. (2015). Exploring the pedagogical meaning and implications of the 4Cs "super skills" for the 21st century through Bruner's 5E lenses of knowledge construction to improve pedagogies of the new learning paradigm. *Creat. Educ.* 6:224. doi: 10.4236/ce.2015.62021

Kokotsaki, D., Menzies, V., and Wiggins, A. (2016). Project-based learning: a review of the literature. *Improv. Schools* 19, 267–277. doi: 10.1177/1365480216659733

Kory, J., and Breazeal, C. (2014). "Storytelling with robots: learning companions for preschool children's language development," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (Edinburgh), 643–648. doi: 10.1109/ROMAN.2014.6926325

Kory-Westlund, J. M., and Breazeal, C. (2019). A long-term study of young children's rapport, social emulation, and language learning with a peer-like robot playmate in preschool. *Front. Robot. AI* 6:81. doi: 10.3389/frobt.2019.00081

Lai, E., DiCerbo, K., and Foltz, P. (2017). *Skills for Today: What We Know about Teaching and Assessing Collaboration*. Pearson.

Leite, I., McCoy, M., Lohani, M., Ullman, D., Salomons, N., Stokes, C., et al. (2015). "Emotional storytelling in the classroom: individual versus group interaction between children and robots," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, OR), 75–82. doi: 10.1145/2696454.2696481

Leyzberg, D., Spaulding, S., Toneva, M., and Scassellati, B. (2012). "The physical presence of a robot tutor increases cognitive learning gains," in *Proceedings of the Cognitive Science Society*, *34*. Retrieved from: http://escholarship.org/uc/item/7ck0p200

Li, J. (2015). The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *Int. J. Hum. Comput. Stud.* 77, 23–37. doi: 10.1016/j.ijhcs.2015.01.001

Luria, M., Hoffman, G., and Zuckerman, O. (2017). "Comparing social robot, screen and voice interfaces for smart-home control," in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems—CHI'17* (New York, NY), 580–628. doi: 10.1145/3025453.3025786

Matsuyama, Y., Akiba, I., Fujie, S., and Kobayashi, T. (2015). Four-participant group conversation: a facilitation robot controlling engagement density as the fourth participant. *Comput. Speech Lang.* 33, 1–24. doi: 10.1016/j.csl.2014.12.001

Matsuyama, Y., Taniyama, H., Fujie, S., and Kobayashi, T. (2008). "Designing communication activation system in group communication," *Humanoids 2008 - 8th IEEE-RAS International Conference on Humanoid Robots* (Daejeon), 629–634. doi: 10.1109/ICHR.2008.4756016

Moust, J. H. C., and Schmidt, H. G. (1994). Facilitating small-group learning: a comparison of student and staff tutors' behavior. *Instruct. Sci.* 22, 287–301. doi: 10.1007/BF00891782

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., and Dong, J.-J. (2013). A review of the applicability of robots in education. *J. Technol. Educ. Learn.* 1:13. doi: 10.2316/Journal.209.2013.1.209-0015

Okita, S. Y., Ng-Thow-Hing, V., and Sarvadevabhatla, R. (2009). "Learning together: ASIMO developing an interactive learning partnership with children," *RO-MAN 2009-The 18th IEEE International Symposium on Robot and Human Interactive Communication* (Toyama), 1125–1130. doi: 10.1109/ROMAN.2009.5326135

Oliveira, R., Arriaga, P., Correia, F., and Paiva, A. (2019). "The stereotype content model applied to human-robot interactions in groups," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Daegu), 123–132. doi: 10.1109/HRI.2019.8673171

Park, H. W., Rosenberg-Kima, R., Rosenberg, M., Gordon, G., and Breazeal, C. (2017). "Growing growth mindset with a social robot peer," in *12th ACM/IEEE International Conference on Human Robot Interaction* (Vienna: ACM). Retrieved from: http://robotic.media.mit.edu/wp-content/uploads/sites/14/2017/01/park2017hri-mindset.pdf

Patton, M. Q. (2014). *Qualitative Research & Evaluation Methods: Integrating Theory and Practice*. SAGE Publications.

Pendry, L. F., and Salvatore, J. (2015). Individual and social benefits of online discussion forums. *Comput. Hum. Behav.* 50, 211–220. doi: 10.1016/j.chb.2015.03.067

Pfeifer A., and Lugrin, B. (2018). "Female robots as role-models? - The influence of robot gender and learning materials on learning success," in *Artificial Intelligence in Education. AIED 2018.* Lecture Notes in Computer Science, Vol 10948, eds C. Penstein Rosé, et al. (Cham: Springer), 276–280. doi: 10.1007/978-3-319-93846-2_51

Phillips, L. D., and Phillips, M. C. (1993). Faciliated work groups: theory and practice. *J. Operat. Res. Soc.* 44, 533–549. doi: 10.1057/jors.1993.96

Reeves, B., and Nass, C. I. (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press.

Rosenberg-Kima, R. B., Koren, Y., Yachini, M., and Godon, G. (2019). "Human-Robot-Collaboration (HRC): social robots as teaching assistants for training activities in small groups," in *Proceedings of the 2019 ACM/IEEE International Conference on Human-Robot Interaction* (Daegu). doi: 10.1109/HRI.2019.8673103

Saavedra, A. R., and Opfer, V. D. (2012). Learning 21st-century skills requires 21st-century teaching. *Phi Delta Kappan* 94, 8–13. doi: 10.1177/003172171209400203

Saenab, S., Yunus, S. R., Saleh, A. R., Virninda, A. N., Hamka L., and Sofyan, N. A. (2018). Project-based learning as the atmoshphere for promoting students' communication skills. *J. Phys.* 1028:012026. doi: 10.1088/1742-6596/1028/1/012026

Scassellati, B., Boccanfuso, L., Huang, C.-M., Mademtzi, M., Qin, M., Salomons, N., et al. (2018). Improving social skills in children with ASD using a long-term, in-home social robot. *Sci. Robot.* 3:eaat7544. doi: 10.1126/scirobotics.aat7544

Schmidt, S. M. P., and Ralph, D. L. (2016). The flipped classroom: a twist on teaching. *Contemp. Issue Educ. Res.* 9, 1–6. doi: 10.19030/cier.v9i1.9544

Shen, S., Slovak, P., and Jung, M. F. (2018). "Stop. I see a conflict happening: a robot mediator for young children's interpersonal conflict resolution," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY), 69–77. doi: 10.1145/3171221.3171248

Shiomi, M., Kanda, T., Howley, I., Hayashi, K., and Hagita, N. (2015). Can a social robot stimulate science curiosity in classrooms? *Int. J. Soc. Robot.* 7, 641–652. doi: 10.1007/s12369-015-0303-1

Short, E., Swift-Spong, K., Greczek, J., Ramachandran, A., Litoiu, A., Grigore, E. C., et al. (2014). "How to train your DragonBot: socially assistive robots for teaching children about nutrition through play," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (Edinburgh), 924–929. doi: 10.1109/ROMAN.2014.6926371

Shulman, L. S. (1986). Those who understand: knowledge growth in teaching. *Educ. Res.* 15, 4–14. doi: 10.3102/0013189X015002004

Sisman, B., Gunay, D., and Kucuk, S. (2019). Development and validation of an educational robot attitude scale (ERAS) for secondary school students. *Interact. Learn. Environ.* 27, 377–388. doi: 10.1080/10494820.2018.1474234

Spaulding, S., Gordon, G., and Breazeal, C. (2016). "Affect-aware student models for robot tutors," in *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, 864–872. Retrieved from: http://dl.acm.org/citation.cfm?id=2937050

Strohkorb, S., Leite, I., Warren, N., and Scassellati, B. (2015). "Classification of children's social dominance in group interactions with robots," in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction* (New York, NY), 227–234. doi: 10.1145/2818346.2820735

Tahir, Y., Rasheed, U., Dauwels, S., and Dauwels, J. (2014). "Perception of humanoid social mediator in two-person dialogs," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY), 300–301. doi: 10.1145/2559636.2559831

ten Dam, G., and Volman, M. (2004). Critical thinking as a citizenship competence: teaching strategies. *Learn. Instruct.* 14, 359–379. doi: 10.1016/j.learninstruc.2004.01.005

Thomas, G., and Thorpe, S. (2019). Enhancing the facilitation of online groups in higher education: a review of the literature on face-to-face and online group-facilitation. *Interact. Learn. Environ.* 27, 62–71. doi: 10.1080/10494820.2018.1451897

Trilling, B., and Fadel, C. (2012). *21st Century Skills: Learning for Life in Our Times*. John Wiley & Sons.

Wainer, J., Feil-Seifer, D. J., Shell, D. A., and Mataric, M. J. (2006). "The role of physical embodiment in human-robot interaction," in *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication* (Hatfield), 117–122. doi: 10.1109/ROMAN.2006.314404

Yang, D., Wen, M., Howley, I., Kraut, R., and Rose, C. (2015). "Exploring the effect of confusion in discussion forums of massive open online courses," in *Proceedings of the Second (2015) ACM Conference on Learning @ Scale* (New York, NY), 121–130. doi: 10.1145/2724660.2724677

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 18 February 2020
doi: 10.3389/frobt.2020.00015

Child-Robot Collaborative Problem-Solving and the Importance of Child's Voluntary Interaction: A Developmental Perspective

Vicky Charisi[1], Emilia Gomez[1,2], Gonzalo Mier[3], Luis Merino[3] and Randy Gomez[4]*

[1] Centre for Advanced Studies, Joint Research Centre, European Commission, Seville, Spain, [2] Music Technology Group, Universitat Pompeu Fabra, Barcelona, Spain, [3] Service Robotics Lab, School of Engineering, Universidad Pablo de Olavide, Seville, Spain, [4] Honda Research Institute Japan Co., Ltd., Saitama, Japan

The emergence and development of cognitive strategies for the transition from exploratory actions towards intentional problem-solving in children is a key question for the understanding of the development of human cognition. Researchers in developmental psychology have studied cognitive strategies and have highlighted the catalytic role of the social environment. However, it is not yet adequately understood how this capacity emerges and develops in biological systems when they perform a problem-solving task in collaboration with a robotic social agent. This paper presents an empirical study in a human-robot interaction (HRI) setting which investigates children's problem-solving from a developmental perspective. In order to theoretically conceptualize children's developmental process of problem-solving in HRI context, we use principles based on the intuitive theory and we take into consideration existing research on executive functions with a focus on inhibitory control. We considered the paradigm of the Tower of Hanoi and we conducted an HRI behavioral experiment to evaluate task performance. We designed two types of robot interventions, "voluntary" and "turn-taking"—manipulating exclusively the timing of the intervention. Our results indicate that the children who participated in the voluntary interaction setting showed a better performance in the problem solving activity during the evaluation session despite their large variability in the frequency of self-initiated interactions with the robot. Additionally, we present a detailed description of the problem-solving trajectory for a representative single case-study, which reveals specific developmental patterns in the context of the specific task. Implications and future work are discussed regarding the development of intelligent robotic systems that allow child-initiated interaction as well as targeted and not constant robot interventions.

Keywords: child-robot interaction, problem solving, self-initiated interaction, robotics, education

Edited by:
Dirk Ifenthaler,
Universität Mannheim, Germany

Reviewed by:
Sean Andrist,
Microsoft Research, United States
Hirotaka Osawa,
University of Tsukuba, Japan

***Correspondence:**
Vicky Charisi
vasiliki.charisi@ec.europa.eu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *25 October 2019*
Accepted: *27 January 2020*
Published: *18 February 2020*

Citation:
Charisi V, Gomez E, Mier G, Merino L and Gomez R (2020) Child-Robot Collaborative Problem-Solving and the Importance of Child's Voluntary Interaction: A Developmental Perspective. Front. Robot. AI 7:15. doi: 10.3389/frobt.2020.00015

1. INTRODUCTION

The emergence and development of problem-solving cognitive strategies are fundamental mechanisms for human evolution. In the case of childhood, these mechanisms allow children to generate and develop novel mental representations and schemata through playful exploratory activities which gradually transform into deliberate problem solving strategies. These cognitive mechanisms are dominant in a child's development as a combination of a series of interrelated, domain general cognitive skills associated with the prefrontal cortex, such as inhibitory control, shifting, working memory, and others which appear under the umbrella term of Executive Functions (EFs). During the last few decades, EFs have gained increasing attention in developmental and educational research (Keen, 2011; Warneken et al., 2014) and often they are associated with playful and exploratory activities (Best and Miller, 2010).

One of the core elements for the development of exploratory actions is a child's curiosity and intrinsic motivation (Oudeyer and Smith, 2016; Twomey and Westermann, 2018). This allows for the child to exhibit sustained task attention and to proceed from exploratory actions to intentional ones developing the necessary planning skills. Planning, as a prototypical EF, is a high-level cognitive process, which includes goal-directed action sequencing and inhibition of competing impulses (Blakey et al., 2016). Although the growth of EFs follows a common trend, it has been indicated that their components do not develop as a unit; rather, each individual EF follows its own trajectory which might differ among individuals (Diamond, 2006; Best and Miller, 2010; Friedman and Miyake, 2017). Thus, an increasing body of research on child development and learning focuses not only on learning outcomes but on the individual differences of learning process and the transition from one developmental stage to another (Siegler and Crowley, 1991; Best and Miller, 2010; Brock and Taber, 2017).

Among the prevalent methods used for the depiction of child's developmental process is the microgenetic analysis (Piaget and Cook, 1952; Siegler and Crowley, 1991; Lavelli et al., 2005; Montes et al., 2017). Microgenetic analysis focuses on the collection of micro-behavioral data in a dense way in order to capture the emergence and the dynamicity of cognitive development. Individuals are observed over a period of developmental change and the observations are conducted before, during, and after an intervention to capture the process of change. Observed behaviors are intensively analyzed, both qualitatively and quantitatively with the aim to identify the processes that give rise to the developmental change. The microgenetic approach has been used in various contexts such as inhibitory control (Flynn et al., 2004), memory (Schlagmüller and Schneider, 2002), mathematics (Van der Ven et al., 2012), and music-making (Charisi et al., 2018).

While individual trajectories are important for the understanding of child's cognitive development, existing theories highlight the role of social interaction in child's learning (Bandura, 1971; Vygotsky, 1978; Tomasello, 1995). For young children, the development of effective strategies for problem-solving is often associated to scaffolding from the social environment (Tomasello, 1995; Cragg and Chevalier, 2012); collaboration is particularly beneficial for low-ability children when there is an ability asymmetry (Sills et al., 2016).

Based on the above-mentioned theoretical accounts and paradigms, the field of child-robot interaction has examined the ways in which robotic agents might be suitable social learning companions for children in various age-groups and in different contexts such as in second language learning (Kennedy et al., 2016; Kory-Westlund and Breazeal, 2019), in inquiry learning (Wijnen et al., 2019), handwriting learning (Lemaignan et al., 2016), story telling (Leite et al., 2017), problem-solving (Ramachandran et al., 2018), and creativity (Alves-Oliveira et al., 2017). As a recent review on social robots in education (Belpaeme et al., 2018) indicates, social robots have consistently been proved that might be helpful in immediate learning gains in the specific contexts.

However, there are also some suggestions that robots that support child learning should limit their social behavior at targeted times based on the cognitive load and the engagement of the child (Kennedy et al., 2015; Belpaeme et al., 2018). However, the majority of the existing research in the field of child-robot interaction refers to studies with imposed canonical robot interventions which do not allow children to develop their exploratory skills and exhibit self-initiated voluntary interaction. In addition to this, most of the existing work focuses only on the learning outcomes (Charisi et al., 2016) without examining the development of learning process as it occurs and the emergence of possible patterns.

Taken together, the current research in developmental psychology and educational sciences indicate the importance of child's exploratory actions as a core strategy for the development of problem-solving skills. However, existing studies that investigate the impact of social robots in child's learning have mainly focused on imposed robot interventions. As a result, one open and important question is whether a voluntary interaction associates with child's problem-solving process and performance and what are the possible emerging patterns and trajectories of problem-solving process in the case of a canonical, such as turn-taking, and on-demand robot intervention. To address this question we conducted a two conditions repeated sessions study in which children solve a problem together with a robot. The rest of the paper presents the methodology, the data analysis and results of the study, which are discussed against the existing literature.

2. METHODOLOGY

2.1. Research Question

There are many factors that influence children's problem-solving process. In the context of voluntary child-robot interaction, this study explores how the robot's intervention style of voluntary (on-demand) interaction affect child's problem-solving process and task performance in contrast to a canoninical intervention in the form of turn-taking setting.

2.2. Hypotheses

To address the above mentioned research question, based on the existing theoretical and empirical work we develop a set of hypotheses:

- H1: In a child-robot interaction problem-solving activity, children that voluntarily interact with a robot are more likely to show better performance and improvements in their performance than children who interact in a turn-taking setting. We expect that because children that voluntarily interact with the robot might have more opportunities for exploration in problem-solving process.

- H2: Children who voluntary interact and who faced more difficulties in problem-solving (e.g., younger children) would ask more frequently for help by a present robot. We expect this because children in challenging situations look to learn from others (Vygotsky, 1978; Gelman, 2009).
- H3: In developmental problem-solving tasks, it is likely that patterns of solution strategies emerge over time. We expect that because of prior work on child's construct emergence (Gerstenberg and Tenenbaum, 2017).

2.3. Research Design

A behavioral exploratory study is designed to illustrate the relationship between child-robot voluntary interaction and child's problem-solving process and performance, focusing on the importance of exploration. In addition to being one of the first studies that implement child's voluntary interaction and exploration while interacting with robots, this study adopts the developmental design of a microgenetic approach, which allows for patterns of child's problem solving process to emerge and involves the understanding of the "how" of the learning process rather than only its outputs. This involves studying change while it is occurring (Siegler and Crowley, 1991).

The micro-genetic approach is characterized by (i) observations that span a period of rapidly changing competence; (ii) the density of observation within this period is high, relative to the rate of change of the knowledge or skills of interest; and (iii) the observations of changing performance are analyzed intensively, with the goal of inferring the representations and processes that gave rise to them. For this reason, the sample size in microgenetic approaches is typically small and possible comparisons among conditions are approached mainly in a descriptive and qualitative manner. The activities that are designed for microgenetic analysis are characterized by spontaneous and exploratory actions which gradually transform into organized and deliberate behavioral manifestations and contribute to the transition from sensori-motor to symbolic representations (Siegler and Crowley, 1991).

The current educational literature is in consensus about the role of exploration as one of the fundamental processes of child's problem solving in contrast to guided instruction (Dewey, 1902; Whitebread et al., 2012). For this reason, the study follows a two-condition design to contrast guided intervention with child's self-initiated interaction with the robot.

We manipulate robot's intervention as follows: (i) in condition 1 (Cond1), the child is instructed to solve the task in collaboration with the robot in a "turn-taking" scenario, which results in a canonical cognitive intervention by the robot and (ii) in Condition 2 (Cond2) the child is instructed to solve the task independently, having the option to ask the help of the robot whenever (if) this is needed, which results in an "on demand" intervention by the robot. Thus, the children have a self-initiative role and they are free to select if and when the robot would contribute to the solution of the task.

For the execution of this study, the participants are administered the Tower of Hanoi (ToH) task (Hinz et al., 2013) which is characterized by incremental task complexity. They are individually tested in 4 sessions of approximately 10 minutes each (**Figure 1**). First, a *Baseline Session* (BL) is conducted without any robot intervention. This session is followed by two *Intervention Sessions* during which the robot intervenes by suggesting the next optimal movement in the ToH after a child's movement; finally, a fourth session is conducted as *Evaluation Session* (EV) without the presence of the robot. An experimenter is present during the sessions who follows a predefined protocol (see complementary material); however, her role is restricted to provide initial instructions only. In order for us to eliminate any possible procedural bias during the experimental session, the experimenter is sitting inside the room avoiding exhibiting any attention to the child's interaction with the robot and the task performance.

2.4. Participants

Participants includes $N = 20$ (13 boys) typically developing children with an average age of $m = 7.7$, $SD = 1.4$). Of those children, four are 6 years old (yo), seven 7 yo, three 8 yo, three 9 yo, two 10yo, and one 11yo. Given the developmental

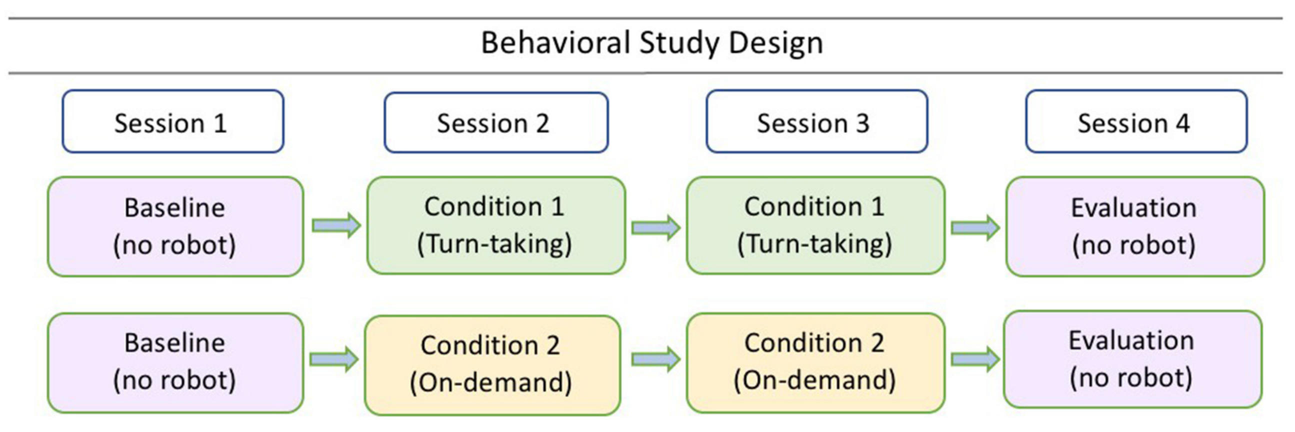


FIGURE 1 | Behavioral study design with four repeated sessions per condition.

nature of the current study with $N = 4$ repeated sessions per child, the sample size is identified to 20 children. The decision for the specific sample size is supported by the fact that this microgenetic exploratory study requires a different approach than experimental studies with detailed analysis of children's behavior development and typically is performed with smaller sample sizes than the ones in experimental research. Lastly, the selected sample size is in accordance to similar previous long-term child-robot interaction studies e.g., $N = 19$ children (Leyzberg et al., 2018) and $N = 14$ children (Kory-Westlund and Breazeal, 2019). One child from Cond2 (voluntary interaction) did not complete the evaluation session and six children from Cond2 did not complete the baseline session; however, we decided to take into consideration their performances during the intervention sessions, since this would provide further input to our observations of the developmental process. The children in this age differ in the degree of intrinsic motivation for task engagement and cognitive abilities. This variability in the sample provides further opportunities for the identification of various developmental patterns during problem-solving activities, which is one of the purposes of this study.

Our analysis includes 72 sessions with 113 tasks from 20 children. Of those, 10 children (4 females and 6 males) ($M = 7.9$, $SD = 1.44$) are assigned to Cond1 and 10 (3 females, 7 males) ($M = 7.6$, $SD = 1.57$) to Cond2. None of the children has any previous experience with the chosen task and any robotic platform; to eliminate any possible novelty effect, we conduct an introductory session with all participant children during which we perform the manipulation check to examine the legibility of robot behaviors.

This research study was approved by the committee on the Use of Humans as Experimental Subjects of the Joint Research Center of the European Commission; parental informed consent was obtained for all participants and all children assented to participate.

2.5. Materials

2.5.1. The Robot

The robot platform considered in the study is Haru (Gomez et al., 2018), a tabletop robot for research on social robotics (**Figure 2**). It presents different modalities for actuation. It can move in 5 degrees of freedom (base, neck, eyes tilt, eyes roll, eyes stroke). The eyes have LCD screens that can play any video. LEDs are present in the mouth and eyebrows of the robot, and it incorporates a set of speakers and microphones. Besides the microphones, the robot uses an external Kinect camera for perception.

The different actuators can be controlled in real time. Also, all these elements can be combined to generate open-loop robot macro-actions mixing movement, eye motion or sounds (see **Figure 3**). These macro-actions are denoted behaviors in this study.

The robot is tele-operated from a control station making use of the Wizard of Oz (WoZ) technique. In the field of human-robot interaction, the WoZ technique is commonly used when the focus of the research is on the interaction design as a step before the development of the autonomous system (Steinfeld et al., 2009; Hoffman, 2016). The control station receives the images from the Kinect camera, and can be used to control directly the different actuators of the robot. Furthermore, the station allows activating pre-designed macro-actions (behaviors). For our study, we designed a set of minimally social behaviors combining the different actuators, as described below in section 2.7. This station is used in the study by the Wizard of Oz (WoZ) to control the robot, mainly by activating the corresponding behaviors adequately.

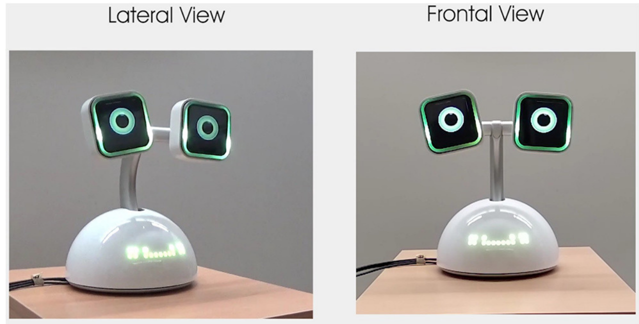


FIGURE 2 | The Haru Robot with mouth LEDs in green and normal eyes on LCDs.

2.5.2. Apparatus: The Tower of Hanoi

The task considered is the Tower of Hanoi (ToH), which has a rich history in cognitive science as a problem-solving task (Simon, 1975). It involves three vertical pegs and a fixed number of colored disks with graduated sizes that fit on the pegs. At the outset, all the disks are pyramidally arranged on one of the pegs with the largest disk on the bottom (**Figure 4**). It requires the arrangement of disks from an initial starting point to a specified end point in the minimum number of moves, allowing the move of one disk at a time and never stacking a larger disk on a smaller one. Any number of disks may be used; the minimum number of moves for a solution is $2^d - 1$, where d is the number of disks.

The ToH task has been used to measure children's planning abilities as well as inhibitory control; for the optimal solution, it requires the use of goal management, in which participants involve inhibition of impulsive moves that bring the child superficially closer to the goal, but are unhelpful for the longer-term solution. However, Miyake et al. (2000) note that participants may use simpler *perceptual* strategies making successive moves that lead to the display looking more like the desired end state.

The solution of the ToH requires that the child sets necessary subgoals which gradually lead to the solution of the task.

2.6. Settings and Procedure

The study has been conducted in a primary school during a summer campus in Spain. A classroom is especially arranged for the setting of the study (see **Figure 5**). In the setting, a table is placed on which the physical instrument of the ToH (see section 2.5.2) and the Haru robot (see section 2.5.1) are deployed, and where the children play the game.

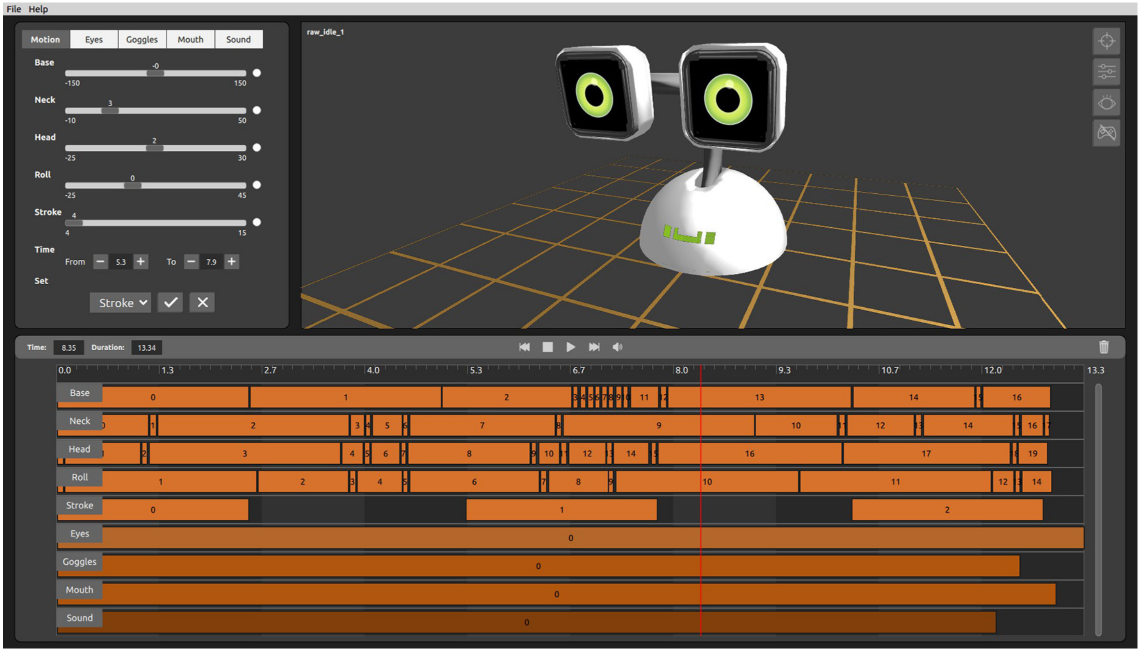


FIGURE 3 | A routine generator application is used to combine the different actuators of the robot (motion of 5 degrees of freedom, eye videos, sound, etc.) to define the open-loop behaviors.

FIGURE 4 | The Tower of Hanoi task.

The teleoperation station for the Wizard of Oz is placed at a corner of the classroom, hidden from the children participating in the study. In order for us to minimize any deception effect because of the teleoperation of the robot, in the beginning of the study, we inform the children about the manner the robot is programmed and teleoperated. However, we have the teleoperation station hidden from the children in order to avoid any distraction. A Kinect camera is located on a vantage point. The video from this camera is fed to the WoZ to facilitate the teleoperation of the robot and the status of the game. Finally, a video camera for data collection is also placed in a reasonable distance from the child to eliminate any possible distraction.

Each individual child participated in four sessions over 1 week. Before the sessions all the children participated to a familiarization session (see section 2.7.3) and they were given to complete the manipulation check (see section 2.7.2). Each of the four sessions of the study was 10–15 min long. In the first session (baseline), the experimenter welcomed the child and asked for his/her assent to participate to the study; she then introduced the Tower of Hanoi rules and let the child solve the game alone. Each time the child completed a task the experimenter was asking whether the child would want to repeat the same task or continue with a more challenging one gradually increasing the number of the disks. In the end of the session the experimenter asked the child whether he/she wanted to continue the next day. In the second session the experimenter introduced the robot and explained the role of the robot depending on the condition. After the end of the four sessions, the children were interviewed about their perceptions of the robot's social competence. However, those results are not reported in this study.

Regarding the procedure followed by the robot, the WoZ is in charge of activating the corresponding robot behaviors (described in the next section) in a timely manner. Furthermore, the WoZ estimates the state of the game to indicate the robot suggestion during its turn in Cond1, or to provide help when asked in Cond2 (the next movement is provided by the robot through its LEDs, indicating the color of the disk and the peg to be moved to). The WoZ also determines when a child is asking for help in Cond2 either verbally or through a button.

2.7. Design of Robot Behaviors

2.7.1. Overview of Robot Behaviors

We constructed a simple behavioral repertoire for the robot that supported the illusion of agency focusing mainly on goal-directed actions and avoiding expressive actions. The robot behavior design was considered as a combination of (i) the type of the behavior and (ii) the timing of the behavior performance.

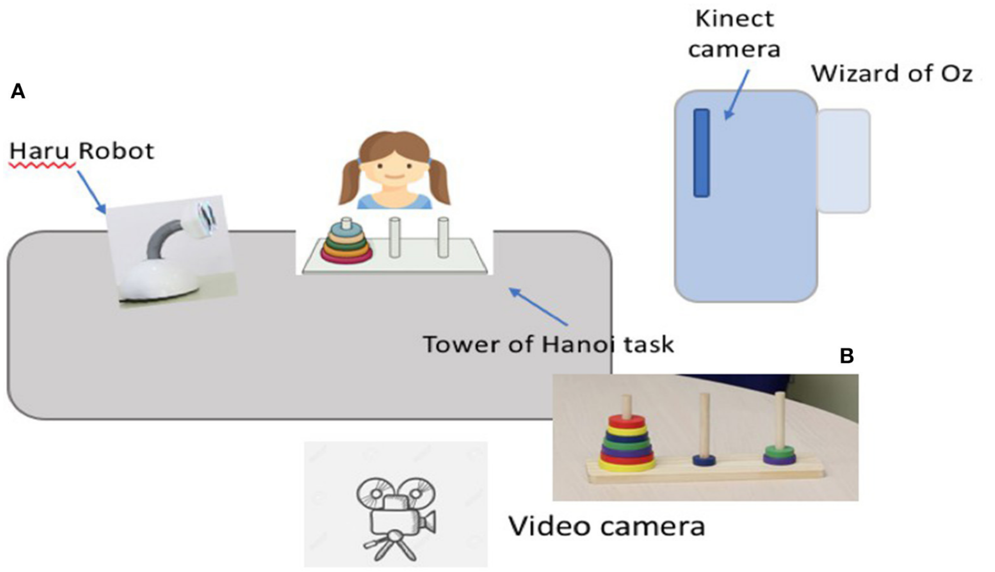


FIGURE 5 | The setting of the study. The children play at a table where the robot **(A)** and the ToH instrument **(B)** are deployed. The Wizard of Oz sits in a hidden location with the teleoperation computer, and it is supported by a Kinect camera. An additional video camera is used to record data.

Regarding the type of behaviors, for the purposes of the current study we designed a set of sonic and gestural non-verbal robot behaviors. In order for us to minimize any possible effect on the expectations that verbal interaction might elicit and on children's intention for initiation of the interaction with the robot, we included only non-verbal behaviors.

More specifically, for the design of different types of behaviors, we used body and eye movements as gestural robot behaviors, sounds, and LED lights to design eight robot behaviors (see **Table 1**). These behaviors were all functional, targeting mainly child's cognitive engagement with the task - with the exception of the starting and ending greetings. The set of behaviors included two types of greetings, two types of providing feedback, three types of task-related behaviors, and one type for indicating that the robot was processing information. In order for us to eliminate any effect which could be related to the type of behaviors of the robot, we kept to the minimum variation of the types of robot behavior. The design of all behaviors was based on previous literature from the field of HRI, design and psychology and were based on minimalistic principles (Saulnier et al., 2011; Cha et al., 2016).

Regarding the timing of intervention, it only related to the cognitive task-related suggestion of the next optimal movement. The robot could give suggestions either in a turn-taking setting (Cond1) or in a setting of voluntary interaction (on-demand, Cond2). For the turn-taking setting, the robot would intervene in turns with the child by providing feedback on the previous movement made by the child, followed by a suggestion for the next task-related optimal movement. For the voluntary interaction setting, the robot would intervene only in the case in which a child would ask for help. The child could ask for help either verbally or with the use of a help button.

TABLE 1 | Robot behavior repertoire.

Robot intention	Robot executed behavior
Greeting hello to the kid	The robot rotates the basis (45° right, 90° left, 90° right, 45° left) it stands still, it rotates the eyes (45° right, 90° left, 90° right, 45° left), it performs sound
Indicating the start of the game	The robot performs a dancing movement and looks at the task
Indicating child's turn	Robot looks at the task and looks at the child
Indicating processing current information	Looks at the task, sequence of different colors LED around the eyes moving toward outside
Suggesting the next movement	Instant suggestion of the color of the Disk and the number of Peg (visual projection on the screen of the eyes and the body
Informing that the movement was optimal	Looks at the task—looks at the child and green happy (once)
Informing that the movement was suboptimal	Repeatedly looks at the task looks at the child (x2)-wiggles NO
Greeting goodbye to the kid	Repeated rotation of the eyes (45° right, 90° left, 90° right, 45° left, LED white softer, fading sound

2.7.2. Manipulation Check

We conducted a manipulation check for the confirmation of the legibility of the robot's behavior. It was designed as a single group-session with all participant children. During the session, the robot performed the designed behaviors and we asked the children to indicate their perceived robot intention in the form of a written task. The experimenter and the research group facilitated the session by introducing the robot, explaining the purpose of the session and guiding the performance of the behaviors and the gathering of children indications. In total

$N = 20$ children participated to the manipulation check. We included four action-directed behaviors (Greeting hello to the kid, Informing that the movement was optimal, Informing that the movement was suboptimal, Greeting goodbye to the kid); finally we examined two additional expressive behaviors (happy and sad) in order to justify children's understanding of the current task. The results of the manipulation check show that $71,43\%$ of children's answers were accurate regarding the legibility of robot's behaviors. More specifically the behavior with the higher percentage of legibility was the Sad behavior with 100% correct answers. These results justified that the children understood the current task. This was followed by the "Informing that the movement was optimal" behavior with $91,67\%$ correct answers. The least legible behavior was the "Greeting hello to the kid" behavior with 50%, which was confused with the "happy" expressive behavior. However, given that the manipulation check was performed in a de-contextualized manner we expected that there might be a confusion between the goal-directed and the expressive actions which are not mutually excluded.

2.7.3. Familiarization Phase

For the elimination of any novelty effect on children's behavior, following the manipulation check, we allowed the children to informally interact with the robot. This informal activity lasted 10 min and was designed to be unstructured. Each child was free to interact with the robot at his or her willingness and the researchers did not impose any kind of interaction. All the children remained into the classroom for the informal activity.

3. METRICS AND ANALYSIS

Audio and video recordings of the study sessions were recorded with two cameras for later transcription and off-line analysis. A first iteration of the recorded sessions observation as well as the initial hypotheses of the study lead us to the development of the annotation scheme. As it was expected, since the robot did not exhibit any verbal behavior, the child-robot verbal interaction was minimum. The only case the children were addressing verbally to the robot was during the Condition 2 of the voluntary interaction when they asked for help verbally—in addition to the option of asking for help with the use of a button. For this reason, the verbal interaction data reported in this paper only includes child's verbal behavior of "asking for help."

The recorded video was used to transcribe children's task-related behavior as well as social interaction with the robot and verbatim. However, for the purposes of the research question addressed in this paper, we only report the task-related behavior. Participants' behaviors were manually annotated off-line by an instructed annotator. Because of the objective nature of our coding scheme (disk movements, asking for help Cohen's K and breaking the rules), which did not require any subjective interpretation of children's behavior, we run a set of sessions during which the two coders annotated the same extracts. During those sessions, any minor disagreement was discussed with the first author of the paper, which resulted in a consensus of the coding.

We annotated in total 72 individual sessions. In each one, more than one task could be included, depending on the duration of child's task performance. This resulted in the annotation of 113 tasks from 3 to 7 disks of the Tower of Hanoi. The annotation scheme included (i) the occurrence of task-related actions (disk movement); (ii) the use of help button or child's verbal asking the robot for help; and (iii) the instances of breaking the rules of the game. In addition, we chose $N = 4$ case studies (see section 4.3), which correspond to 16 sessions (BL, Interventions and EV) to annotate the characterization of the child's task-related action (optimal or suboptimal, see below).

We observed that because of the canonical robot intervention in the sessions of the turn-taking condition, the sessions in Cond1 lasted longer than the ones in Cond2. For this reason, we normalized the sessions duration taking into consideration the optimal number of movements per task. However, we did not consider the sessions duration in our data analysis because robot's canonical interventions in Cond1 (turn-taking) and the on-demand intervention in Cond2 would create an imbalanced comparison between the two conditions. For this reason, our data analysis only focused on children's task-related actions.

Using the off-line video annotation tool ELAN[1], we manually annotated the data according to the annotation scheme.

3.1. Task Performance

To measure the performance of a given task, we annotate individual movements and compute the difference between the number of movements L and the optimal number of movements O_d for the number of disks d of the task. In order to compare this metric for tasks with different number of disks, we then normalize this value by the optimal number of moves in task as follows:

$$K = \frac{\Delta L}{O_d} = \frac{L - O_d}{O_d} \tag{1}$$

where d varies from 3 to 7 disks and $O_d = 2^d - 1$[2]. Please note that, as defined, higher values of the metric K indicate lower performance.

Since the child was free to choose whether after the completion of one task she/he would continue to the next task by increasing the number of disks or not, we considered for our analysis the total number of tasks per disk, which might exceed the number of participants.

In the following analyses, we ran a Kolmogorov–Smirnov test to check the normality of the data. Based on the Komogorov-Smirnov result, we used non-parametric Wilcoxon test for paired samples to check the difference in children's problem-solving performance (learning) by comparing the results of the baseline and the evaluation session, and Mann–Whitney's U-test for non-paired samples to compare the results of the evaluation session between the two conditions.

[1]https://tla.mpi.nl/tools/tla-tools/elan/

[2]Values in this study are $O_3 = 7$, $O_4 = 15$, $O_5 = 31$, $O_6 = 63$, $O_7 = 127$.

3.2. Voluntary Interaction

In Cond2, we designed a voluntary HRI setting with an "on demand" robot intervention as an indicator of child's intrinsic motivation for problem-solving. We annotated and counted the instances in which the child was explicitly asking the robot for help either verbally or using a help-button. We normalized the number of instances with respect to the number of optimal moves per disk O_d to obtain a measure for help H. This normalization was done to obtain comparable values for different tasks.

3.3. Task Improvement

As being a developmental study, we also analyse the development of child's task performance K along the different sessions. Thus, we analyzed the improvement between (i) the first and the last task of the baseline; and (ii) the last task of each session and the first task of the following session, by computing the difference of normalized extra moves metric K in both cases.

3.4. Developmental Process

The Tower of Hanoi game can be represented as a graph (the Hanoi graph) (Knoblock, 1990; Hinz et al., 2013), as illustrated in **Figure 6**, in which each node represents a legal disposition of the disks on the pegs (for instance, for the 3-disk case, the node 112 represents the smallest disk in the 2nd peg, and the other disks in the 1st peg) and edges represent valid movements between nodes.

For d disks, there are 3^d nodes. Under a certain positioning of the nodes, the graphs resemble the Sierpinski gasket (Hinz et al., 2013). We used this model to manually annotate each sequence of movements and relate it to the optimal path, understood as the solution with the minimum amount of moves O_d. Due to the high cost of manual annotation, we carried out these annotations for all tasks but only for subgoals with a maximum of 5-disks, i.e., $d \leq 5$.

We annotated the type of movements as follows: (i) Optimal/sub-optimal: optimal moves refer to the moves which are on the optimal path toward the solution or function as recovery actions (sub-optimal) toward the optimal path; and (ii) Auxiliary: Auxiliary movements refer to those that use a third peg as a scaffold for the optimal solution of the task. Within a task of a certain number of disks there are subgoals; these are instances of milestones of a subpyramid that leads to the task solution (see **Figure 6**).

4. RESULTS

We used the above-mentioned metrics to address the research questions and the corresponding hypotheses as follows.

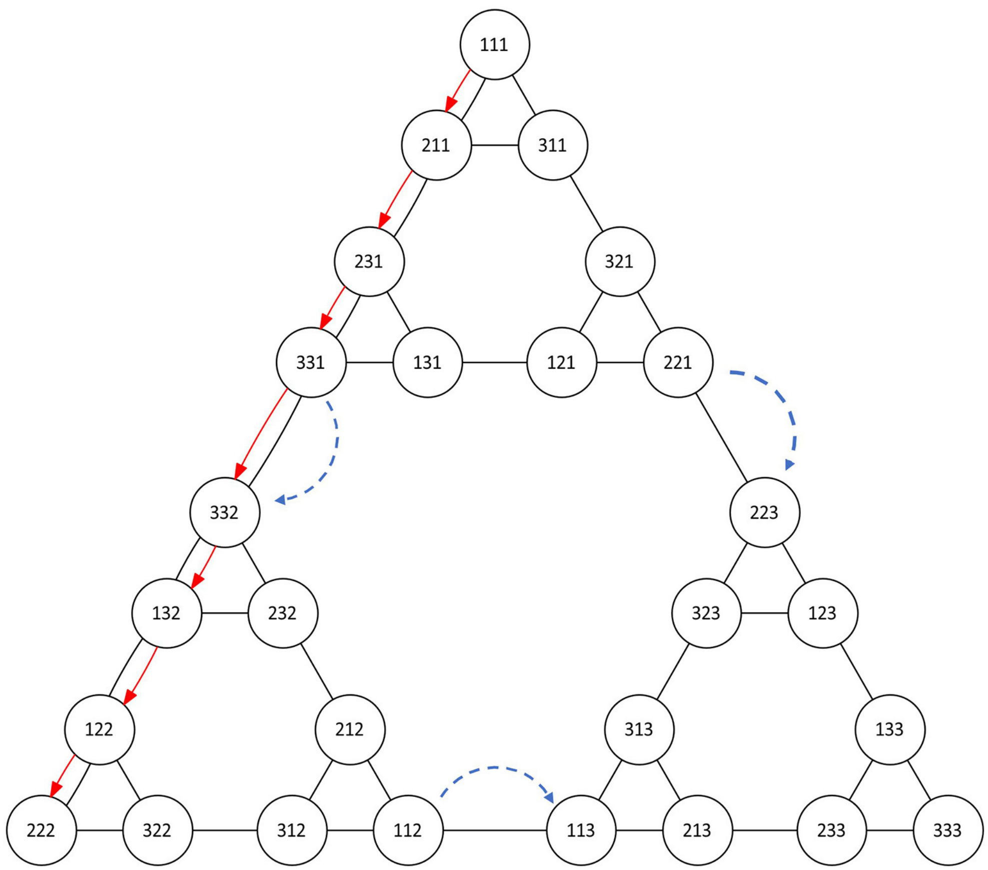


FIGURE 6 | Graph representation of the ToH game for $d = 3$ disks. Each node represents one disposition of the disks. Red color represents the optimal path between the initial disposition and one solution. In blue dashed, auxiliary movements (movements between sub-graphs leading toward the solution).

4.1. Task Performance in Turn-Taking and Voluntary Interaction (Hypothesis 1)

We hypothesize that children in Cond2 (voluntary interaction) would be more likely to show better performance in the evaluation session than children in Cond1 (turn-taking). To explore this hypothesis, we consider (i) children's task performance and (ii) task improvement over the four sessions, with a focus on the evaluation session. We note that, during the intervention sessions, children in Cond2 had the opportunity to perform more movements than children in Cond1 since in Cond1 the robot provided canonical intervention in a turn-taking setting.

4.1.1. Task Performance

For each task, we assess the value of K (normalized extra moves), as detailed in section 3.1.

Mean performance metrics are summarized in **Table 2** (BL and Cond1) and **Table 3** (Cond2). The average is presented in relation to the incremental task complexity (number of disks d with $d = 3...7$). For the baseline, we integrate the performance of participants in both conditions as there is no difference in the setting.

As expected, in the baseline, we observe an increase in the normalized extra moves K for the task with increased difficulty (more disks), ranging from 0.40 ($d = 3$) to 1.17 ($d = 5$). In the intervention session of Cond1 the normalized number of extra movements seem to not be associated with the increased difficulty of the task ranging from 0.12 ($d = 6$) to 0.43 ($d = 3$) with relatively small deviation from the optimal solution path during the robot's canonical intervention. However, in the intervention session of Cond2, the extra movements range was between 0.45 ($d = 4$) to 1.18 ($d = 6$) which is a larger deviation from the optimal path than in Cond1. As expected, the task performance is linked to the difficulty of the task in terms of number of disks.

Interestingly, in the evaluation session in Cond1, task performance K ranges from 1.59 ($d = 7$) to 3.51 ($d = 5$), and the deviation from the optimal solution is higher in the first task of the session than in later stages with increased difficulty. However, in the evaluation session of Cond2, K is smaller than for Cond1, ranging from 0.6 to 2.55 which indicates a smoother transition from the intervention to the evaluation session in the voluntary interaction case. It should be noted that in many cases the Standard Deviation of the selected metrics is relatively large, which indicates a large distribution, most likely because of the small sample size.

4.1.2. Task Improvement

Figure 7 shows the distribution of task performance for BL and EV sessions considering both conditions. We observe that the median value is higher in evaluation than in baseline, indicating an overall learning effect. In addition, we computed individual differences in task performance ΔK between EV and BL session. It should be noted that, since we measure the difference in normalized extra movements, a negative difference indicates task improvement. Our descriptive results show an average $\Delta K = -0.326$ (or 35% if instead of difference we compute the percentage decrease), which also reflects a better average performance in the EV session. However, the results from the Wilcoxon test between BL and EV value distributions showed no statistical significance ($p = 0.286$, $a = 0.05$). As discussed below, due to the increasing difficulty of the task in the EV session, our findings might indicate a learning tendency.

In addition, we performed a Mann–Whiteney's U-test to check the statistical difference between EV sessions in Cond1 and Cond2. Statistical distributions are illustrated in **Figure 8**. Results show significance of $p = 0.038$ ($a = 0.05$). The interval of confidence for the difference between Cond1 and Cond2 is between 0.020 and 1.237, which means that the performance is significantly higher in Cond2 than Cond1.

Lastly, we looked at the possible association of those results with the age of the children (**Figure 9**). Our results show that in the evaluation session of Cond1 most of the children of any age perform larger numbers of movements than in Cond2.

TABLE 3 | Task performance metrics ΔL and K in Cond2 (voluntary interaction) per session and per number of disks.

d	Cond2-Intervention			Cond2-Evaluation		
	t	ΔL	*K* Mean (SD)	t	ΔL	*K* Mean (SD)
4	7	6.86	0.45 (0.17)	1	9	0.6 (0)
5	12	20.5	0.66 (0.10)	2	31	1 (0.15)
6	10	74.5	1.18 (0.11)	2	161	2.55 (1.41)
7	–	–	–	4	–	1.15 (0.38)

t represents the number of considered tasks.

TABLE 2 | Task performance metrics ΔL and K in Baseline (left column) and in Cond1 (turn-taking) per session and per number of disks.

d	Baseline			Cond1-Intervention			Cond1-Evaluation		
	t	ΔL	*K* Mean (SD)	t	Δ L	*K* Mean (SD)	t	ΔL	*K* Mean (SD)
3	18	2.78	0.40 (1.24)	1	3	0.43 (0)	–	–	–
4	16	15.38	1.03 (0.79)	6	4.67	0.31 (0.17)	–	–	–
5	5	36.40	1.17 (0.99)	13	4.69	0.15 (0.10)	2	109	3.52 (2)
6	–	–	–	14	7.28	0.12 (0.11)	4	112	1.78 (0.33)
7	–	–	–	–	–	–	5	201.6	1.59 (0.45)

t represents the number of considered tasks.

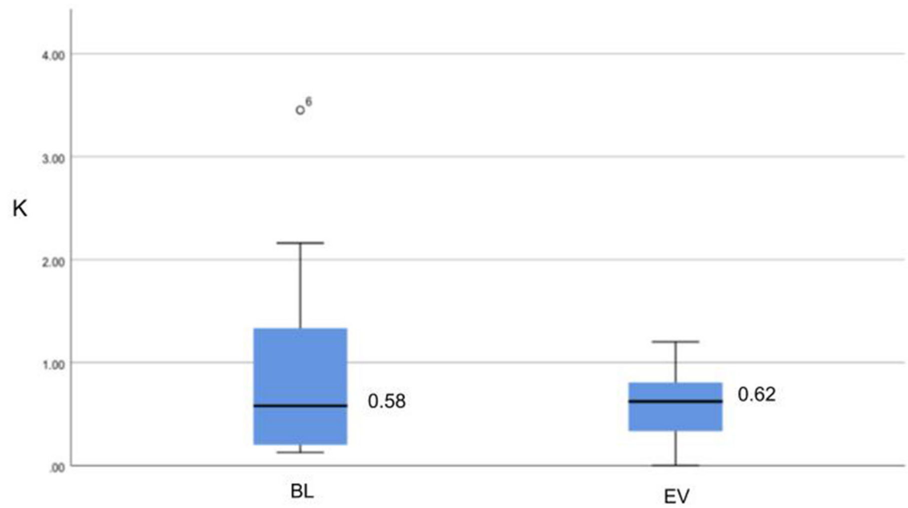


FIGURE 7 | Distribution of task performance *K* for BL and EV sessions in both conditions. Median values are displayed.

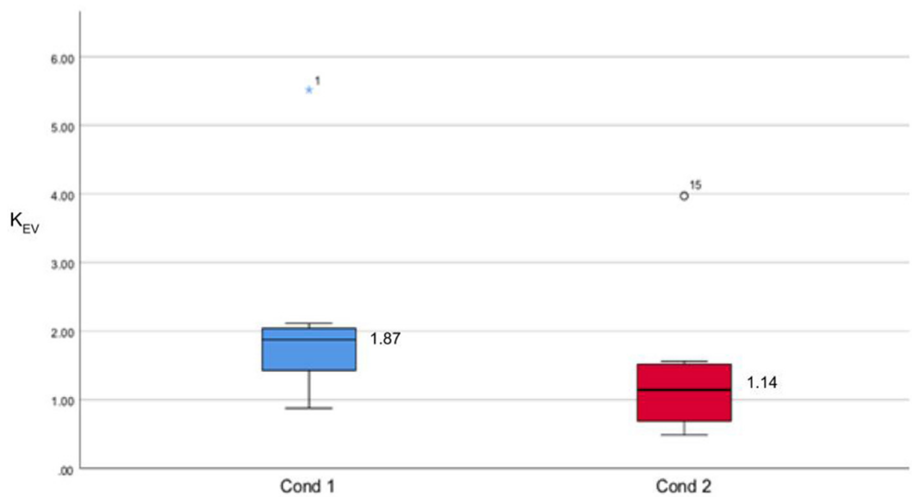


FIGURE 8 | Statistical distribution of task performance *K* in evaluation sessions of both conditions. Median values are displayed.

4.2. Children's Tendency for Self-Initiated Robot Interaction (Hypothesis 2)

We hypothesize that in the voluntary interaction the participant children who faced more difficulties in solving the task (e.g., younger children) were more keen to ask the robot for help. We expected this because child's learning often occurs in collaborative settings with the scaffolding by others (Vygotsky, 1978). To explore this hypothesis, we considered (i) the number of the instances the individual child asked for help (asking for help) in relation to the task performance (extra movements) and (ii) the age of the child. Because of the incremental nature of the task, there were more opportunities for children to ask for help; for this reason, we normalized the scores in order to be comparable.

We assess child's voluntary interaction in relation to the task performance during Cond2 as well as the task improvement in the evaluation session.

During the robot intervention in voluntary interaction, we observed a trend for more instances of asking for help in less demanding tasks by younger children (**Figure 10**). More specifically, for $d = 4$ tasks, children of average age 6.8 years exhibit $H = 0.26$ asking for help behavior. For $d = 5$ tasks, children of average age 7.5 years exhibit $H = 0.16$ asking for help behavior. For D6, children of average age 8.3 asked for help $H = 0.03$ instances and for $d = 7$ the only child that managed to perform the task with 7 disks in the intervention session was a 10 year-old who didn't ask for help at all.

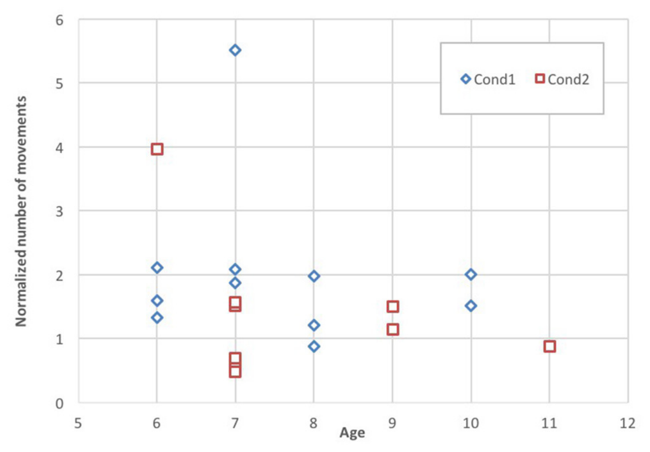


FIGURE 9 | Individual task performance K versus age in the evaluation session for Cond1 and Cond2.

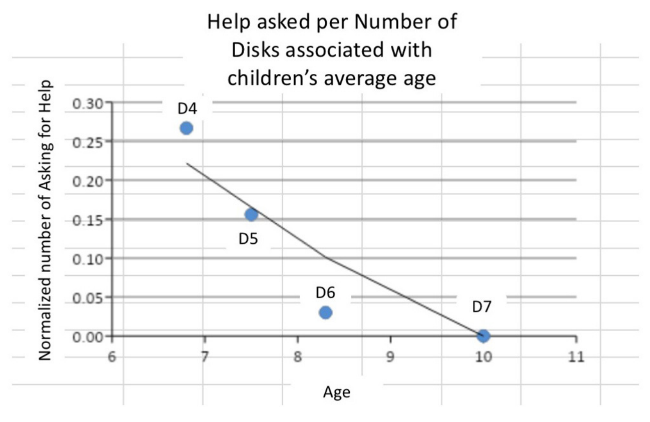


FIGURE 10 | Average H metric of child's asking for help in Cond2 vs. child's age, presented according to number of disks.

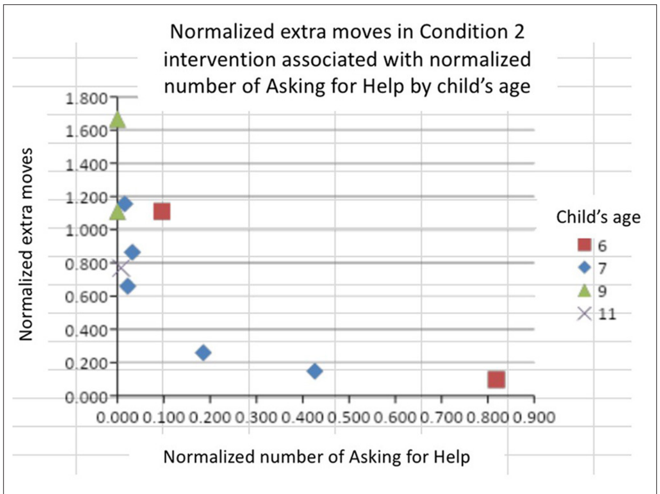


FIGURE 11 | Representation of K vs. H to represent normalized number of movements with respect to help by the robot.

In addition, we considered the score of extra movements for these children. As shown in **Figure 11** in total 9 out of 10 children asked for help during the robot intervention. Of those, 6 children showed increased number of extra movements ranging from $K = 1.11$ to $K = 1.66$ with low number of instances of asking for help (normalized range from $H = 0.00$ to $H = 0.096$). On the contrary, three children exhibit increased number of instances of asking for help, ranging from $H = 0.18$ to $H = 0.82$, with decreased number of extra movements, ranging from $K = 0.25$ to 0.097.

4.3. A Single Case-Study, Pattern Emergence, and Inter-individual Differences (Hypothesis 3)

To gain a more refined understanding of the problem-solving process and to identify possible patterns in action sequences, we map the developmental trajectories of the task solution for $N = 4$ selected children. The selection of the specific case studies was based on their representiveness in terms of the solution path that the children followed during the sessions. In this section we analyse one single case study and we selectively juxtapose instances from the remaining three case studies.

For our analysis, we assessed all movements as optimal or suboptimal and mapped it to the visual representation of the ToH solution presented above (see **Figure 6**). We used the visualization to map the sequence of child's task-related actions and to define possible emerging patterns.

4.3.1. Baseline Session

4.3.1.1. Optimal performance

The child "Sophie," aged 8 years, participated in Cond1 of the study. During the baseline session, without the presence of the robot, Sophie understood the rules of the game and showed a positive stance toward the game and the activity. She started solving the task with $d = 3$ disks, without facing any difficulty. We observed that toward the end of the solution, Sophie increased the pace of her task-related actions. This has been registered as a typical behavior that was observed repeatedly in all participant children and can be explained by the cognitive theories that describe child's *perceptual* strategies making successive moves that lead to the display looking more like the desired end state (Miyake et al., 2000).

4.3.1.2. Deviation

Then, Sophie proceeded to the next task with $d = 4$ disks. While in the beginning of the task, we observed an increased pace in her actions, after the movement 4, the solution pace was diminished and, as shown in **Figure 12**, she started deviating from the optimal solution path. The point she started to deviate was the instance where she should perform an auxiliary movement and inhibit inappropriate move selection. This demand appears at specific points where there is a mismatch between the end goal

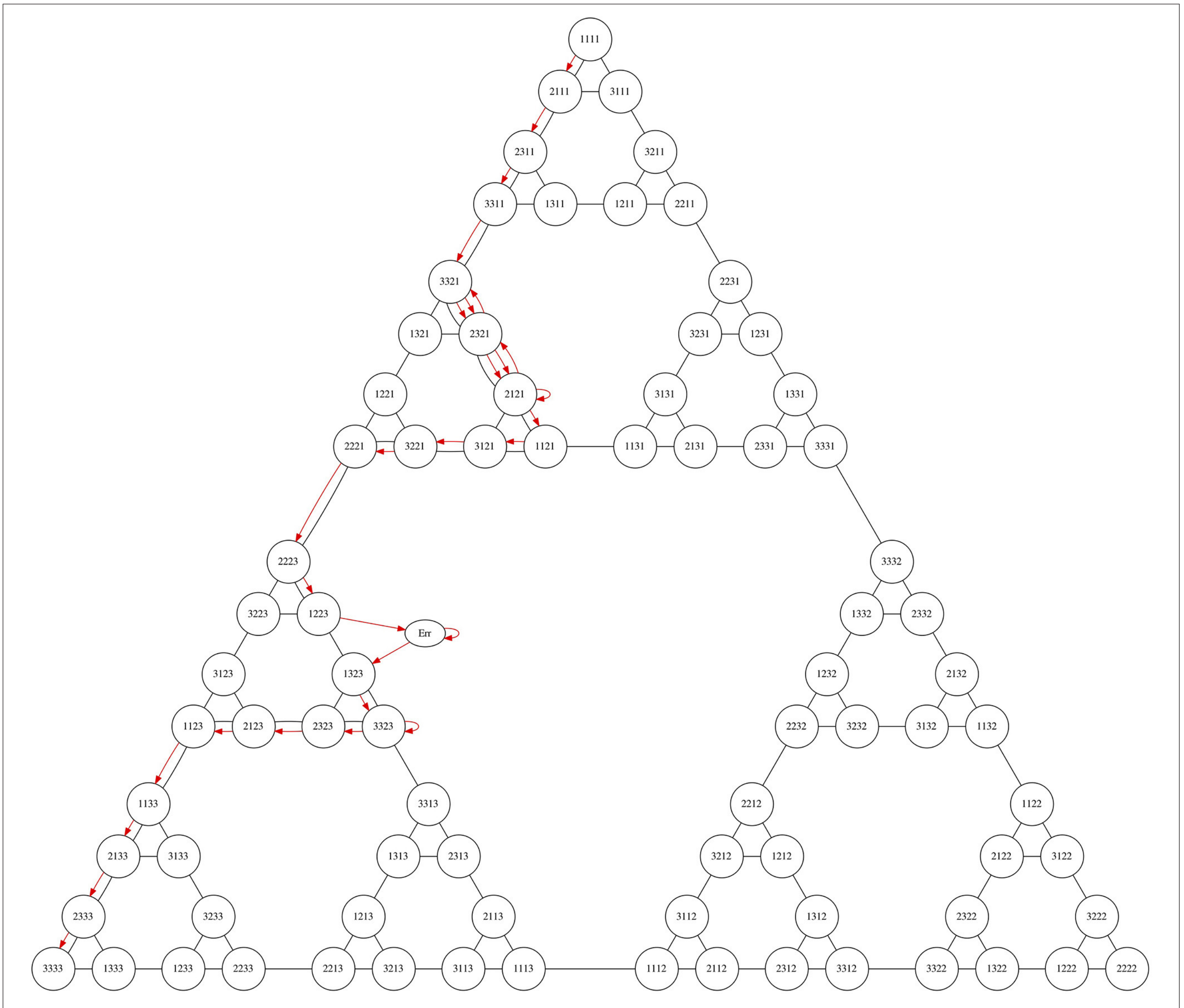


FIGURE 12 | A developmental representation of individual child's problem-solving path of the baseline. The figure shows a pattern of frequent deviation from the optimal solution path.

of the problem and a current subgoal. This was a typical behavior that appeared in the Baseline session in all the four case studies we evaluated.

4.3.1.3. Recovery

After four movements, Sophie understood that she was not on the optimal solution path of the task and she started performing recovering actions. We observed an increased pace of her actions during the recovery which might be explained by theories that focus on executive function of planning (e.g., Miyake et al., 2000).

4.3.1.4. Inhibitory control points

The solution of the $d = 4$ disks ToH task requires from the child at least three instances of inhibitory control. At those points the child should perform an auxiliary movement in order not to deviate from the optimal solution path. However, Sophie did not make use of the auxiliary movement which resulted in a canonical deviation from the optimal solution path as appears in **Figure 12**.

4.3.1.5. Child-initiated interaction

In condition 2, we annotated the child initiated interaction indicated by the instances of the child's asking for help as described in section 4.2. The microgenetic assessment provides further insights on the timing of child-initiated interaction. As expected, we observed that the majority of the instances appear on the nodes where the child had more than one options to perform the next movement with higher probability to deviate from the optimal path. This coincides with the auxiliary actions that indicate the child's inhibitory control.

4.3.1.6. Pattern emergence of developmental sequences

The pattern which appears in Sophie's baseline for the $d = 4$ disks task appeared in all the four case studies we analyzed. In a similar way, a typical solution path in turn-taking condition consisted of optimal moves only following the diagonal axis of the triangle. In the evaluation session, the child exhibits canonical deviation from the optimal solution path with an improvement from the baseline, and a more frequent deviation from the optimal path than the one exhibited in the intervention session.

4.3.1.7. Pattern emergence of temporal aspects

To illustrate how the problem-solving trajectory develops over time, we examine the speed of the moves throughout the task. **Figure 13** shows a selected set of representative examples from the analyzed cases with the duration of each move (in seconds), in addition to a moving average of the last three movements. We observe an increase in speed (short duration) in the last movements of a subgoal for all the analyzed children. Additionally, we observe that the increase in the speed of movements toward the final solution of the tasks is associated with optimal movements, while increase in the speed of movements between subgoals of the same number of disks is associated with suboptimal movements such as exploratory actions.

4.3.2. Robot Intervention Session

In the second and the third session Sophie participated to the robot intervention session in Cond1 in which Sophie was instructed to solve the ToH task together with the robot in a turn-taking setting. Sophie looked engaged with the robot and she clearly perceived all the intended behaviors of the robot. She selected to repeat the task with $d = 4$ disks which she solved in the optimal way in collaboration with the robot. As shown in **Figure 12**, her performance was optimal in the task with $d = 5$ disks as well when solving the task together with the robot.

While an optimal solution was typical for all children in Cond1, children in Cond2 showed different patterns of task solution, which differ depending on the frequency child asked the robot for help. In the examined case studies we observed solutions with (i) canonical deviation from the optimal solution at the points which required auxiliary movements (ii) instances of child breaking the rules of the game and (iii) solutions with extensive exploratory actions which lead to a final solution with the use of large number of extra movements.

4.3.3. Evaluation

In the last session, Sophie selected to solve again the $d = 5$ disks task, without the help of the robot. While in the intervention session, Sophie solved the task together with the robot following the optimal path, in the evaluation session she regularly deviated by the optimal as shown in **Figure 12**. The pattern of Sophie's deviation from the optimal path in the evaluation resembles the one in the Baseline with the four disks task, in having the critical points of the use of auxiliary movements as necessary for the continuation of the optimal solution. However, Sophie's pattern of solution seems improved in the evaluation session since she achieved to use inhibitory control in four out of eight critical points. This indicates the dynamic nature of problem-solving

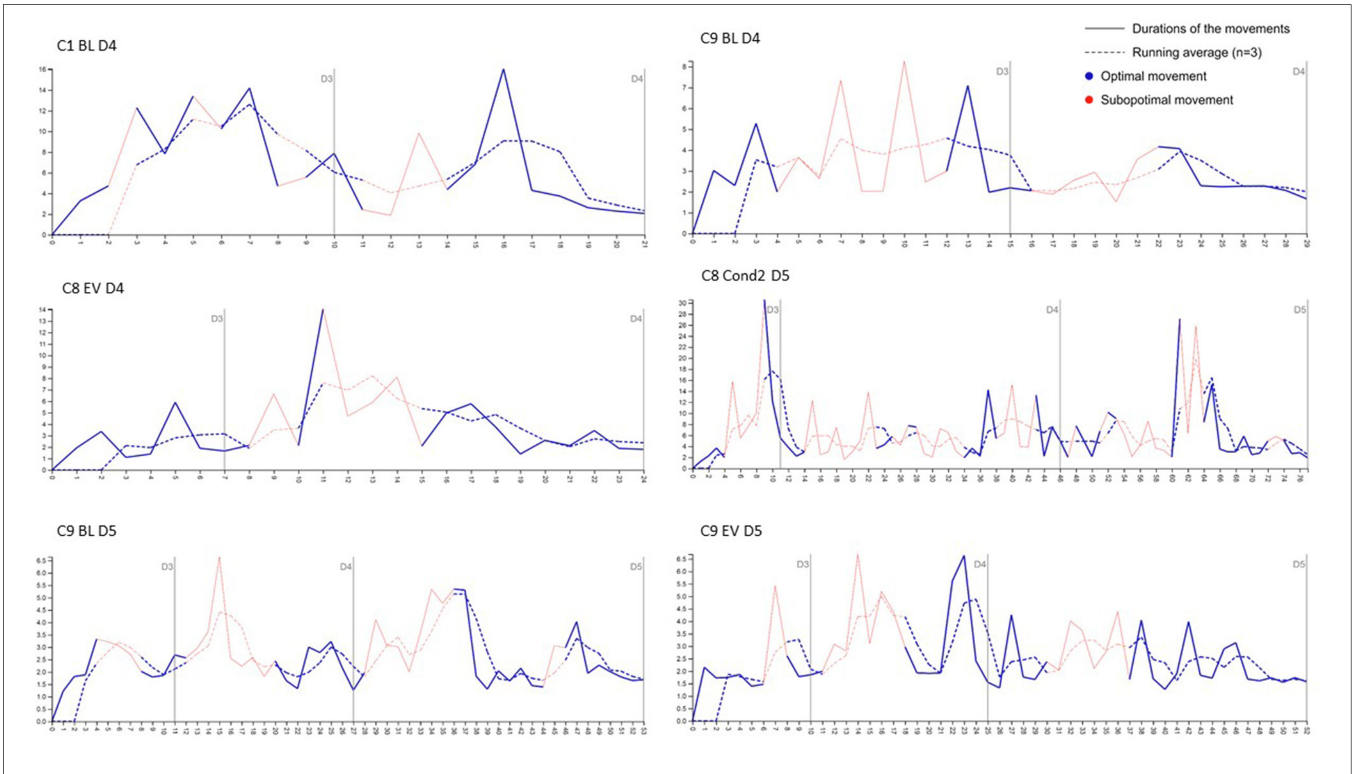


FIGURE 13 | Examples of speed of optimal and suboptimal actions in association with the subgoals of the problem-solving task.

process in incremental tasks which requires special attention to the design of robot intervention.

5. DISCUSSION

In the current study, two main topics were addressed: First, we evaluated children's problem-solving task performance in a "voluntary" HRI condition in contrast with a "turn-taking" condition in a longitudinal setting. Second, we examined the developmental trajectory of the process of problem-solving via possible patterns of the sequence of actions over multiple sessions. To address the first topic we captured children's performance of the ToH task in an incremental manner looking at the role of the robot intervention on the task performance. To address the second topic, we considered children's deviation from the optimal path of the solution which allowed us to highlight the heterogeneity of children's problem-solving trajectories. Our goal was to observe children's trajectories of problem solving, and to create an HRI setting that allowed for voluntary childrobot interaction with child-initiated robot intervention. Below we discuss the main findings:

5.1. Exploration in Young Children's Problem-Solving

Our results indicate that participants in the "turn-taking" condition exhibit less exploratory movements than in the "on-demand" robot intervention condition. However, in challenging tasks, young children that participated in the "on-demand" robot intervention and had the possibility to perform more exploratory actions outperformed young children that participated in the "turn-taking" condition in terms of deviation of the optimal moves. Thus, our findings provide initial indications regarding young children's need for exploratory actions in problem-solving process in HRI settings and the efficacy of those actions in challenging task performance.

5.2. Inhibitory Strategy Emergence and Development

The cognitive strategy of inhibition has been characterized as one of the main strategies used for the optimal solution of the ToH task (Goel and Grafman, 1995). This strategy allows the child to inhibit moves directly to the goal in order to make the counter-intuitive move that leads to the optimal solution. We identify the use of inhibitory strategy in all observed optimal moves excluding the moves leading to a subgoal or the final solution of the ToH. Our design allowed us to observe that this strategy is not apparent to all young children, especially in the more challenging tasks. However, the fact that our cases increased the speed of their optimal movements only toward the reach of a subgoal indicates that the analyzed children used additional strategies for the task solution such as implicit learning. Typically, this procedural learning is observed by continuous improvement in performance over repeated administrations of the same ToH problem, as shown by our analysis of the learning effect.

5.3. Designing Robot Behaviors to Scaffold Child's Exploration

For the current study we used the Haru robot with minimally designed social behaviors. Since our main focus was on the type and timing of robot cognitive intervention rather than on robot's social behaviors, on purpose, we restricted the robot behaviors into cognitive interventions providing suggestions in a neutral non-verbal manner and feedback related to the task performance only. Maintaining the same behavioral principles, we designed an "on demand" robot intervention. This is one of the few studies in HRI that provide children the space to voluntarily initiate the robot intervention. Our results indicate that there is a relationship between children's intrinsic motivation for exploration and robot intervention, since in many cases the participant children did not ask for help by the robot and preferred exploration which lead to increased task performance. Additionally, the "on demand" intervention allowed for inter-individual variability to be observed, with some younger children being inclined for more exploratory actions, which might require personalized robot interventions.

However, we observed that children's deviation from the optimal solution path in the specific task comes with certain patterns. From a pedagogical perspective, these patterns can be utilized in order for designers to develop *targeted robot interventions* which allow the child to explore and experience self-initiated interactions. In addition, at targeted instances of the task, the robot intervenes in order to provide recovery in child's actions and scaffold the child's problem-solving process which would lead to better learning experience and outcomes for the child.

This paper contributes to the field of HRI as one of the few developmental studies which focuses on the process rather than only on the final outcome of child's activity and provides indications about not only the *what* but the *why* of collaborative problem solving in child-robot interaction. Further, the suggestions of voluntary interaction contributes to the current dialogue about the ways we need to develop value-centered intelligent systems. In this way the child has the freedom to initiate the interaction according to her needs.

6. LIMITATIONS AND FUTURE WORK

Deeper insight into the trajectories of children's problem-solving will allow us to construct dedicated theoretic models for the emergence and development of children's complex strategies. In similar fashion to the work by Oudeyer and Smith (2016) on modeling curiosity development, in future work, we also intend to computationally model and simulate problem-solving processes of increasingly complex tasks. Toward this end, we intend to develop a robotic companion for dynamic assessment and support of children's tendency for exploration as one of the catalytic stages for the emergence and development of relevant cognitive strategies for problem solving. From a methodological perspective, whilst most of the current longitudinal studies with children in HRI include relatively small sample (i.e., Leyzberg et al., 2018), we aim to investigate child-robot collaboration in problem-solving tasks in a longitudinal study with a larger sample. In this way, we will be able to contribute to the dialogue regarding child development in HRI settings with generalizable results. In addition to this, we acknowledge that between the interaction design of the two conditions lie further possibilities for child-robot interaction in the context of collaborative problem-solving activities. Our plans for future work include additional possibilities for further types of interaction design.

Regarding the robotic system itself, we are currently developing a fully autonomous system for the dynamic assessment and autonomous robot intervention for the ToH task to carry out a larger scale study considering a fully autonomous interaction. This requires, from the perception part, to estimate the state of the game, the individual child problem-solving abilities and other individual characteristics. Tracking the state of the game makes it possible for the robot to automatically evaluate the task progress and thus take decisions accordingly.

Deeper data-driven analyses may further reveal characteristics and causes of child development and the transition from primitive cognitive and social actions toward more complex behaviors. As discussed before, all children did not have explicit conceptualized knowledge and strategies for problem-solving of the ToH task. So interacting with this task could be considered as a novel activity with many exploratory opportunities, which is still an open area of research for HRI. At the same time, it will be interesting to further investigate what design principles would be applied in developing robots that scaffold children to effectively transit from exploratory actions to intentional behaviors. Individual pace differences of this transition will require for the robot to be adaptively intelligent by dropping old solutions when something shifts in the child's behavior, the task or the context. This demands a dynamic approach of the conceptualization of problem-solving cognitive activity in child-robot cognitive collaboration. Taken together, our results are initial steps toward creating flexible autonomous agents that self-supervise in realistic physical environments by supporting human tendency for self-directed problem-solving activities.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the committee on the Use of Humans as Experimental Subjects of the Joint Research Centre of the European Commission. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin and all children assented to participate.

AUTHOR CONTRIBUTIONS

VC conceived, designed, executed the study, analyzed the data, and wrote the paper. EG executed the study, analyzed the data, and wrote the paper. LM and GM contributed towards robot development and technical support, executed the study, and wrote the paper. RG overviewed the study, contributed toward robot development, and wrote the paper.

ACKNOWLEDGMENTS

We thank the teachers and the participant children. We also thank M. Escobar-Planas, E. Moreno-Cruz, and A. Bolivar-Gonzalez for their contribution to the data annotation and analysis. This research was supported by the HUMAINT project of the Centre for Advanced Studies of the JRC, European Commission.

REFERENCES

Alves-Oliveira, P., Arriaga, P., Paiva, A., and Hoffman, G. (2017). "Yolo, a robot for creativity: a co-design study with children," in *Proceedings of the 2017 Conference on Interaction Design and Children* (Stanford, CA: ACM), 423–429.

Bandura, A. (1971). *Social Learning Theory*. Morristown, NJ.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: a review. *Sci. Robot.* 3:eaat5954. doi: 10.1126/scirobotics.aat5954

Best, J. R., and Miller, P. H. (2010). A developmental perspective on executive function. *Child Dev.* 81, 1641–1660. doi: 10.1111/j.1467-8624.2010.01499.x

Blakey, E., Visser, I., and Carroll, D. J. (2016). Different executive functions support different kinds of cognitive flexibility: evidence from 2-, 3-, and 4-year-olds. *Child Dev.* 87, 513–526. doi: 10.1111/cdev.12468

Brock, R., and Taber, K. S. (2017). The application of the microgenetic method to studies of learning in science education: characteristics of published studies, methodological issues and recommendations for future research. *Stud. Sci. Educ.* 53, 45–73. doi: 10.1080/03057267.2016.1262046

Cha, E., Matarić, M., and Fong, T. (2016). "Nonverbal signaling for non-humanoid robots during human-robot collaboration," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (Christchurch: IEEE Press), 601–602.

Charisi, V., Davison, D., Reidsma, D., and Evers, V. (2016). "Evaluation methods for user-centered child-robot interaction," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (New York, NY: IEEE), 545–550.

Charisi, V., Liem, C. C., and Gomez, E. (2018). "Novelty-based cognitive processes in unstructured music-making settings in early childhood," in *2018 Joint IEEE 8th International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob)* (Tokyo: IEEE), 218–223.

Cragg, L., and Chevalier, N. (2012). The processes underlying flexibility in childhood. *Q. J. Exp. Psychol.* 65, 209–232. doi: 10.1080/17470210903204618

Dewey, J. (1902). *The Child and the Curriculum*. Chicago, IL: University of Chicago Press.

Diamond, A. (2006). The early development of executive functions. *Lifespan Cogn.* 210, 70–95. doi: 10.1093/acprof:oso/9780195169539.003.0006

Flynn, E., O'Malley, C., and Wood, D. (2004). A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. *Dev. Sci.* 7, 103–115. doi: 10.1111/j.1467-7687.2004.00326.x

Friedman, N. P., and Miyake, A. (2017). Unity and diversity of executive functions: individual differences as a window on cognitive structure. *Cortex* 86, 186–204. doi: 10.1016/j.cortex.2016.04.023

Gelman, S. A. (2009). Learning from others: children's construction of concepts. *Annu. Rev. Psychol.* 60, 115–140. doi: 10.1146/annurev.psych.59.103006.093659

Gerstenberg, T., and Tenenbaum, J. B. (2017). "Intuitive theories," in *Oxford Handbook of Causal Reasoning*, ed M. R. Waldmann (New York, NY: Oxford University Press), 515–548.

Goel, V., and Grafman, J. (1995). Are the frontal lobes implicated in "planning" functions? interpreting data from the tower of hanoi. *Neuropsychologia* 33, 623–642. doi: 10.1016/0028-3932(95)90866-P

Gomez, R., Szapiro, D., Galindo, K., and Nakamura, K. (2018). "Haru: hardware design of an experimental tabletop robot assistant," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL: ACM), 233–240.

Hinz, A. M., Klavzar, S., Milutinovic, U., and Petr, C. (2013s). *The Tower of Hanoi: Myths and Maths*. Basel: Birkhäuser.

Hoffman, G. (2016). "Openwoz: a runtime-configurable wizard-of-oz framework for human-robot interaction," in *2016 AAAI Spring Symposium Series* (Palo Alto, CA).

Keen, R. (2011). The development of problem solving in young children: a critical cognitive skill. *Annu. Rev. Psychol.* 62, 1–21. doi: 10.1146/annurev.psych.031809.130730

Kennedy, J., Baxter, P., and Belpaeme, T. (2015). "The robot who tried too hard: social behaviour of a robot tutor can negatively affect child learning," in *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Portland, OR: IEEE), 67–74.

Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2016). "Social robot tutoring for child second language learning," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (Christchurch: IEEE Press), 231–238. doi: 10.1109/HRI.2016.7451757

Knoblock, C. A. (1990). "Abstracting the tower of hanoi," in *Working Notes of AAAI-90 Workshop on Automatic Generation of Approximations and Abstractions* (Boston, MA: Citeseer), 13–23.

Kory-Westlund, J. M., and Breazeal, C. (2019). A long-term study of young children's rapport, social emulation, and language learning with a peer-like robot playmate in preschool. *Front. Robot. AI* 6:81. doi: 10.3389/frobt.2019.00081

Lavelli, M., Pantoja, A. P., Hsu, H., Messinger, D., and Fogel, A. (2005). "Using microgenetic designs to study change processes," in *Handbook of Research Methods in Developmental Science*, ed D. M. Teti (Hoboken, NJ: Blackwell Publishing Ltd), 40–65. doi: 10.1002/9780470756676.ch3

Leite, I., McCoy, M., Lohani, M., Ullman, D., Salomons, N., Stokes, C., et al. (2017). Narratives with robots: the impact of interaction context and individual differences on story recall and emotional understanding. *Front. Robot. AI* 4:29. doi: 10.3389/frobt.2017.00029

Lemaignan, S., Jacq, A., Hood, D., Garcia, F., Paiva, A., and Dillenbourg, P. (2016). Learning by teaching a robot: the case of handwriting. *IEEE Robot. Autom. Mag.* 23, 56–66. doi: 10.1109/MRA.2016.2546700

Leyzberg, D., Ramachandran, A., and Scassellati, B. (2018). The effect of personalization in longer-term robot tutoring. *ACM Trans. Hum. Robot Inter.* 7:19. doi: 10.1145/3283453

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734

Montes, J., van Dijk, M., Puche, R., and van Geert, P. (2017). Trajectories of scientific reasoning: a microgenetic study on children's inquiry functioning. *J. Pers. Orient. Res.* 3, 67–85. doi: 10.17505/jpor.2017.07

Oudeyer, P.-Y., and Smith, L. B. (2016). How evolution may work through curiosity-driven developmental process. *Top. Cogn. Sci.* 8, 492–502. doi: 10.1111/tops.12196

Piaget, J., and Cook, M. (1952). *The Origins of Intelligence in Children, Vol. 8*. New York, NY: International Universities Press.

Ramachandran, A., Huang, C.-M., Gartland, E., and Scassellati, B. (2018). "Thinking aloud with a tutoring robot to enhance learning," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL: ACM), 59–68.

Saulnier, P., Sharlin, E., and Greenberg, S. (2011). "Exploring minimal nonverbal interruption in HRI," in *2011 RO-MAN* (Atlanta, GA: IEEE), 79–86. doi: 10.1109/ROMAN.2011.6005257

Schlagmüller, M., and Schneider, W. (2002). The development of organizational strategies in children: evidence from a microgenetic longitudinal study. *J. Exp. Child Psychol.* 81, 298–319. doi: 10.1006/jecp.2002.2655

Siegler, R. S., and Crowley, K. (1991). The microgenetic method: a direct means for studying cognitive development. *Am. Psychol.* 46:606. doi: 10.1037/0003-066X.46.6.606

Sills, J., Rowse, G., and Emerson, L.-M. (2016). The role of collaboration in the cognitive development of young children: a systematic review. *Child* 42, 313–324. doi: 10.1111/cch.12330

Simon, H. A. (1975). The functional equivalence of problem solving skills. *Cogn. Psychol.* 7, 268–288. doi: 10.1016/0010-0285(75)90012-2

Steinfeld, A., Jenkins, O. C., and Scassellati, B. (2009). "The OZ of wizard: simulating the human for interaction research," in *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (La Jolla, CA), 101–107.

Tomasello, M. (1995). "Joint attention as social cognition," in *Joint Attention: Its Origins and Role in Development*, eds C. Moore and P. J. Dunham (Hillsdale, NJ: Lawrence Erlbaum Associates Inc.), 103–130.

Twomey, K. E., and Westermann, G. (2018). Curiosity-based learning in infants: a neurocomputational approach. *Dev. Sci.* 21:e12629. doi: 10.1111/desc.12629

Van der Ven, S. H., Boom, J., Kroesbergen, E. H., and Leseman, P. P. (2012). Microgenetic patterns of children's multiplication learning: confirming the overlapping waves model by latent growth modeling. *J. Exp. Child Psychol.* 113, 1–19. doi: 10.1016/j.jecp.2012.02.001

Vygotsky, L. (1978). Interaction between learning and development. *Read. Dev. Child.* 23, 34–41.

Warneken, F., Steinwender, J., Hamann, K., and Tomasello, M. (2014). Young children's planning in a collaborative problem-solving task. *Cogn. Dev.* 31, 48–58. doi: 10.1016/j.cogdev.2014.02.003

Whitebread, D., Basilio, M., Kuvalja, M., and Verma, M. (2012). The importance of play. *Brussels: Toy Industries of Europe*.

Wijnen, F. M., Davison, D. P., Reidsma, D., van der Meij, J., Charisi, V., and Evers, V. (2019). Now we're talking: learning by explaining your reasoning to a social robot. *ACM Trans. Hum. Robot Interact*. 9:5. doi: 10.1145/3345508

Conflict of Interest: RG was employed by the Honda Research Institute, Tokyo, Japan, but no commercial or financial relationships were involved for this study.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 06 March 2020
doi: 10.3389/frobt.2020.00029

Iterative Design and Evaluation of a Tangible Robot-Assisted Handwriting Activity for Special Education

Arzu Guneysu Ozgur[1], Ayberk Özgür[1], Thibault Asselborn[1], Wafa Johal[2], Elmira Yadollahi[1], Barbara Bruno[1,3], Melissa Skweres[4] and Pierre Dillenbourg[1]*

[1] Computer-Human Interaction in Learning and Instruction Laboratory (CHILI), EPFL, Lausanne, Switzerland, [2] University of New South Wales, Sydney, NSW, Australia, [3] MOBOTS Group of the Biorobotics Laboratory (BIOROB), EPFL, Lausanne, Switzerland, [4] Center for Learning Sciences (LEARN), EPFL, Lausanne, Switzerland

OPEN ACCESS

Edited by:
Agnese Augello, Institute for High Performance Computing and Networking (ICAR), Italy

Reviewed by:
Ilaria Bortone, National Institute of Gastroenterology S. de Bellis Research Hospital (IRCCS), Italy
Marcos Faundez-Zanuy, Escuela Superior Politécnica, Universidad Pompeu Fabra, Spain

***Correspondence:**
Arzu Guneysu Ozgur
arzu.guneysu@epfl.ch

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: 28 November 2019
Accepted: 19 February 2020
Published: 06 March 2020

Citation:
Guneysu Ozgur A, Özgür A, Asselborn T, Johal W, Yadollahi E, Bruno B, Skweres M and Dillenbourg P (2020) Iterative Design and Evaluation of a Tangible Robot-Assisted Handwriting Activity for Special Education. Front. Robot. AI 7:29. doi: 10.3389/frobt.2020.00029

In this article we investigate the role of interactive haptic-enabled tangible robots in supporting the learning of cursive letter writing for children with attention and visuomotor coordination issues. We focus on the two principal aspects of handwriting that are linked to these issues: Visual perception and visuomotor coordination. These aspects, respectively, enhance two features of letter representation in the learner's mind in particular, namely the shape (*grapheme*) and the dynamics (*ductus*) of the letter, which constitute the central learning goals in our activity. Building upon an initial design tested with 17 healthy children in a preliminary school, we iteratively ported the activity to an occupational therapy context in 2 different therapy centers, in the context of 3 different summer school camps involving a total of 12 children having writing difficulties. The various iterations allowed us to uncover insights about the design of robot-enhanced writing activities for special education, specifically highlighting the importance of ease of modification of the duration of an activity as well as of adaptable frequency, content, flow and game-play and of providing a range of evaluation test alternatives. Results show that the use of robot-assisted handwriting activities could have a positive impact on the learning of the representation of letters in the context of occupational therapy ($V = 1,449, p < 0.001, r = 0.42$). Results also highlight how the design changes made across the iterations affected the outcomes of the handwriting sessions, such as the evaluation of the performances, monitoring of the performances, and the connectedness of the handwriting.

Keywords: handwriting, occupational therapy, tangible robots, iterative design, robots for education, haptic devices, interactive learning, special education

1. INTRODUCTION

Handwriting is a complex perceptual-motor skill consisting of visuomotor integration, motor planning, visual-spatial abilities, visual perception, as well as responsiveness to tactile and kinesthetic stimuli (Maeland, 1992; Amundson and Weil, 1996; Feder and Majnemer, 2007). It is a fundamental ability which has a great impact on a wide range of tasks such as communicating and recording our knowledge, emotions, ideas and opinions. Unsurprisingly, it has been shown that handwriting is a critical skill to be acquired for the academic and behavioral development of students (Berninger et al., 1997; Feder and Majnemer, 2007; Christensen, 2009). Hence, there is an ongoing research effort dedicated to empowering students with effective writing skills and highlighting the challenges students face to master handwriting.

In recent years, several studies have been conducted exploring the processes engaged in handwriting and the learning effects of different technologies on the handwriting process. Feder and Majnemer (2007) suggested that handwriting difficulties do not resolve without intervention. Considering that up to 25% of the school-aged population is affected by handwriting difficulties (Smits-Engelsman et al., 2001; Charles et al., 2003), there is a need to develop technologies that support intervention methods for typically developing and high-risk populations. One example where technology can be useful in this domain is the usage of digital tablets to detect handwriting difficulties. They made possible the evaluation not only of the final product of handwriting (the static image), but also its dynamics (Asselborn, T. et al., 2018; Zolna et al., 2019). For example, Pagliarini et al. (2017) used digital tablets to collect data on handwriting ability before handwriting is performed automatically. Thanks to quantitative methods, they could find patterns indicating potential future writing impairments at a very early age. Mekyska et al. (2016) used a supervised learning model to detect dysgraphia. The authors included 54 third-grade Israeli children in the study and used a 10-item Handwriting Proficiency Screening Questionnaire (HPSQ) (Rosenblum, 2008) to identify poor writing.

Rosenblum et al. (2019) in their study of handwriting investigated how certain low-level and high-level processes differ between children with ASD and typically developing children. Their findings have clinical implications which can inspire the development of technologies to help children with executive function deficiencies. These results indicate that the accurate assessment performed by therapists to identify the deficits and to determine the appropriate handwriting intervention customized to the individual have considerable importance.

In a related study, Asselborn, T. et al. (2018) focused on the detection of severe handwriting difficulties such as dysgraphia, using a digital approach that identifies and characterizes handwriting difficulties (Asselborn, T. et al., 2018; Zolna et al., 2019). Their approach was inspired by the original standardized test devised by therapists to detect handwriting difficulties. Their tablet-based test can have direct implications on developing educational technologies for children, either typically developing or with handwriting difficulties. Several other tablet-based applications can be found in the literature that remediate handwriting difficulties; the main advantages of these tablet-based applications is that they allow the display of additional visual information to provide immediate adaptive feedback and instructions to the learner, while capturing the handwriting data accurately to be processed in real-time or afterwards (Yamasaki et al., 1990; Lee and Lim, 2013).

Furthermore, a growing number of studies aim at helping children with developmental disorders by incorporating robots to help handwriting (Chandra et al., 2019; Kim et al., 2019). For instance, the Cowriter project (Hood et al., 2015; Chandra et al., 2019) exploits the social capabilities of a humanoid robot to teach handwriting in an original way. Based on the learning-by-teaching approach, the child becomes the teacher of a robot "requiring help" to improve its handwriting and this role reversal results in several powerful effects including motivation gain and de-dramatization of the child's problems.

From a learning goals perspective, in order to have a complete letter representation, a child should acquire the visual perception of the letter, called the *grapheme*, but also the visuomotor coordination associated with it, i.e., the dynamics of the movement, called *ductus* (Bara and Gentaz, 2011). To enhance the visual perception as well as the visuomotor coordination, it is shown that using more sensory information ranging from audio and visual to kinesthetic feedback is important (Hayes, 1982; Bluteau et al., 2008; Bara and Gentaz, 2011; Danna and Velay, 2015). Because of this reason, teachers commonly use techniques allowing children to experience various sensory information when learning how to write. These techniques include drawing letters in sand or semolina, touching and sensing the shape of letters carved in a piece of wood, verbally describing the letters or building the letter with play-dough (Berninger et al., 1997; Arslan, 2012).

Indeed, kinesthetic real-time feedback is shown to be paramount sensory information needed during the process of handwriting (Laszlo and Bairstow, 1984; Laszlo and Broderick, 1991). To fill this gap in robot-assisted and digital technologies, several recent studies are using haptically active training programs in order to teach handwriting. Bara and Gentaz (2011) compared a visual-haptic to a visual only program to teach five different letters to a group of 21 first-grade children. The authors showed that the combination of visual with haptic information is more efficient than visual only information since it improves both perceptual and visuo-motor skills.

Palluel-Germain et al. (2007) showed the use of visual-haptic feedback to teach handwriting to kindergarten children where they present a device, "*Telemaque*," that incorporates a programmable force-feedback pen that can be guided along a letter model (which is "*not only static (the shape) but also dynamics (rules of motor production)*") in order to enhance the visuomotor perception of the letters targeted. In their study, the authors focused on six cursive letters ("a," "b," "f," "i," "l," "s") and showed significant improvement of the handwriting's legibility for all trained letters after the visual-haptic training with respect to the control group.

Garcia-Hernandez and Parra-Vega (2009) proposed a haptic tele-operated training method aiming to improve motor skill acquisition. A master helps an apprentice by showing the desired path (a letter) using a robot end-effector, whose motion is sensed by the learner via the haptic device. The authors showed better and faster learning of motor control compared to the condition using visual information only.

Even though these devices have brought very promising results, a strong limitation to their widespread use comes from their very high cost, that makes them unaffordable for most schools. In addition and to the best of our knowledge, there is currently no haptic system providing collaborative handwriting activities in classrooms, which typically requires one set of equipment per learner. For this reason, one of the goals of this article is to present a system for teaching handwriting that relies on low-cost equipment, while also allowing haptic feedback in single- and multi-participant collaborative learning activities.

Collaborative learning appears in situations where two or more people attempt to learn something together (Dillenbourg, 1999). Even if no general assumption can be made concerning the benefits of collaborative learning (because it is strongly dependent on the designed activity), Kreijns et al. (2003) summarize the positive effects that sometimes arise with collaborative learning as a deeper level of learning, critical thinking, shared understanding, and long term retention of the learned material. Moreover, according to the therapists' feedback in the occupational therapy centers, children may benefit from group therapy sessions by modeling their peers, learning how to cooperate, acknowledging each other's strengths. Lastly, group occupational therapy or group physical therapy may provide beneficial social interaction to children: they can not only communicate their ideas with each other, but also improve their self-esteem by achieving skills and tasks in front of their peers. For these reasons, activities and tasks that are planned for the group session should be fun, flexible, exciting and novel as well as in line with the children's goals, preferences and attitudes to minimize the number of children who refuse to participate or exhibit non-compliant behavior[1].

Our research effort, described in the current and the previous studies (Asselborn, T.* et al., 2018), aims to enhance these sensory information by using the tangible, haptic-enabled, low cost, small-sized Cellulo robots (Özgür et al., 2017). While these robots move on a sheet of paper displaying the letter's visual representation (see **Figure 1A**), the learner can observe the ductus of the letter (the trajectory followed by the robot between the starting and ending points of the letter), as well as the grapheme of the letter (printed directly on the sheet of paper on which the robot moves). Moreover, the haptic and visual capabilities of the robots allow for increasing the sensory information provided to the learner during the activity. In this article, we hypothesize that training with the robot can effectively convey the procedural knowledge of the grapheme and the ductus of the letter in our context of interest. At the same time, using multiple robots and their synchronized behaviors, we aim to show that it is possible to design collaborative learning activities in the aforementioned fun, flexible and inclusive manner.

The primary aim of this article is to support handwriting learning, with a specific focus on special education, by designing tangible robot-mediated, interactive, collaborative activities. In previous work, we performed a content analysis to target specific skills involved in the handwriting processes and based on that designed an activity flow composed of 4 sub-activities, which was tested and validated in a public school. In this article, we refine and adapt the activity to a therapy context over a number of experiments in different therapy centers, with the close collaboration of therapists and children in need of occupational therapy.

During this iterative design process, we identify the key design aspects to be taken into consideration when addressing occupational therapy scenarios and evaluate the effect of the tested variants on learning using qualitative and quantitative methods. We discuss several key take-home lessons and conclude with shortcomings and future work.

2. MATERIALS AND METHODS

2.1. Cellulo Robotic Platform

Cellulo robots are low-cost, small-sized tangible mobile robots that can operate on printed sheets of paper covered with a dot pattern that enables fast (>90 Hz) and accurate localization (sub-mm) of each robot without any calibration (Hostettler et al., 2016; Özgür et al., 2017). This design allows for the recording of rich interaction-related information during the activity, such as user's motion trajectory, accuracy of the motion, etc. The robot's holonomic motion system provides autonomous motion capability, as well as robustness against human manipulation (Özgür et al., 2016). The overall design of the robot allows easy set-up and use in classroom and therapy environments thanks to the plug-and-play nature of its ecosystem. The proposed writing activity is composed of Cellulo robots and several shapes printed on paper sheets, displaying letters and cues related to the letter's ductus (see **Figures 1A,B**). The haptic, audio, visual and synchronization capabilities of the Cellulo robots allow us to provide real-time multi-sensory feedback during the handwriting task at the individual learner level as well as at the group level during collaborative handwriting activities. Lastly, each robot can be programmed to have a passive, active or semi-active role, which helps us design a pool of different activities where the role of the children can switch in between active and passive.

2.2. Iterative Design Methodology

The design of learning activities for children requiring special education within an occupational therapy session brings about many challenges and unknowns, such as orchestration, use of space and choice of grouping of children with vastly different learning objectives and activities. Furthermore, the design of activities for special education involves the crucial participation of a wide spectrum of stakeholders, including teachers, therapists and the children themselves. Given these factors, it becomes impractical to imagine a classical study scenario where a working design can be made and tested successfully to show that it yields positive learning outcomes: As we show below, there are typically many failures, lessons that must be learned and interactions that must be made with the stakeholders in order to improve the existing design and bring it to an acceptable level of operability and adaptability.

For these reasons, we opted to follow an iterative design methodology where we tested and improved the design repeatedly at different stages of maturity and practicality. At each iteration, we make/refine the design and attempt to verify it rigorously with a study in order to reveal flaws and gather observations that may aid in improving it further. First, we start our design by aiming to meet the learning objectives with healthy children in a typical school environment. This is labeled as Iteration 1, and aims to yield a base-level usable activity that we can iterate over; this iteration is previously

[1] https://www.yourtherapysource.com/blog1/2019/04/24/tips-for-successful-pediatric-group-therapy-sessions/ (accessed November, 2019).

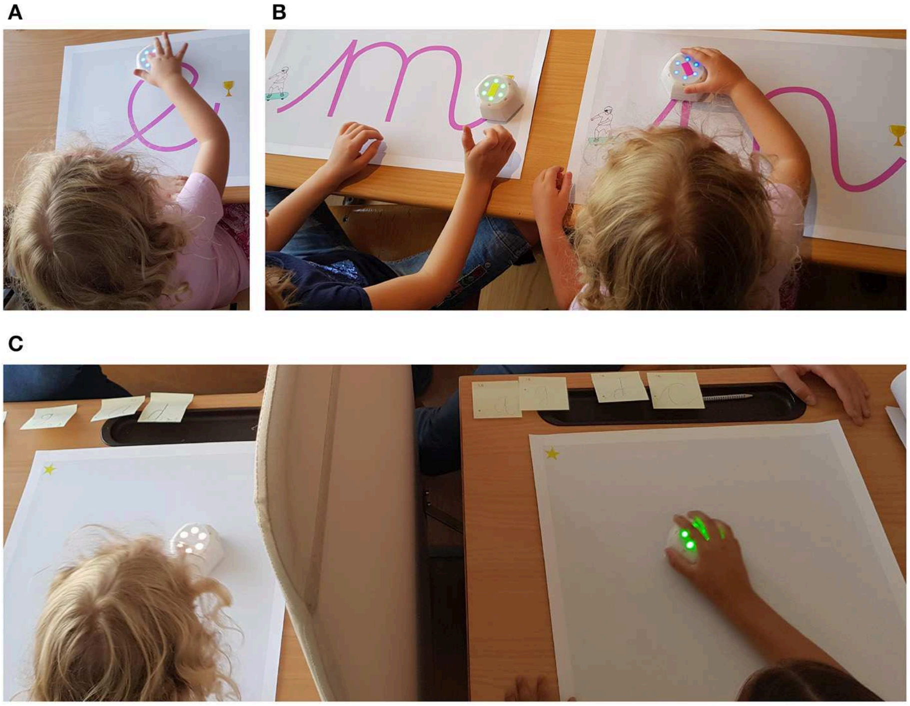


FIGURE 1 | Different sub-activities of tangible robot assisted handwriting activity in a therapy session. **(A)** Feel the robot, **(B)** drive the robot, **(C)** guess the letter game without grapheme.

published in Asselborn, T.* et al. (2018). This also allowed us to eliminate the usability flaws before launching the activity in a therapy environment. In section 3, each iterative step of our iterative design methodology is explained in detail to reflect the design changes, feedback and observations affecting the next steps.

2.3. Participants

During the iterative design process, one public school and two therapy centers were involved in our activity design and evaluation. Initial design and evaluation were done in the public school with the contributions of teachers and the participation of 17 healthy children with a mean age of 5.5. During the integration of the system into the therapy setting, therapists of each center gave feedback on the application before the testing stage. In the first center, we conducted one training session with 5 boys. In the second therapy center, we first conducted 3 training sessions with 3 girls and then 2 training sessions with 4 boys. The children attending the sessions had a variety of problems such as difficulty in concentration, fine motor dexterity issues, poor attention, etc. The detailed information about the symptoms and problems of each child indicated by the corresponding therapists of each group can be seen in **Table 1**. The clinical and neuropsychological assessment data belonging to the participant children are provided directly by the therapists. These assessments include proper clinical diagnoses (reported as ASD and ADHD), but also other potential problems related to handwriting observed by the therapists. These problems may simply be outside of clinical diagnosis scope (reported as e.g., "does not like to write") or may potentially eventually lead to the discovery of clearly diagnosable disorders in the child (reported as e.g., "motor coordination/activity problems"). In the latter case, the clinical diagnoses were not yet attempted on the children by their legal guardians. We opted to report all of these cases as they were highly beneficial in being the primary guiding factor in both the design phase and the application phase, i.e., when the actual interaction with the affected child took place.

Testing of our system was part of the three different summer school camps for fine motor and handwriting skills. These camps were aimed at helping with different aspects of handwriting and included varying activities to assist: (1) Core body strength and shoulder stability, (2) Body posture and hand positioning, (3) Manual dexterity and pencil grasp, (4) Fluidity of writing movements, (5) Handwriting legibility, (6) Typing, (7) Sensory awareness, (8) Graphomotor skills, (9) Concentration and attention, (10) Social skills. In the second therapy center, therapists were also providing support for the development of gross motor skills with outdoor activities.

This study was carried out in accordance with the recommendations of the Human Research Ethics Committee (HREC) at EPFL. The protocol was approved by the HREC (No. HREC 008-2018/16.02.2018). All subjects' parent or legal guardian gave written informed consent in accordance with the Declaration of Helsinki. All child participants gave a recorded assent and were informed of their right to stop the experiment at any time.

TABLE 1 | Child participants to occupational therapy sessions.

Group	Child id	Age	Symptoms or problems indicated by the corresponding therapists
Group 1	F	7	ASD, losing motivation quickly, problems in visual construction, does not like to write
	A	7	ADHD, attention problem, sensitivity to auditive stimulation
	X	6	ADHD, attention problem, does not like to write
	B	7	Visuomotor coordination problems, poor fine motor dexterity, problems in line following
	V	7	Visuomotor coordination problems, poor fine motor dexterity, problems in line following
Group 2	J	7	Handwriting problems, poor fine motor skills, poor precision, functional problems, high intelligence assessment, moves a lot and is disturbed quickly, poor concentration
	C	8	Problems in fine motor skills, poor attention, not totally concentrated, focused or engaged while handwriting
	I	5.5	Poor gross and fine motor skills, robot activity is first experience with cursive letters
Group 3	O	7	Problems in handwriting skill and fine motor activity, difficulty in visual perception, line following and drawing
	S	7	Handwriting problems
	K	8	Hyperactive, sensory problems
	M	7	High potential, fine motor skill difficulties, handwriting problems, hyperactive

2.4. Data Analysis

In order to explore the added value of our robot-assisted writing activities to the handwriting learning process, we want to assess the visual perception (representation of the letter's grapheme) and the visuomotor coordination (representation of the letter's ductus) aspects of the learners in detail. In other words, we want to assess the quality of the letter representation in the child's mind in terms of ductus and grapheme.

Children participated in each activity session in the following way: First, they did a pre-test with a pressure-sensitive pen & tablet (Wacom Cintiq Pro in the public school, Lenovo ThinkPad X1 Yoga in the therapy centers) in order to measure their handwriting proficiency before the activity. Then, they participated in the tangible robot-enabled activity, namely the main writing session with the Cellulo robots. Finally, they did a post-test in a similar way to the pre-test to measure their progress after our activity.

Initially, we asked experts to grade each letter from every child in terms of the ductus quality between 0 (for totally wrong ductus) and 3 (perfect ductus with proper start and end points and directions) but the inter-rater agreement of the experts was found to be too low. One of the contributing factors was the high variance between the hand writing performance (ductus, grapheme and cursiveness quality) of children in therapy centers. Another was that during the initial phases of the experiment, tests were mis-perceived by some children who started to fill the letter graphemes as if it was a line following (in the form of painting) activity rather than a writing activity. Ranking between 0 and 3 was also not reflecting the improvement in writing performance of children who were previously unable to write at all: There were instances where the pre-test performance was not gradeable (no sensible letter was written) and the post-test performance was very low but comparably closer to actual writing. Some learning clearly took place, but both performances received 0 rank.

In order to reliably quantify the letter writing performance by focusing on grapheme and ductus quality, we switched to the Dynamic Time Warping (DTW) technique from Salvador and Chan (2004) (available as a python package under the name of fast-dtw) which allows measuring the distance between two temporal sequences regardless of the speed. Using this technique, we measured the distance between the written letters [taken as an actual time series (x, y, t)] and the ideal letter represented on the activity sheets [taken as an ideal imaginary time series (x, y, t)], which is taken as a factor contributing to performance. This distance is used as an error score for writing performances. For a given letter, a lower error score indicates a closer ductus and grapheme to the expected letter. Furthermore, we calculated the connectedness of the letters (defined as the total number of strokes per letter) as another factor contributing to performance, in order to take into account the possible mis-perception effect mentioned above.

3. ITERATIVE DESIGN OF THE ROBOT-ASSISTED WRITING ACTIVITY

This section explains in detail each step through the iterative design process, starting from the pedagogical design, followed by the various steps of testing in the school and therapy centers, and the adaptation of the system to the new learning environment. The overall flow of the iterations and corresponding group information can be seen in **Figure 2**.

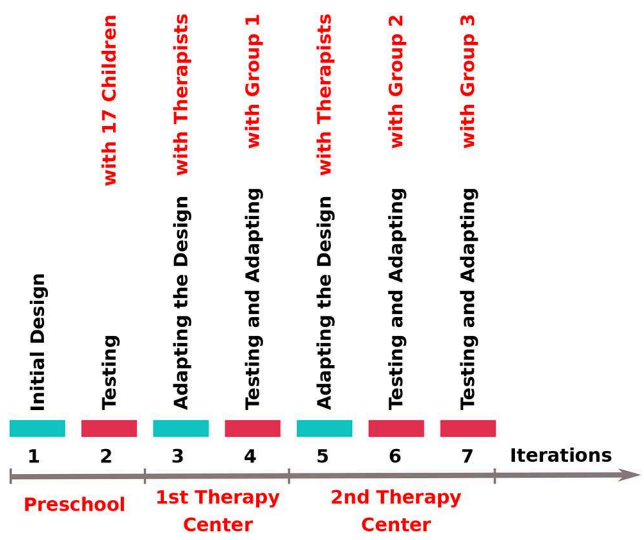


FIGURE 2 | The timeline of the iterative design and testing steps. Group 1, Group 2, and Group 3 are child participants in occupational therapy sessions in **Table 1**.

3.1. Initial Design of the Letter Writing Activity: Iteration 1

3.1.1. Pedagogical Design

In the initial design, previously published in Asselborn, T.* et al. (2018), our focus was on enhancing the knowledge of the grapheme and the ductus of the letter which are correlated with the visual perception and the visuomotor coordination. The content analysis done to determine the specific skills involved has led us to define the following sub-goals:

- *Remembering the Grapheme*: Memorizing the letter's physical representation (Free Recall and Recognition).
- *Remembering the Ductus*: Memorizing the letter's drawing pattern (Imitation).
- *Remembering the Phoneme to Ductus-Grapheme Link*: Memorizing the link between the letter's pronunciation (phoneme) and the corresponding grapheme and ductus.

It is shown that using more sensory information ranging from audio, visual to kinesthetic feedback enhances visual perception as well as the visuomotor coordination (Hayes, 1982; Bluteau et al., 2008; Bara and Gentaz, 2011; Danna and Velay, 2015). Precisely because of this, teachers use techniques allowing children to experience various sensory information during letter learning such as using sand filled boxes for drawing letters in; touching and sensing the grapheme of letters craved in a piece of wood or plastic surface[2]; or building the letter with play-dough or with similar materials that can be shaped by hands (Berninger et al., 1997; Arslan, 2012). There also exist sensory play games used in therapy centers such as *draw on your back* game. Each child takes turns with the teacher or therapist in drawing with their finger on the other's back. The main goal is to try to guess what the other person is drawing or writing. The level of difficulty is easily adjusted by modifying what is drawn - starting with shapes for young children, progressing through letters of their name, numbers, and so on[3]. The design of our robot-mediated activity is inspired from these traditional methods that are already used in classrooms, as well as from discussions with school teachers and therapists on how we can position Cellulo in handwriting activities.

3.1.2. Activity Design

We decided to use three features of the robot, namely haptic information, autonomous motion and synchronized behavior of multiple robots, to increase the multi-sensory feedback via touch, motion and sight. Haptic features allow each child to receive individual real-time feedback, autonomous motion makes the robot reproduce the ductus while synchronization allows collaborative game design. With this in mind, we designed the following sequence of sub-activities:

- Sub-Activity 1: Link between grapheme and ductus - *Watch the Robot* - In this activity, we aim that the child learns the letter's ductus by watching the robot moving on a map with the grapheme of that letter. The robot performs the dynamic that should be done while writing with its autonomous motion as the first representation of the letter's ductus. In addition, the letter's phoneme is generated at the beginning and the end of the writing process, to strengthen the link with the corresponding grapheme. The robot's rotary LEDs turn red and blinking, in order, one by one as a progress indicator while the path is followed, and turn solid green when the end point is reached.
- Sub-Activity 2: Link between grapheme and ductus - *Feel the Robot* - While the child watches the robot only in the first activity, we add another representation of the letter's ductus in this second activity by asking the child to put their hand on the robot while it is drawing the letter. The child does not actively move the robot, but only follows its autonomous motion in a passive way. **Figure 1A** shows an example screenshot of Feel the Robot activity where the child follows her robot with her index finger, while it is performing the ductus of letter "e."
- Sub-Activity 3: Memorizing the ductus of the letter - *Drive the Robot* - In this activity, the child actively drives the robot in order to produce the ductus of the letter. The grapheme of the letter is drawn on a map as seen in **Figure 3A**, the design of which includes a car racing theme with the start and end points that the writing should follow. Each child moves with their own speed since the robot is in passively drivable mode. The robot provides assistive haptic feedback by moving the child's hand toward the expected path if the child moves away from it. In order to discriminate the active and passive roles of the children in sub-activity 2 and 3, we assigned different colors to the LED's while robot is on the path. The robot's LEDs are blue while the correct path is followed, turn red if it is out of the letter path and turn green when the end point is reached. These

[2] Such as the one in https://www.etsy.com/listing/453872176/cursive-alphabet-wood-tracing-board (accessed September, 2019).

[3] Described in https://childhood101.com/sensory-play-ideas-games-to-develop-the-sense-of-touch (accessed September, 2019).

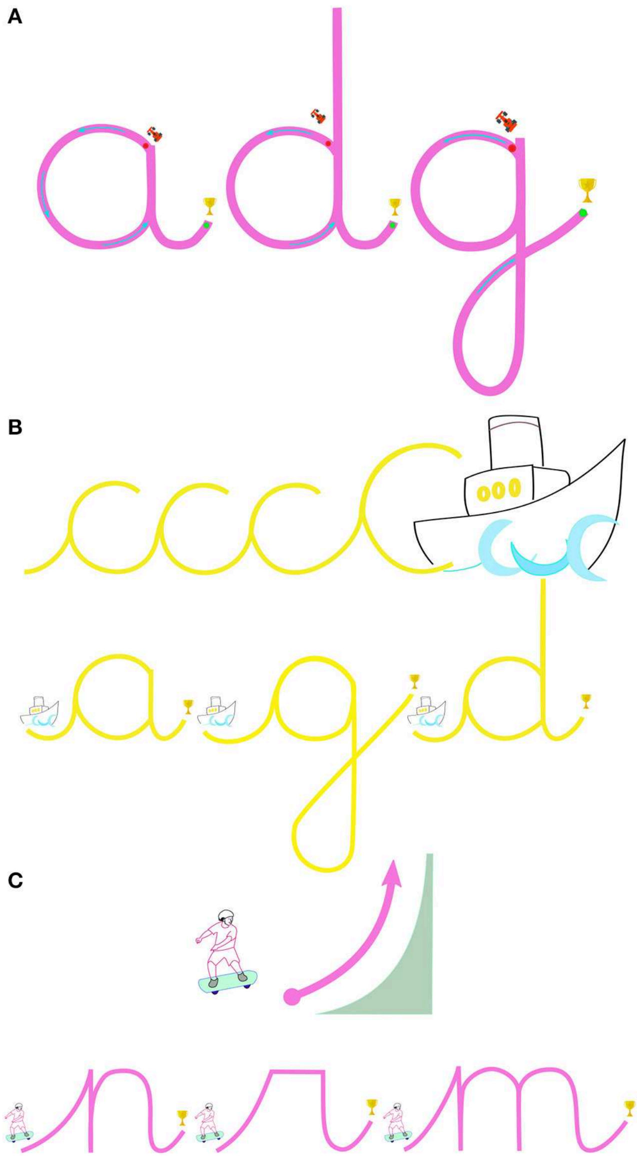


FIGURE 3 | (A) Includes round letter maps with the racing theme, letter "a" used in the public school (Iterations 1 and 2) and all three letters are used in first therapy center (Iterations 3 and 4). **(B,C)** Are cursive letters and their initial strokes for practice used in the second therapy center (Iterations 5, 6, and 7). **(B)** Includes examples of wave letters, **(C)** includes examples of skateboarding letters. Adapted from the ABC Boum + teaching approach of graphomotricity.

feedback elements condition the child to recognize errors, and serve as extrinsic motivation for drawing correctly. **Figure 1B** shows an example screenshot of Drive the Robot activity where the child on the left reached the end of the letter "m" (the robot's LEDs turn green) and the child on the right drives her robot on the correct path (the robot's LEDs are blue).

- Team Activity: Recalling grapheme by watching ductus - *Guess the Letter* - In this team activity, children form groups where one child takes turns at drawing a letter with a robot. Each time, the other children have to guess which letter is being drawn. In the group, the two guesser children sit together, with the writer separated from the other two by a physical barrier in order to ensure that they cannot see each other. The writer has one robot, and the two guessers have one robot (or one robot each depending on the size of the workspace) that reproduces whatever movement the first robot performs. In the beginning of the activity, the writer is shown (privately) the map of the letter that indicates only the grapheme, which they then have to draw with their robot. The other children watch their robot reproduce the letter drawn on their empty map. Then, the guessers have to choose the correct letter by recalling the letters they learned or selecting among given graphemes. An illustration of this activity can be seen in **Figure 4** where the two children on the left of the barrier are the guessers and the child on the right side is the writer.

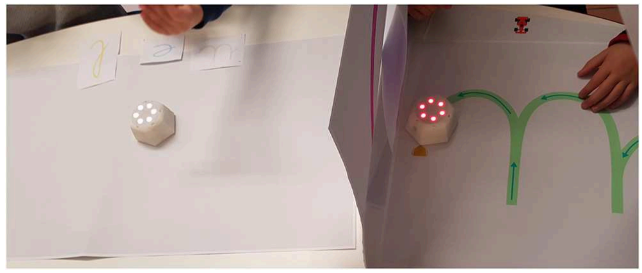


FIGURE 4 | Guess the Letter Game with Grapheme: The children on the left side of the barrier are the guessers and the child on the right side is the writer who just finished writing 'u' with the map having grapheme of the letter and waiting for the guessers to guess the written letter.

3.1.3. Performance Evaluation Design

In order to explore the added value of our system to handwriting learning, the visual perception and the visuomotor coordination aspects should be assessed in detail. Therefore, we focused on assessing the quality of the letter representation in the child's mind in terms of ductus and grapheme. Three sub-skills mentioned above are evaluated in a software application developed in Python that runs on a graphic tablet (Wacom Cintiq Pro). The use of the graphic tablet allowed us to save various data concerning the child's handwriting: The x and y coordinates of the pen were recorded as well as the pressure and the pen tilt for every time frame at a sampling rate of 200 Hz.

- *From Phoneme to Grapheme & Ductus*: In this test, we aimed to assess if the child remembers both the grapheme and the ductus of the letters: The child hears the phoneme of a letter (upon pressing button 1 in **Figure 5A**) and is asked to draw the grapheme on the tablet. As the link between the grapheme and the phoneme of the letter might not yet be fully operational, we offer the child the possibility to see the grapheme of the letter (only the grapheme and not the ductus) during 1 s, upon pressing button 2. As the child might want to have access to the grapheme even though they have the representation of the letter in their mind (just to make sure they are writing correctly or to ameliorate the letter), we ensured throughout the test

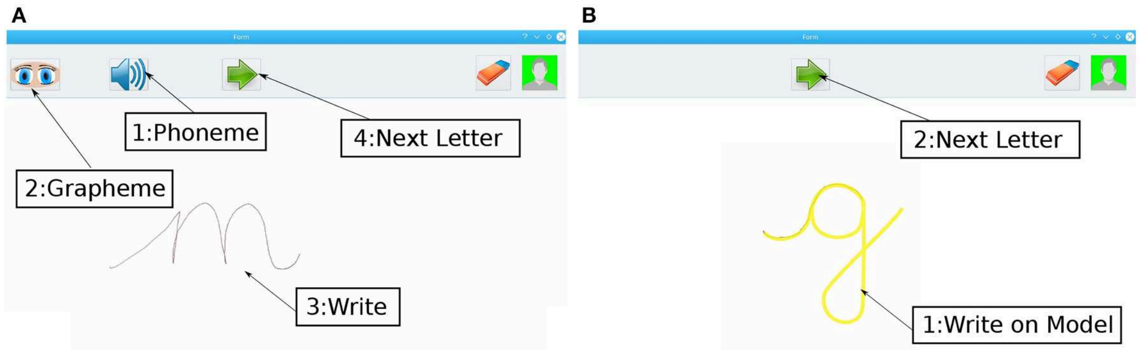


FIGURE 5 | Pre/post-test software: **(A)** Test without letter grapheme to assess the link between the phoneme of the letter and its associated grapheme and ductus. By pressing button #1, the child hears the phoneme of the letter. With button #2, the child has access to the grapheme of the letter during 1 s. #3 is the grapheme drawn by the child. Once finished, button #4 is used to save the data and move to the next letter. **(B)** Test with letter grapheme to assess the link between the grapheme of the letter and its associated ductus. #1 is the letter drawn by the child on the grapheme. Once finished, button #2 is used to save the data and move to the next letter.

that they can press the button only if they have not memorized the grapheme of the letter at all. Since the model of the letter grapheme is not given as default in this test, we referred this pre/post test as *test without grapheme* in this paper.

- *From Grapheme to Ductus*: This test is aimed to evaluate the grapheme-ductus link: The letter's grapheme is displayed on the tablet's screen (see **Figure 5B**), and the child is expected to draw the letter directly on top of the grapheme. The specific path between the start and end points of the letter is assessed during the test.
- *From Phoneme to Grapheme*: The goal of this test is to evaluate the visual perception which helps the child to find the right grapheme after hearing the phoneme of a letter among other letters. Concretely, the child has to press a button to hear the phoneme of a letter and find the associated grapheme among a choice of given letters.

3.2. Initial Testing in Public School: Iteration 2

With the activity and evaluation design done in Iteration 1, initial experiments were conducted with 17 five-year-old children in a public school. The students were split in two learning groups in order to explore the potential benefit of teaching sessions involving the robots compared to teaching session run with more traditional methods. Furthermore, research was done to inspect how these two teaching methods (with the robots and with traditional methods) can be combined together.

Results show a clear potential of our robot-assisted learning activity, with a visible improvement in certain skills of handwriting, most notably in creating the ductus of the letters, discriminating a letter among others and in the average handwriting speed. Moreover, we show that the benefit of our learning activities to the handwriting process increases when it is used after traditional learning sessions. These results were previously published in Asselborn, T.* et al. (2018) in detail; in this paper, we only focus on the insights and observations contributing to future design. Notably, we received the following feedback:

- *Difficulty of Feel the Robot*: The children were frequently having problems in doing this activity due to excessive downward force they applied to the robot which blocked its motion. This required the experimenters to intervene and show the child the proper way to do the activity. Even though initially we decided to abandon this sub-activity in the future designs, discussions with the therapists revealed that the feedback loop provided by the robot not moving while the child is applying too much pressure could be useful for conditioning some children in reducing this pressure. More detail is provided in the corresponding iteration description below.
- *Pre/post-test duration*: Even tough inspecting the learning performance for each learning goal is crucial, collecting data through several pre/post-tests, which must be done for each child participating to the sessions, were observed to be very time consuming. Due to this, we decided to adopt fewer, more focused learning goals in order to be able to design shorter evaluations per session in the future.
- *Confusing visual feedback color*: During the activities, it was observed that one particular child became mad at her robot since it was giving red visual feedback during the View the Robot sub-activity. She said that her robot is misbehaving and not working properly, rightly on her part, since the Draw the Robot sub-activity uses red color to create the negative reinforcement feedback. It was a usability flaw which was fixed for the subsequent iterations.

3.3. Adapting the Activity to Occupational Therapy: Iteration 3

3.3.1. Overview

The re-design process comprised of a number of successive iterations with the participation and feedback of several therapists from 2 different therapy centers during 3 summer school activities including multiple groups of children. In this section, the adaptation of the activity to the first therapy center is described, which started with preliminary meetings with therapists in order to do the adaptations specific to the therapy center's teaching methodologies and learning objectives. Taking into account the specific stage the child participants and the therapists were at during this time, it was decided to work on round cursive letters "a," "d," and "g." The previously designed racing theme was kept, as can be seen **Figure 3A**.

3.3.2. Change in Context, Frequency and Pre/Post-tests Due to Time Limitation

The principal change was on the total time of the activity and limitation of the time spent on pre/post-test evaluation: We decided on less repetition on Watch the Robot, Feel the Robot and Drive the Robot sub-activities and on using only one of the pre/post-tests that focuses on ductus learning evaluation. We chose the From Grapheme to Ductus test since it showed the most clear contribution of our system during Iteration 2.

3.3.3. Removing the Grapheme From the Guess the Letter Activity to Focus on the Goal of Recalling Grapheme & Ductus

Another crucial change done was with the Guess the Letter game activity where the grapheme was removed on the writer side to force the child to remember the letter grapheme, which was hypothesized to be more effective to learn the letter compared to providing the grapheme. Variation between two maps can be compared by checking with grapheme version in **Figure 4** and without grapheme version in **Figure 1C**. This adaptation does not change the learning objective for the guessers but changes the learning objectives for the writer by contributing to the final goal of our learning objective: To encourage the child to remember both grapheme and ductus. If the writer cannot write the letter properly, by definition the guessers cannot guess correctly. This becomes a feedback mechanism for the writer to rewrite the letter by paying better attention to the writing process. Since the grapheme is not there anymore, it further allows us to track the progress of the child through the writing trajectory data which does not necessarily follow the correct path. Some example trajectory results of this feedback mechanism are given in section 4.2.

3.4. Testing in the First Therapy Center: Iteration 4

The activity is tested within the first day of the summer school with 5 boys for 1 h. The information related to this group can be seen in **Table 1** - Group 1. We encountered a number of problems in the pre/post-test application, gathered observations that highlight the added value of the activity, and feedback from the therapists, which are summarized as follows:

- *Problem of sequential testing design*: The activity is started with testing with the grapheme. After second child's tests, the rest of the group were bored of waiting for 10 min and the pre-test was not completed. The sequential design (pre-test one child at a time) did not work due to the limited attention span of the target child group. For the following summer school sessions we decided to do the testing while the other children doing another group activity and not waiting for each other.
- *Added value of the Feel the Robot activity for sensing self-applied force*: Child F has the problem of discriminating the relationship between his touch sensation and visual perception. Therapists indicated that Feel the Robot activity is very useful for children having such problems to train on exerting the right amount of force by improving the connecting between sight and touch sensations. As it is observed with Child F, while the robot was blocked by putting too much force on it, in order to observe the motion of the robot, the child was encouraged to balance and reduce this force. In doing so, he was training in controlling it.
- *Motivation and engagement*: The overall group motivation was observed to be high and the attentive time spent on our activity was observed to be longer compared to other writing activities. In particular, the total time child F was attentive was considerably high according to therapists, since he does not like to write and he did not previously focus on a writing task for such a long period of time. He was observed to be highly willing to write with Cellulo and he readily completed all of the tasks.

3.5. Adapting the Activity to the Second Therapy Center: Iteration 5

3.5.1. Overview

Apart from the necessities of integration to the occupational therapy environment, it was observed in the previous iteration that there may be a need for further adaptation to each therapy center to be compatible with their learning methodologies. In this iteration, besides doing this, we also integrated the previously suggested changes by therapists to our activity design and flow. These are discussed below.

3.5.2. Re-designing Visual Cues to Be In-line With the Present Teaching Methodology

The main change in this iteration was to adapt the cursive letter shapes and visual cues on the map designs as the way of teaching cursive letters differed from the previous therapy center: The letters were adapted to also include the connecting strokes from the previous (imaginary) letter as the initial stroke, which was not previously present in our design.

Furthermore, the methodology designed by the ABC Boum+ company[4] was adopted as it was being used in the therapy center. This method combines visual, conceptual and sound cues with the initial strokes of the cursive letters in order to reinforce the learning, where the letters are divided into different conceptual groups according to their initial strokes. We designed new maps with these new cues instead of the car racing theme, except the trophy icon at the end which was kept. We also designed new maps consisting of only the cues to teach the initial stroke of the corresponding letter group. Three example map designs from "wave letters" and "skateboarding letters" groups and their corresponding initial cues can be seen in **Figures 3B,C**, respectively.

[4] ABC Boum teaching approach of graphomotricity, https://abcboum.net (accessed September, 2019).

3.5.3. Knowledge Transfer From "Large Letters With the Robot" to "Small Letters With the Pencil"

In order to reinforce the ductus and grapheme representation learning, therapists suggested to add writing activities with a pencil and post-it after each letter practiced with the robot. The second reason for this addition was to switch between a gross motor activity to a fine motor activity to help mapping the learned shape to actual handwriting practice. This allowed us to confirm whether the writing practice with the robot, with the pencil and with the ThinkPad pen are similar or not. See section 4.4 for writing performance comparison and discussion.

3.6. Testing in Second Therapy Center: Iteration 6

The activity was tested during the each day of the first summer school (3 days) with 3 girls. The information related to this group can be seen in **Table 1** - Group 2. Our findings are as follows:

- *Effect of summer school context on engagement*: For the first day of the activity, 4 letters were selected, which made the activity length roughly 50 min in total. This duration was quite long in comparison with the duration of the other activities within the summer school. Since it was a summer school including both gross and fine motor skills, there were several active game-play sessions including games in the playground, jumping, climbing etc. Within this context, the duration of the activity with the robots played a crucial role to keep the attention and engagement of the children stable. For instance, child J did not want to continue the activity because the other activities were more fun in her view. For this reason, from the second day on, we reduced the number of letters to 3 in the activity, which reduced its length to roughly 40 min.
- *Group therapy including children with different abilities*: Child I (age 5.5) was younger than the rest of the group (mean age 7) and she had not received any lesson on cursive writing before our activity. Even though it was her first time writing cursive letters, she was observed to perform well albeit with the help of graphemes provided to her on a post-it, which was not given to other children. Child C has problems with fine motor skills, attention, and organization. Therapists indicated that sometimes while writing she is not totally concentrated or focused. She engaged a lot during the robot mediated activities and liked the game. Each day she wanted to continue to do more exercise with the robots.
- *Pre/post-test mis-perception*: The most discriminative test previously used (writing on top of a grapheme, i.e., From Grapheme to Ductus) that showed best the progress of the healthy children in the school context did not work with some children in this therapy center. These children did not understand the relationship between these letters and the activity, and proceeded to fill the letters as if it was a line following task (see **Figure 6** for sample data): The graphemes on the screen were not perceived as a letter writing grapheme but as a line to be followed and/or an area to be painted. Therapists suggested that if there is no stable grapheme in the test, it might be easier for the children to avoid this confusion. Therefore, from the second day on, we switched to the pre/post-test without a stable grapheme where grapheme was made to appear on the screen for only 1 s after pressing the grapheme cue button (i.e., From Phoneme to Grapheme & Ductus).

3.7. Second Testing in Second Therapy Center: Iteration 7

The activity was further tested during the first 2 days of another summer school approximately 1 week later, with 4 boys. The information related to this group can be seen in **Table 1** - Group 3. We found that:

- *Difficulty of changing the letter maps during the session*: Even though the therapists found the activity useful for children, it was observed that it is difficult for one single teacher/therapist to control the whole activity flow including several letter maps in a session with 4 children. For future use in such group sessions, they proposed using large, thick sheets of paper or paper sheets attached to thin wooden blocks for further ease of changing maps.
- *Need for practice in recalling the grapheme*: Most appreciated feature of the activity by the therapists was having separate sub-activity alternatives, with and without grapheme. It was suggested that a version of the Drive the Robot without the grapheme (i.e., empty map, similar to how it is done in the pre/post-test and the Guess the Letter game) should be added alongside the one with the grapheme, in order to provide an exercise in recalling the grapheme of letters.
- *Loss of motivation due to passive tasks*: In the second day, one child was observed to lose engagement in the Guess the Letter game while he is in the role of guesser which resulted in inattentive and random guessing, as he indicated that he would like to play the writer role instead. This implies that, to be more robust against such cases, more variants should be done to ensure that every sub-activity and every role could be tweaked to include active participation.
- *Need for repeated sessions*: Since children learn script letters before the cursive letters at school, the change could be difficult for them, as expected. Therapists reported that indeed more repetition of the sessions is needed before the ductus knowledge could be fully integrated.

After this session, we also had the chance to get the children's feedback on which part of the game they like the most and the least:

- Child K: He enjoyed every part of the activity, particularly the Guess the Letter game.
- Child O: He enjoyed Guess the Letter game the most and Drive the Robot the least.
- Child S: He enjoyed drawing the letter during Drive the Robot the most, while he enjoyed the rest of the activity in general. He was very attentive during both days and even named his robot.
- Child M: He enjoyed the writer role in the Guess the Letter game the most, but enjoyed the guesser role the least.

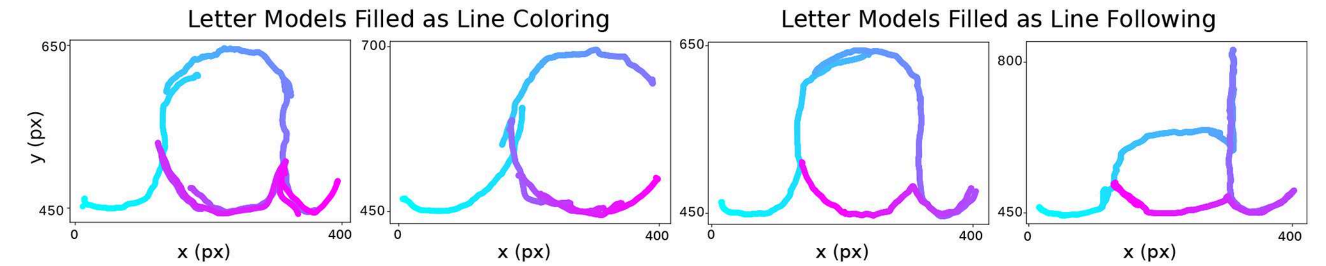


FIGURE 6 | Mis-perception of the letter models, the time dimension in the data is indicated by the color of the stroke going from light blue to light pink: Some children filled the grapheme of the letters as if it is a line coloring/painting activity (on the left, note the strokes going back and forth), or a line following activity (on the right, note the strokes following the grapheme's lines continuously but not in the correct ductus) in the pre/post-test with the grapheme.

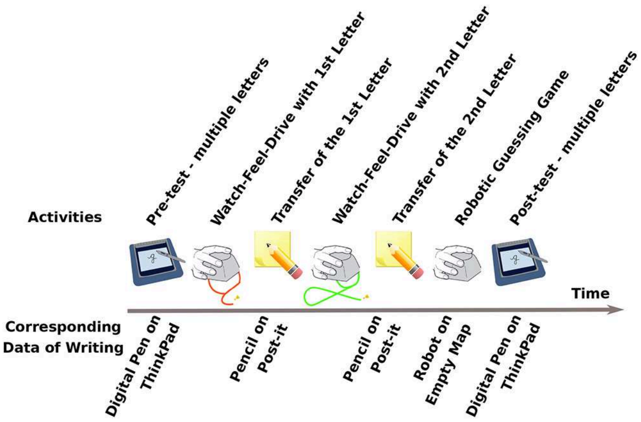


FIGURE 7 | An example activity flow of a robot-assisted writing session with two letters and corresponding handwriting data collected with a different medium in each step.

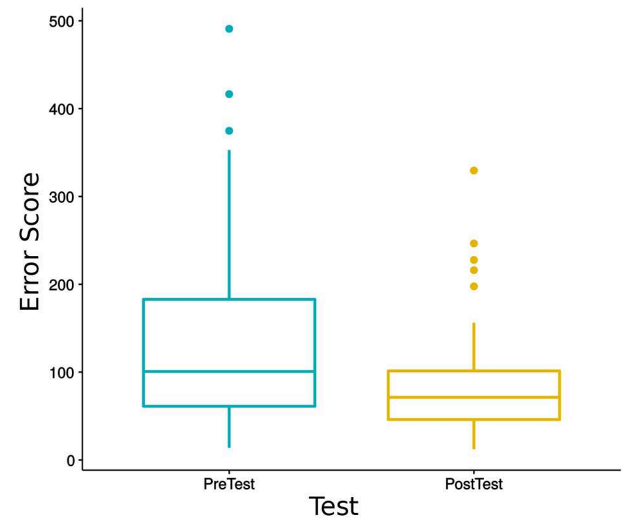


FIGURE 8 | Comparison of DTW error scores of all children for pre-test and post-test. We found that post-test error scores are significantly lower than pre-test error scores ($V = 1449, p < 0.001$).

4. RESULTS

Our iterative design approach allowed us to successfully integrate our robot assisted writing activity into the occupational therapy center by adapting the activity as well as the evaluation methods for different therapy centers and groups. During this process, a number of different letters are practiced with our robotic platform during several sessions. In this section, we first investigate the effectiveness of our activity in teaching the child participants to write letters. Second, we focus on the effect of changes made during the iterative processes on the writing performance. An example activity flow of a refined robot-assisted writing session and corresponding handwriting data collected with different mediums in each steps can be seen in **Figure 7**.

4.1. Overall Learning

During each iteration, before and after the learning session, a pre/post-test is done to measure the progress in letters learnt during the sessions. To analyze if there was overall learning in writing letters for all sessions for all children, we did a Wilcoxon Signed-Ranks Test, which indicated that post-test error scores were significantly lower than pre-test error scores [$V = 1,449, p < 0.001, r = 0.42$ (moderate effect size)], see **Figure 8**.

Since two types of pre/post-test evaluation are used to measure overall learning, we also checked for per-test learning by doing two separate Wilcoxon Signed-Ranks Tests. In the data collected with *the test with grapheme* during Iteration 4 and the first day of Iteration 6, we found a significant decrease in error scores of post-test compared to error scores of pre-test [$V = 160, p < 0.05, r = 0.45$ (moderate effect size)]. Similarly, in the data collected with *the test without grapheme* during day 2 of Iteration 6 and Iteration 7, we found significant decrease in error scores of post-test compared to the pre-test [$V = 681, p < 0.01, r = 0.65$ (large effect size)], see **Figure 9**.

In order to test if there is any significant difference between the average performance of the children in the 3 experimental groups, a Kruskal-Wallis Test was done which revealed no significant difference between groups ($H = 0.17, df = 2, p = 0.92$).

Similarly, we checked child-level difference in overall data including pre/post-test error scores with Kruskal-Wallis Tests and found significant difference ($H = 18.91, df = 8, p < 0.05$).

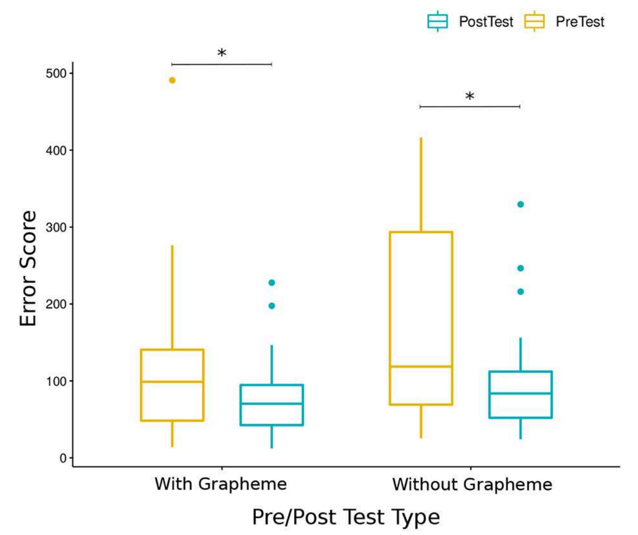


FIGURE 9 | Comparison of DTW error scores of all children for pre-test and post-test using the test with grapheme and the test without grapheme. In both with and without grapheme tests, post-test error scores are significantly lower than pre-test error scores ($V = 160, p < 0.05, r = 0.45$ and $V = 681, p < 0.01, r = 0.65$).

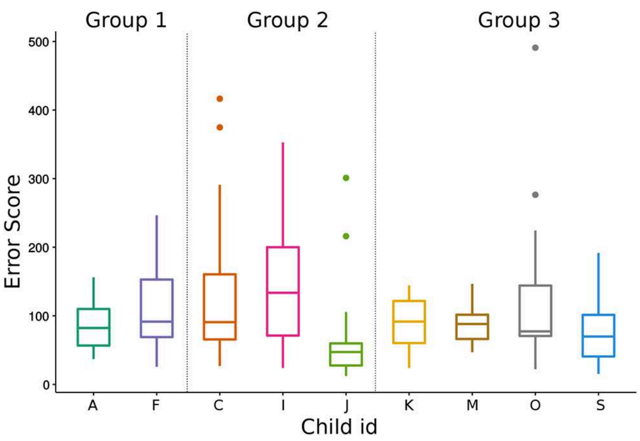


FIGURE 10 | Total pre/post-test error score of each child (Excluding three the children in Group 1 who could not attend pre/post tests due to the time limitation).

In order to identify which pairs of children is different from each other, we did multiple pairwise comparisons between children with a Pairwise Wilcoxon Test and found that error scores of child J is significantly lower than child I ($p < 0.05$), please see **Figure 10** for the comparison.

In order to test if there is a significant difference between children in pre/post-test score difference (improvement in writing), we did a Kruskal-Wallis Test and found no significant difference between the improvements of children ($H = 8.66, df = 8, p = 0.37$).

4.2. Effect of Removing the Grapheme From Guess the Letter Game on Writing Performance

In the initial version of the Guess the Letter game, the writer did not have to reflect on the writing performance as he/she had the grapheme available directly on the activity map. Upon removing the grapheme, the writer was obliged to listen to the feedback given by his/her guesser friends in case they did not understand which letter is drawn due to poor writing. This forces the writer to pay more attention to the discriminative features of letters. To this feedback mechanism, the therapist sometimes contributes additional cues such as, "Write it bigger," "You should make the tail longer," etc. **Figure 11** displays the sample letters written during the Guess the Letter game. Letters indicated as "Trial 1" are the first writing trials of the children which are not understood or not guessed correctly by their peers. Letters indicated as "Trial 2" are the second writing trials of the children just after getting feedback from peers and therapist on the first trials.

4.3. Effect of Removing the Grapheme in Pre/Post-tests on Number of Strokes

Pre/post-test type is changed during Iteration 6 due to the mis-perception of the test with the grapheme. In order to see the effect of this mis-perception on handwriting quality, connectedness of each letter is calculated by counting the number of strokes used to write each letter. We did a Mann-Whitney U-Test to compare the number of strokes to write a letter in the test with grapheme and in the test without grapheme. We found that number of strokes is significantly higher in test with grapheme ($U = 828.5, p < 0.01$), meaning more connected letters were drawn when the grapheme is not provided, see **Figure 12**.

We also looked for the change in number of strokes before and after the writing activity. A Mann-Whitney U Test is used to compare the pre-test and post-test in the number of strokes (using data from both the test with grapheme and the test without grapheme). We found that the change is not affected by the test type (pre-test v.s. post-test), and there is no significant difference in the change of number of strokes ($U = 290.5, p = 0.31$).

4.4. Knowledge Transfer

In Iteration 5, in order to switch between gross and fine motor activities after each letter practice with the robot, the therapists suggested to let the child write the letter in focus with a pencil on a post-it. This also allowed us to monitor the differences and similarities between the writing practice and performance with our robot on an empty map, and with a pencil on a post-it. In this comparison, we also included the writing on the tablet screen with its pen, used in the pre/post-tests, in order to compare and contrast our evaluation practice against the actual writing task. Sample letter performances in this comparison can be seen in **Figures 13, 14**.

In all three variations of the given samples, there is general consistency in the grapheme and ductus. Nevertheless, it is clearly visible that there is increased jerkiness of motion in most of the letters written with the tablet pen compared to the

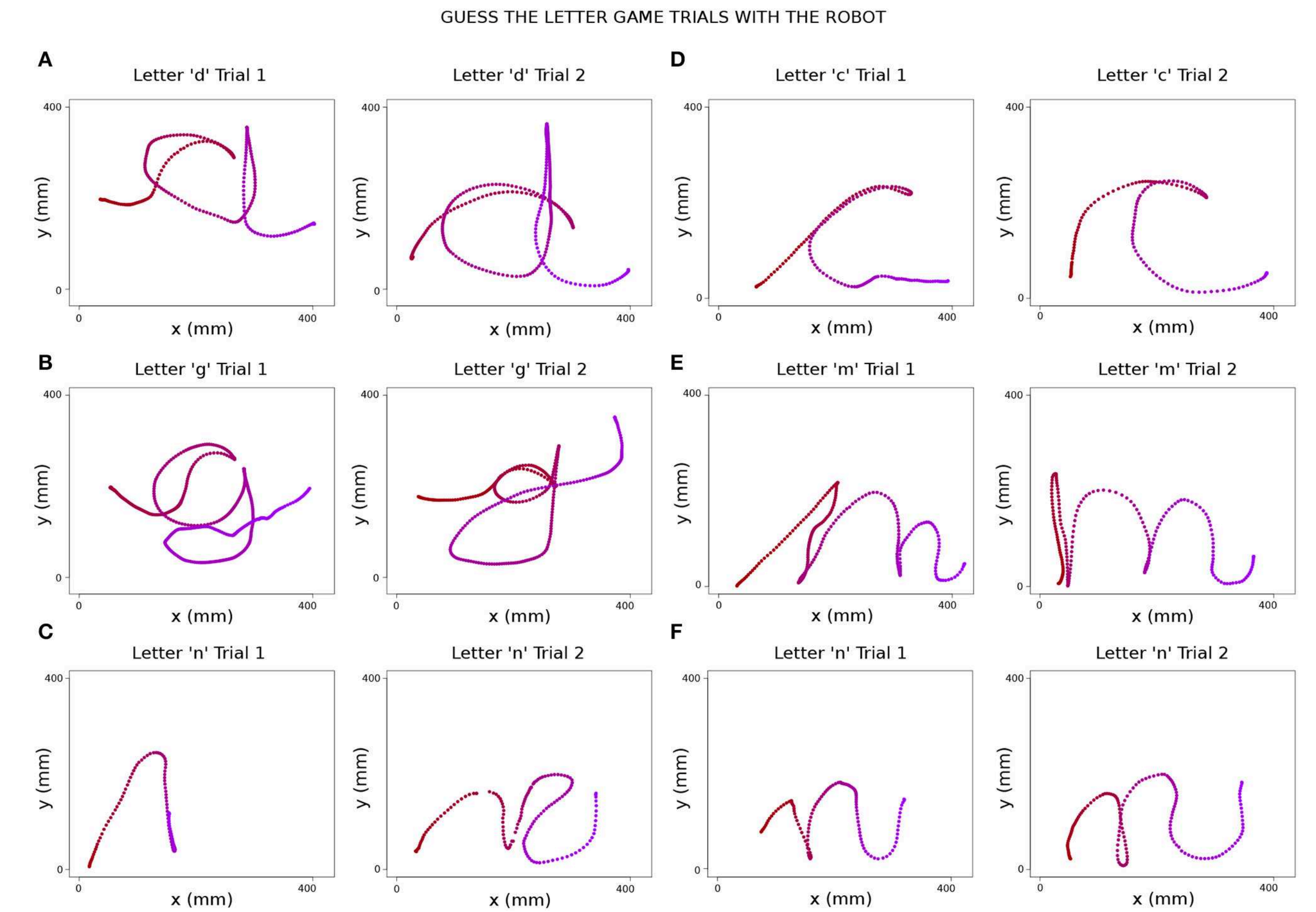


FIGURE 11 | Sample improvements in Guess the Letter game trials. The time dimension in the data is indicated by the color of the stroke going from dark red to purple. **(A)** Trials of writing "d." On the second trial, the child made the upper tail of "d" longer to differentiate it from "a," which was the previous answer from the guessers. **(B)** Trials of writing "g." On the second trial, the child made the curved tail of "g" rounder to be more clear for the other children who were guessing. **(C)** Trials of writing "n." In the second trial, the child wrote a better version of "n" by paying attention to the cursive start. **(D)** Trials of writing "c." On the second trial, the child made the "c" more curvy. **(E)** Trials of writing "m." In the second trial, the child wrote a better version of "m" by paying attention to the proportional size of its different parts. **(F)** Trials of writing "n." The child wrote a better version of "n" in the second trial by paying attention to the direction of the lines and cursiveness.

ones written on paper and written with the robot. There are further slight differences between the methods, such as alignment problems with "a" and "g" in the case of the robot. Finally, the letters written by Child I were observed to be inconsistent in general, which may be due to several reasons including the child's age, her current stage of learning and the nature of the letters. See the corresponding discussion points below on each of these observations.

5. DISCUSSION

5.1. Overall Learning

The presented activity is designed to support children in learning to write cursive letters within occupational therapy sessions. Reported experimental results suggest that children having writing problems are able to improve in letter writing after the use of the system for one session. This was evident by an overall significant decrease in error scores of post-test compared to the error scores of pre-test.

Furthermore, while investigating individual performance differences per child, we found that only the performance of child J and child I were significantly different than each other. As can be seen in **Table 1**, child I was the youngest participant, having her first experience with cursive letters, while child J has the overall best performance and high intelligent assessment.

The score data probing the learning gain differences per child show that even though the levels of the children are different, the learning gains in handwriting are similar, thus suggesting that the activity is inherently adaptive to the learner's abilities and expertise.

5.2. Effect of Removing the Grapheme From Guess the Letter Game on Writing Performance

As results in section 4.2 reflect, providing another version of the game by removing the letter grapheme from the writer side allowed children to learn from their errors when their peers

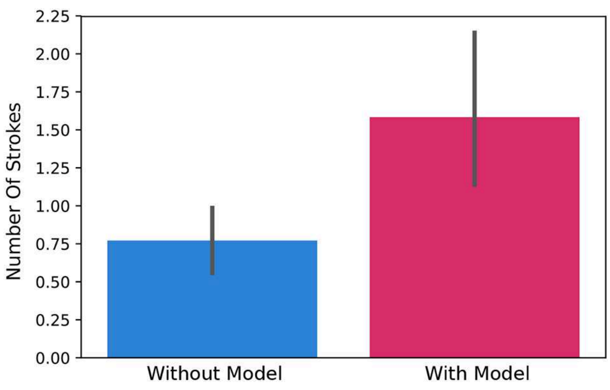


FIGURE 12 | Total number of strokes to write a letter with the test not including a grapheme (on the left) and with the test including the letter grapheme (on the right). We found that the number of strokes is significantly lower in test without the grapheme than in the test with the grapheme ($U = 828.5, p < 0.01$).

could not guess the correct letter. Here, adaptive content of the game allowed us to change the learning goal of the game for the writer, i.e., remembering the ductus in the version with the grapheme v.s. remembering both the grapheme and the ductus in the version without the grapheme. In the version without the grapheme, we observed that the writer child was encouraged to focus on the proportionality of the letter's parts, as well as its discriminative parts from other letters. For instance, in **Figure 11**, the first trial of letter "d" was perceived as an "a" by the guessers and the writer prolonged the upper tail of the letter "d" to make it distinguishable from an "a." The sequence of "n" letters written in the second trials shows the importance of paying attention to the starting gesture and direction of the strokes belonging to the letter.

Even though the learning objective for the writer is changed for the occupational therapy, if desired, within the session, the version with the grapheme can be rapidly switched to, in the case where the learning objective is the ductus only, e.g., in case of a very preliminary learning stage.

The new Guess the Letter game version can also improve children's understanding by their peers in successive trials. Peer collaborative interactions are crucial for a child's learning: Vygotsky (1980) stated that learning awakens in children a variety of internal developmental processes that can operate only when they interact with more competent people in their environment and in cooperation with their peers. The effect of removing the grapheme placed onus on both participants in the Guess the Letter game, brought cooperation to the forefront and was supported by the therapist cues - all highly benefiting the writer in enhancing their learning.

5.3. Added Value of Adaptive Content

The behavioral observations and feedback of therapists through the iterations emphasized the importance of using adaptive interfaces. The unique localization mechanism of the Cellulo platform allowed us to switch from with grapheme to without grapheme versions of the sub-activities, easily adapting to different learning objectives for different letter representations.

The ability to change the number of letters to be learned during sessions and between sessions enabled adapting the activity flow and the total time of the therapy session. We were able to thus tune the duration of Guess the Letter and the total number of times Watch, Feel and Drive the Robot sub-activities by taking the motivation level of the children into consideration. This temporal adaptivity is learned to be crucial in a therapy setting: We experienced a number of failures due to the previously fixed natures of the activity in various iterations, which had been successfully applied previously in a public school environment with typically developing children. In this case, the limited attention and concentration spans of some children in need of such therapy in handwriting absolutely requires this kind of adaptivity; otherwise the activity risks failing at some point. From another perspective, different therapy centers vary in their availability in time and this availability is typically very limited. These two facts further emphasize the importance of temporal adaptability in enabling applicability in a large number of therapy centers, as opposed to being targeted to the scheduling and practices of a single collaborating center.

Adding traditional paper-based activities between each robot-assisted letter activity allowed us to encourage mapping between large and small letters, and between the writing tool used i.e., robot and pen, while allowing us to switch between training gross motor skills and fine motor skills. This change also allowed us to compare and contrast the performances while using different writing media. This is another form of adaptivity of our system design providing another kind of added value, namely adaptability and especially flexibility for integration with traditional practices, which we have previously shown to be potentially useful in improving the gain from the robot-assisted activity (Asselborn, T.* et al., 2018).

5.4. Added Value of Robotic Platform Capabilities

The Cellulo platform allows us to implement parallel robot behaviors where the tasks for different children can be orchestrated simultaneously within the same activity. By scaling up/down the number of robots, the activity can serve to groups of different number of children where one robot can be assigned to one child and be programmed to do the same task as all other robots do. This feature allowed us to design three such sub-activities: Watch, Feel and Drive the Robot.

Tests conducted in therapy environments highlighted an advantage of another feature where a robot becomes active when the corresponding child puts his/her robot on his/her paper map, since the localization is dependent on the paper. This allows parallel activity flow that is compliant to each child's attention and intention to start.

The Cellulo platform also provides synchronized robot behaviors where the behaviors of each robot can depend on each of the other robots, i.e., provides swarm behaviors. Using this attribute, we designed the Guess the Letter game where the writer robot guides the rest of the robots. In different therapy

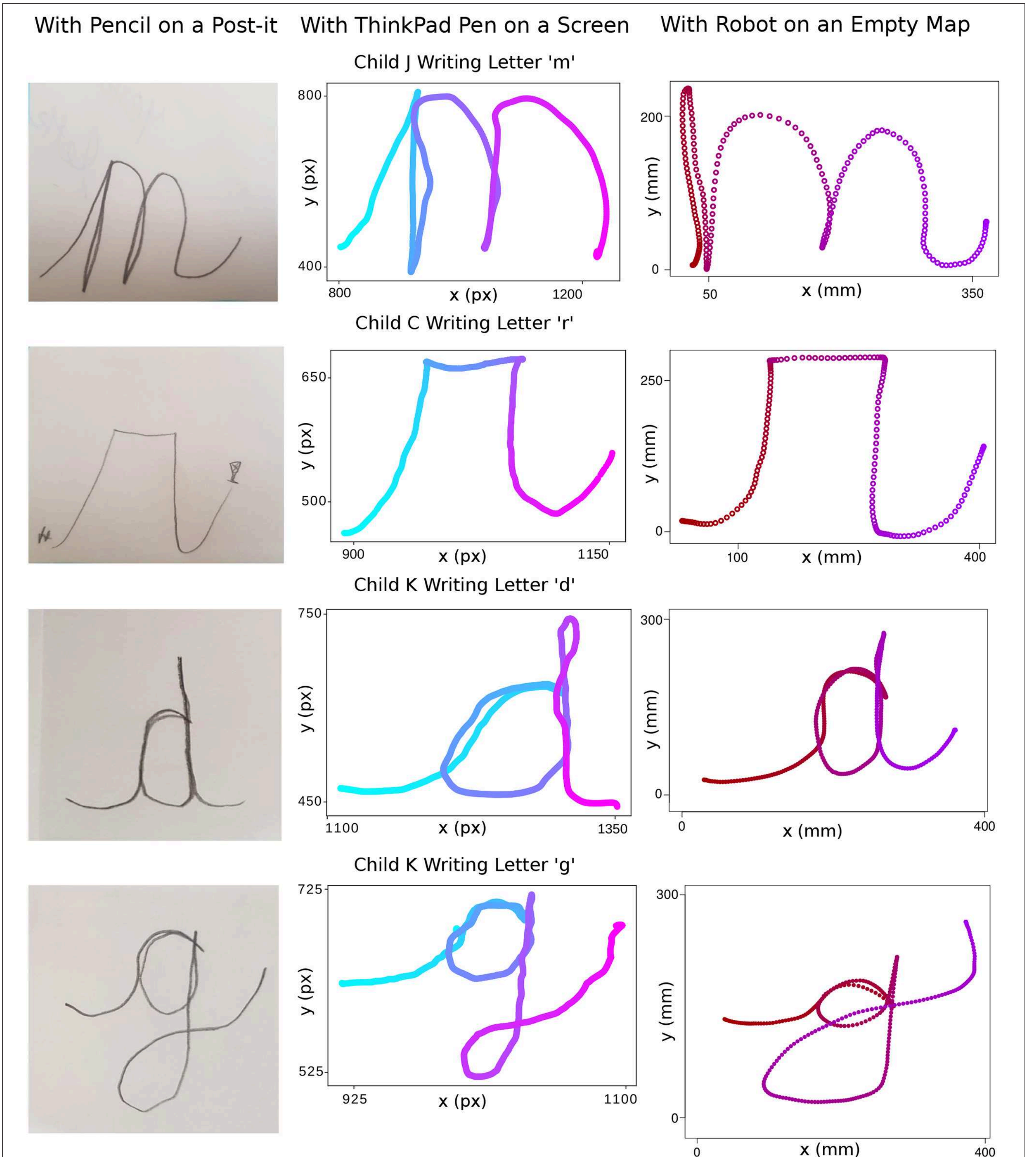


FIGURE 13 | Samples of letters "m," "r," "d," and "g" written with a pencil on a post-it, with ThinkPad pen on a screen and with the Cellulo robot on an empty paper map. The time dimension in the data coming from ThinkPad pen writing is indicated by the color of the stroke going from light blue to light pink. The time dimension in the data coming from the robot writing is indicated by the color of the stroke going from dark red to purple.

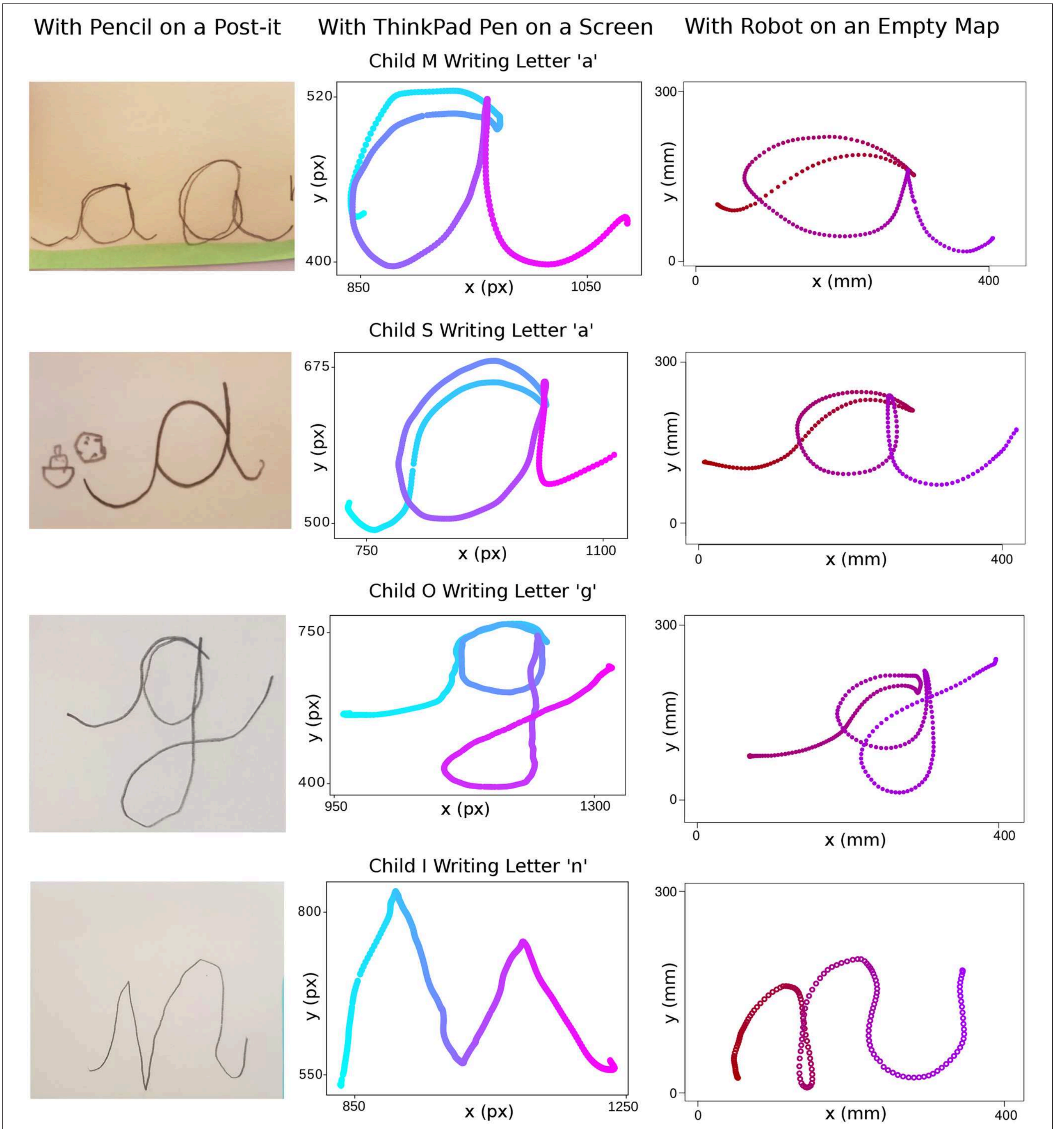


FIGURE 14 | Samples of letters "a," "g," and "n" written with a pencil on a post-it, with ThinkPad pen on a screen and with the Cellulo robot on an empty paper map. The time dimension in the data coming from ThinkPad pen writing is indicated by the color of the stroke going from light blue to light pink. The time dimension in the data coming from the robot writing is indicated by the color of the stroke going from dark red to purple.

centers and sessions, we had varying room settings according to the availability of dissimilar rooms and tables with varying number of children attending the session. Synchronous and parallel capabilities of the robots enable parallel and/or shareable activity workspaces where we can group or separate children, distributing them to different tables with different workspaces as needed. Therefore, the system can be adapted to: (1) The unique room settings of different therapy centers involving the size of the tables, the number of tables and the type of the divider preventing the guessers from seeing the writer during the Guess the Letter game; (2) Number of children attending the session.

5.5. Effect of Pre/Post-test

After using a pre/post-test targeting the measuring of three different learning objectives in Iteration 2, we found that integrating all of these objectives into the therapy environment may not be feasible due to time limitations. Therefore, one test among them that most strongly emphasizes the added value of our activity was selected for use in therapy sessions. However, during the sessions of Iteration 6, the test was mis-perceived as a line following or coloring activity by some children.

This is evidence that perception of such activities by children with attention or visuomotor coordination problems might differ from typically developing children. Even within each group, there may be differences on perception and mapping capabilities. As results in section 4.3 indicate, the device or medium used for pre/post-testing may affect the resulting performance, simply because of this mis-conception of the provided test design. Therefore, for each special group, the system should be able to provide alternative pre/post-test design choices and the designers should question whether they can use the pre/post-tests which are typically designed for regular schools in a special education setting.

5.6. Effect of Writing Tool and Knowledge Transfer

Accommodation of the hand and the grasping and moving styles were different in each medium. Typically for the screen, there were unintended touch events caused by resting one's palm or grazing the fingers over the surface. Presumably because of this phenomenon, some children were observed to adopt an uncomfortable writing position to avoid the unintentional touch event. Another reason for this observation might be the dissimilarity between the friction provided by the tablet surface and its pen, and the typical friction provided by pen and paper surface. A previous study (Annett, 2014) reported that many participants felt that "there was not enough friction between the pen and screen to feel natural" and their hand jerked across the screen as they moved it. This mismatch was also reflected in the number of participants who floated their palms above the surface of the screen which might be due to the different feeling of new pen and screen friction type different than the friction between pencil and paper.

Even though there is jerkiness of motion with the digital pen and screen, the letter shapes in our case are observed to be similar to the ones written on paper, even when we take into account that children are used to writing with pen on paper as typical handwriting practice, and that a digital pen is a new medium for them. Furthermore, the data and the information that a digital platform provides is very valuable from the perspective of teachers and therapists: For instance, this data can be easily made to reflect if the child knows the grapheme and ductus by providing direction information with color coding. For this reason, it must be considered by the designers of handwriting learning activities whether this jerkiness of motion is an important factor or not, and whether it disallows the use of tablets, depending on the needs of the specific application.

In the robot medium, children are using the whole hand to grasp the robot which makes the practice more comparable to gross motor action supported by arm motions, where it is typically easier to control the writing action. This may be a strong reason why we do not observe jerky motion in robot writing. From another perspective, teachers indicated that it is very promising how children can reflect the knowledge of writing onto a robot, which is to a certain extent different than other typical school activities including writing with finger, with pen or with pencil: The robot appears to support the skill transfer from pen and pencil, where all the letters look like written letters when viewed in the same size. This indicates the potential of the robots as an interesting alternative approach providing more feedback than a traditional sandbox or home remedies.

Even though there is a lack of visual feedback of the written letter with the robot (the robot does not leave any "ink" on the paper), the letters written with the robot were observed to be of similar quality to the ones written with the pen on paper. However, the alignment of the strokes which pass over or under previously drawn strokes were more difficult to adjust on an empty map since the previously drawn part of the letter cannot be seen visually. This results in typically more disproportionate parts in letters involving such strokes, such as "g": This is exemplified in **Figures 13**, **14**, which also show a similar problem in the letter "a" whose initial connecting stroke is typically more disproportionately positioned compared to pen and paper where the strokes can be made to pass exactly on top of each other more easily.

Comparing the writing of the letter with a pencil on a post-it, then with the robot on an empty map, and finally with a tablet pen on a screen, gives a useful picture to the child's strengths and points where he/she is having difficulties: Fine motor skills, gross motor skills, child's preference of the tools etc. By learning more about where an exact difficulty or strength may be for given child, a therapist or teacher can add more tools and options to support him/her with overcoming his/her writing difficulties. Furthermore, it gives a variety and interest to the practice that pencil and paper alone cannot provide. It provides an opportunity for the child to work with their favorite writing tool and transfer the grapheme and ductus skills to a less favored tool.

From an individual child level, only the performance of the 5.5-year-old Child I was observed to be inconsistent across the writing tools. This may be due to the lack of orthographic coding of the letter "n" in this child which facilitates forgetting the grapheme of the letter, which could have occurred at different points in time within the activity. It could also be letter dependent since the cursive letter "n" has repetitive bumps making it harder to consistently reproduce. It may be found by future studies that the expectation of mastery in the cursive grapheme and ductus at this age is not be feasible at all.

5.7. Activity Presentation and the Overall Theme

A child who does not enjoy writing was totally engaged for 40 min of the session that was presented within a writing theme (Child F), whereas another child totally lost attention after 30 min of the session that was presented in a more physically active theme (Child J). Here, we observed that the general theme of the therapy session may drastically affect the perception of the proposed activity. For instance, if the robotic writing activity is part of a general writing session, it may boost the child's motivation and engagement. However, presenting it among other activities involving games with more physical activity where children can run, jump, climp etc. can make it more difficult and less motivating for the child to sit down and concentrate for 40 min. Therefore, the general theme within which the writing activity will be proposed should be considered carefully when designing the activity flow, and its duration and composition should be adapted accordingly.

5.8. Novelty Effect

Even though the engagement was observed to be very high for our studies which took 2–3 days, it is not realistic to expect efficiency and engagement in the long term because of the well-known novelty effect typically associated with technologies such as ours. For overcoming this challenge, we hypothesize that the activity could be extended with new drawing concepts and free-drawing sessions. These sessions may include drawing any geometrical shapes, numbers, animals or objects with their model visible in the Feel and Drive the Robot sub-activities or the Guess the Letter game. Guess the Letter can also be modified by various themes, such as:

- Writing the first letter of a friend's/object's/animal's name and guessing who/what it is.
- Free-drawing where the writer child can draw anything they imagine without necessarily using a model: A toy, a house, an umbrella etc.
- Writing the initial cue of the letter (such as wave) while the guessers guess the group of possible letters (such as "a," "c," "d," and "g") as a more advanced sub-activity.

5.9. Limitations and Future Work

Even though writing on an empty map pushes the child to remember and practice what they learned before, the robotic platform lacks visual feedback since it cannot provide the visual output of what is previously drawn by the child. The only feedback is the peers' perception of the writing and the therapist's cues such as: "Your friends didn't understand what you drew, you should write it bigger" or "Please write it as cursive as we learned today."

Another practical limitation of the system is the need to secure paper sheets to the tables, typically done with non-permanent adhesive such as masking tape. For a group of children having attention problems, this alone may create a need for a second therapist since they lose attention quickly while waiting for a preparation process even though it lasts only a few minutes. An alternative is to involve children themselves in this process and having them aid in the preparation and application of the tape, which may also be argued to promote fine motor activity.

The results show the overall effect of the system on progress to handwriting quality of 9 children with visuomotor coordination and attention problems (excluding 3 children whose test data were incomplete). Since the main purpose of the study was to adapt the system to the environment rather than adapting the therapy to the proposed robotic activity, we had a heterogeneous group of children which was a natural aspect of a group therapy session in an occupational therapy center. In order to have more generalizable outcomes, the activity should be further tested in different institutions with more children ranging in age and in difficulties they have. The overall effect of the refined robotic activity (through iterative design process) should be compared with a traditional training process in therapy centers within a study similar to the one conducted in the preschool study. Further research is also needed to investigate the long term improvement and retention in ductus and grapheme learning in such children.

Comparing the results of the children in the therapy center with those of the children in preschool would give us precious insights on the value of the activity with the robot. However, the designs of the activity carried out within school and within the iterations at the therapy centers differ on a number of crucial variables including total duration of the writing activity, mean age of children and device that is used for pre-post test, which reduces the validity of a comparison between the data already collected. Therefore this comparison should be considered for a more controlled follow-up study specifically targeting this question.

The results comparing different writing media and investigating knowledge transfer are limited to observational inputs. In order to explore the knowledge transfer more in depth, more experiments should be conducted where the focus is on transfer learning with quantitative methods.

6. CONCLUSION

Robot-assisted activity was integrated to different occupational therapy sessions and was shown to improve the letter writing performances of children with visuomotor integration and attention problems. This emphasizes the Cellulo interface as a potential tool to conduct handwriting training to teach ductus and grapheme of the letters in multi-child special education environments.

The effective integration of the robot-assisted system into the occupational therapy environment demanded variants of different content throughout the iterations. These modifications included the adaptation of the duration of sub-activities, adaptation of the number of letters and repetitions of each letter in these sub-activities, adaptation of the map graphemes and themes, adaptation of map content (with/without grapheme) within the game, adaptation of game duration and adaptation of activity flow, selectively integrating with traditional practices. These adjustments assisted the consistency and overlap of the learning goals determined by the therapists and the learning goals of the activity. They also allowed adequate engagement of the different groups of children while fitting into the typical timeframe of an occupational therapy session.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Human Research Ethics Committee (HREC) at EPFL. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual(s), and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

AG, TA, and EY: conception and design of the activity, data collection in school, and literature search. TA and AG: design and implementation of the evaluation tests. AG: iterative design and data collection in therapy centers and system design and implementation. AG, WJ, and TA: statistical data analysis. AG, AO, WJ, BB, and MS: interpretation of the results. AG, AO, and BB: drafting of the manuscript. AG, AO, PD, and BB: critical revision of the manuscript for important intellectual content.

FUNDING

This work has been partially supported by the Swiss National Science Foundation and the National Centre of Competence in Research Robotics (NCCR Robotics).

ACKNOWLEDGMENTS

We warmly thank all the children who made this paper possible with their participation, feedback, curiosity, and energy. We also kindly thank their parents for their support for the study and trust in our system. We would like to share our gratitude to Dr. Kristine Jaccoud and her therapist team: Therapist Hélène Arlt and Therapist Pierre Guyot for their valuable constant support, feedback and energy throughout the studies in the occupational therapy center Ergo Leman Sarl. We thank to Therapist Caroline André Carrasco, Therapist Celine Furrer, and Therapist Perrelet Valentine for their enthusiasm, feedback and assistance that facilitated our system to be part of the writing summer school in Ergotherapy Service of CHUV Lausanne. We also would like to thank Therapist Natasha Rouleau for giving authorization for us to integrate ABC Boum cues to our map design to be in line with the Ergo Leman Sarl's teaching methodology.

REFERENCES

Amundson, S. J., and Weil, M. (1996). Prewriting and handwriting skills. *Occupat. Ther. Child.* 3, 524–541.

Annett, M. K. (2014). *The fundamental issues of pen-based interaction with tablet devices* (Ph.D. thesis). University of Alberta, Edmonton, AB, Canada.

Arslan, D. (2012). Examining first grade teachers' handwriting instruction. *Educ. Sci. Theory Pract.* 12, 2839–2846. Available online at: https://eric.ed.gov/?id=EJ1002878

Asselborn, T.*, Guneysu, A.*, Mrini, K., Yadollahi, E., Özgür, A., Johal, W., et al. (2018). "Bringing letters to life: handwriting with haptic-enabled tangible robots," in *Proceedings of the 17th ACM Conference on Interaction Design and Children*, Vol. 12, 219–230. doi: 10.1145/3202185.3202747

Asselborn, T., Gargot, T., Kidziński, Ł., Johal, W., Cohen, D., Jolly, C., et al. (2018). Automated human-level diagnosis of dysgraphia using a consumer tablet. *npj Digit. Med.* 1:42. doi: 10.1038/s41746-018-0049-x

Bara, F., and Gentaz, E. (2011). Haptics in teaching handwriting: the role of perceptual and visuo-motor skills. *Hum. Movem. Sci.* 30, 745–759. doi: 10.1016/j.humov.2010.05.015

Berninger, V., Vaughan, K., Abbott, R., Abbott, S., Woodruff Rogan, L., Brooks, A., et al. (1997). Treatment of handwriting problems in beginning writers: transfer from handwriting to composition. *J. Educ. Psychol.* 89, 652–666. doi: 10.1037/0022-0663.89.4.652

Bluteau, J., Coquillart, S., Payan, Y., and Gentaz, E. (2008). Haptic guidance improves the visuo-manual tracking of trajectories. *PLoS ONE* 3:e1775. doi: 10.1371/journal.pone.0001775

Chandra, S., Dillenbourg, P., and Paiva, A. (2019). Children teach handwriting to a social robot with different learning competencies. *Int. J. Soc. Robot.* 1–28. doi: 10.1007/s12369-019-00589-w

Charles, M., Soppelsa, R., and Albaret, J.-M. (2003). *BHK: Echelle d'èvaluation Rapide de l'ècriture Chez l'Enfant*. Paris: ECPA, Les éditions du Centre de Psychologie Appliquée.

Christensen, C. A. (2009). "The critical role handwriting plays in the ability to produce high-quality written text," in *The SAGE Handbook of Writing Development* (Los Angeles, CA), 284–299. doi: 10.4135/9780857021069.n20

Danna, J., and Velay, J.-L. (2015). Basic and supplementary sensory feedback in handwriting. *Front. Psychol.* 6:169. doi: 10.3389/fpsyg.2015.00169

Dillenbourg, P. (ed.). (1999). "What do you mean by collaborative learning?," in *Collaborative-Learning: Cognitive and Computational Approaches* (New York, NY: Pergamon), 1–19.

Feder, K. P., and Majnemer, A. (2007). Handwriting development, competency, and intervention. *Dev. Med. Child Neurol.* 49, 312–317. doi: 10.1111/j.1469-8749.2007.00312.x

Garcia-Hernandez, N., and Parra-Vega, V. (2009). "Active and efficient motor skill learning method used in a haptic teleoperated system," in *The 18th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN 2009* (Toyama), 915–920. doi: 10.1109/ROMAN.2009.5326132

Hayes, D. (1982). Handwriting practice: the effects of perceptual prompts. *J. Educ. Res.* 75, 169–172. doi: 10.1080/00220671.1982.10885375

Hood, D., Lemaignan, S., and Dillenbourg, P. (2015). "When children teach a robot to write: an autonomous teachable humanoid which uses simulated handwriting," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (Portland, OR), 83–90. doi: 10.1145/2696454.2696479

Hostettler, L., Özgür, A., Lemaignan, S., Dillenbourg, P., and Mondada, F. (2016). "Real-time high-accuracy 2D localization with structured patterns," in *2016 IEEE International Conference on Robotics and Automation (ICRA)* (Stockholm), 4536–4543. doi: 10.1109/ICRA.2016.7487653

Kim, M.-G., Park, M., Choi, Y. S., Kim, K.-H., Lee, N., and Sohn, D.-S. (2019). "Understanding work environment of therapists to incorporate robots into education for children with developmental disorders," in *2019 16th International Conference on Ubiquitous Robots (UR)* (Jeju: IEEE), 603–608.

Kreijns, K., Kirschner, P. A., and Jochems, W. (2003). Identifying the pitfalls for social interaction in computer-supported collaborative learning environments: a review of the research. *Comput. Hum. Behav.* 19, 335–353. doi: 10.1016/S0747-5632(02)00057-2

Laszlo, J. I., and Bairstow, P. J. (1984). Handwriting: difficulties and possible solutions. *School Psychol. Int.* 5, 207–213. doi: 10.1177/0143034384054004

Laszlo, J. I., and Broderick, P. (1991). "Drawing and handwriting difficulties: reasons for and remediation of dysfunction," in *Development of Graphic Skills: Research, Perspectives and Educational Implications* (London: Academic Press), 259–280.

Lee, H. W., and Lim, K. Y. (2013). Does digital handwriting of instructors using the iPad enhance student learning? *Asia-Pacif. Educ. Res.* 22, 241–245. doi: 10.1007/s40299-012-0016-2

Maeland, A. F. (1992). Handwriting and perceptual-motor skills in clumsy, dysgraphic, and 'normal' children. *Percept. Mot. Skills* 75(3 Suppl.):1207–1217. doi: 10.2466/pms.1992.75.3f.1207

Mekyska, J., Faundez-Zanuy, M., Mzourek, Z., Galaz, Z., Smekal, Z., and Rosenblum, S. (2016). Identification and rating of developmental dysgraphia by handwriting analysis. *IEEE Trans. Hum. Mach. Syst.* 47, 235–248. doi: 10.1109/THMS.2016.2586605

Özgür, A., Johal, W., and Dillenbourg, P. (2016). "Permanent magnet-assisted omnidirectional ball drive," in *Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Daejeon). doi: 10.1109/IROS.2016.7759180

Özgür, A., Lemaignan, S., Johal, W., Beltran, M., Briod, M., Pereyre, L., et al. (2017). "Cellulo: versatile handheld robots for education," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna), 119–127. doi: 10.1145/2909824.3020247

Pagliarini, E., Scocchia, L., Vernice, M., Zoppello, M., Balottin, U., Bouamama, S., et al. (2017). Children's first handwriting productions show a rhythmic structure. *Sci. Rep.* 7, 1–10. doi: 10.1038/s41598-017-05105-6

Palluel-Germain, R., Bara, F., De Boisferon, A. H., Hennion, B., Gouagout, P., and Gentaz, E. (2007). "A visuo-haptic device-telemaque-increases kindergarten children's handwriting acquisition," in *Second Joint EuroHaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems (WHC)* (Tsukaba: IEEE), 72–77. doi: 10.1109/WHC.2007.13

Rosenblum, S. (2008). Development, reliability, and validity of the handwriting proficiency screening questionnaire (HPSQ). *Am. J. Occupat. Ther.* 62, 298–307. doi: 10.5014/ajot.62.3.298

Rosenblum, S., Ben-Simhon, H. A., Meyer, S., and Gal, E. (2019). Predictors of handwriting performance among children with autism spectrum disorder. *Res. Autism Spectr. Disord.* 60, 16–24. doi: 10.1016/j.rasd.2019.01.002

Salvador, S., and Chan, P. (2004). "FastDTW: toward accurate dynamic time warping in linear time and space," in *3rd Workshop on Mining Temporal and Sequential Data, ACM KDD'04* (Seattle, WA).

Smits-Engelsman, B., Niemeijer, A., and van Galen, G. (2001). Fine motor deficiencies in children diagnosed as DCD based on poor grapho-motor ability. *Hum. Movem. Sci.* 20, 161–182. doi: 10.1016/S0167-9457(01)00033-1

Vygotsky, L. S. (1980). *Mind in Society: The Development of Higher Psychological Processes*. London: Harvard University Press.

Yamasaki, T., Hattori, T., and Inokuchi, S. (1990). "An adaptive real time training system for beautiful handwriting of japanese letters," in *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* (Los Angeles, CA), 88–93. doi: 10.1109/ICSMC.1990.142066

Zolna, K., Asselborn, T., Jolly, C., Casteran, L., Johal, W., Dillenbourg, P., et al. (2019). The dynamics of handwriting improves the automated diagnosis of dysgraphia. *arXiv: 1906.07576*.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 12 March 2020
doi: 10.3389/frobt.2020.00033

A Playful Experiential Learning System With Educational Robotics

Antonella D'Amico[1,2,3], Domenico Guastella[2] and Antonio Chella[3,4,5]*

[1] Dipartimento di Scienze Psicologiche, Pedagogiche, dell'Esercizio Fisico e della Formazione, Università degli Studi di Palermo, Palermo, Italy, [2] MetaIntelligenze Onlus, Palermo, Italy, [3] Centro Interdipartimentale di Tecnologie della Conoscenza, Università degli Studi di Palermo, Palermo, Italy, [4] RoboticsLab, Dipartimento di Ingegneria, Università degli Studi di Palermo, Palermo, Italy, [5] Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR), Consiglio Nazionale delle Ricerche, Palermo, Italy

This article reports on two studies that aimed to evaluate the effective impact of educational robotics in learning concepts related to Physics and Geography. The reported studies involved two courses from an upper secondary school and two courses from a lower secondary school. Upper secondary school classes studied topics of motion physics, and lower secondary school classes explored issues related to geography. In each grade, there was an "experimental group" that carried out their study using robotics and cooperative learning and a "control group" that studied the same concepts without robots. Students in both classes were subjected to tests before and after the robotics laboratory, to check their knowledge in the topics covered. Our initial hypothesis was that classes involving educational robotics and cooperative learning are more effective in improving learning and stimulating the interest and motivation of students. As expected, the results showed that students in the experimental groups had a far better understanding of concepts and higher participation to the activities than students in the control groups.

Keywords: educational robotics, metacognition, physics, geography, playful-based learning

OPEN ACCESS

Edited by:
Dirk Ifenthaler,
Universität Mannheim, Germany

Reviewed by:
Alessandra Sciutti,
Italian Institute of Technology, Italy
Giulia Perugia,
Uppsala University, Sweden

****Correspondence:***
Antonella D'Amico
antonella.damico@unipa.it

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *30 September 2019*
Accepted: *25 February 2020*
Published: *12 March 2020*

Citation:
D'Amico A, Guastella D and Chella A (2020) A Playful Experiential Learning System With Educational Robotics. Front. Robot. AI 7:33. doi: 10.3389/frobt.2020.00033

INTRODUCTION

Technological development in the twenty first century has led to the introduction of new types of technologies in the educational field. One exciting technological innovation is educational robotics (ER), i.e., the application of robotics in an educational context. In this approach, students acquire specific skills (e.g., knowledge of electricity, electronics, robotics) and develop strategic and dynamic capabilities in a playful context that is supposed to increase a learner's motivation and engagement, and facilitate learning. Robotics, indeed, allows the application of the principles of constructivism (Piaget and Inhelder, 1966), constructionism (Papert, 1980, 1993), and embodied cognition (Shapiro, 2010) to learning.

The Theoretical Background

According to Piaget's principles, cognition develops as an active process in the mental construction of knowledge related to concrete objects in the environment (Piaget and Inhelder, 1966). Papert (1980), starting from Piaget's principles, claimed that learning should not simply be considered the acquisition of behaviors or skills. Instead, it is a subjective process of structuring knowledge, facilitated and enriched by the environmental production of concrete objects defined by the author as "objects to think with" (Papert and Harel, 1991). Papert assigned a high value to concrete thinking, the physical dimension and tangible products of human intelligence (Turkle and Papert, 1992). From his "learning by doing" perspective, ideas are formed and transformed when they are expressed through different tools, when they are used in particular contexts and when individual minds elaborate them. The author deals with the concept of "construction sets": every mental construction (or robot components, in the case of educational robotics) can be metaphorically associated with parts assembled and built together (Papert, 1980), allowing people "to think with" technological artifacts.

Recently, Raskin (2002) claimed that knowledge is not built through ontogenetically programmed stages of learning, but that it is developed through continuous actions and doing, and the adaptation of the child in the environment with which they interact. In this framework, learning is stimulated by approaches focused on "doing" and on the production of tools that encourage the learner to activate this "constructive" way of learning. Suitable learning contexts should be able to promote: (a) the use of functional strategies to achieve pre-established goals; (b) exposure to different points of view; (c) the involvement of students as an active part of the educational activity; (d) the role of the teacher as a facilitator of the "source of knowledge"; (e) cooperation through the social negotiation of meanings (Vygotsky, 1978); and (f) the use of investigation methodologies as proceeding by trial and error and the activation of problem-solving. Herrington and Kervin (2007) suggest an extension to constructionism by inserting the construct of "authentic activity," or "poorly defined" activities, requiring students to define the necessary tasks and sub-tasks to complete them. The authors also specify that some kind of "poorly defined" activities may include complex tasks to be investigated over an extended period, which offer students the opportunity to examine the task from different perspectives by using a variety of resources and by providing an opportunity to collaborate and to reflect. We claim that ER satisfies all these requirements for such useful learning contexts.

The importance of educational robotics is also supported by the embodied cognition theoretical approach (Shapiro, 2010), which emphasizes the value of experiential activities in teaching. Embodied cognition is a multidimensional and interdisciplinary construct developed through the contribution of scientific disciplines such as neuroscience, psychology, philosophy, and cognitive sciences. The embodied cognition approach overcomes the debate about the role of the brain/body or the environment in the development of the human mind (Gibbs, 2005). Simply, it considers the body and the environment as an "extension" of the mind. From the perspective of embodied cognition, any kind of human cognition is embodied. In the study of the mind, the role of the body and its interaction with the environment is thus essential. The body ensures coordination between cognition and action, and it facilitates or hinders cognition. Similarly, in ER, learners experience a direct connection between body and mind, and the way the characteristics of the body affect and are affected by the functioning of the mind.

Two main typologies of robots are traditionally employed in the ER field. The first type is classified as a humanoid or zoomorphic robot, as NAO (Shamsuddin et al., 2012) or Pleo (Kim et al., 2013). Robotics construction kits such as LEGO® kits, are a different kind of tool. Humanoid or zoomorphic robots are employed to study human-robot interaction and to improve social skills in children, since they reproduce human or animal-like interactive behavior. Social robots have the advantage of being able to show "human social" characteristics, such as emotions and autonomous language simulation; they can establish/maintain social relationships; they employ natural cues (gaze, gestures, etc.), and learn/develop social competencies (Fong et al., 2003). Social robots are employed in the educational field, for instance, as tutors/teachers or as peers in an educational context (see Belpaeme et al., 2018, for a review). The user does not have the ability of modifying the robot bodies, however, and they are generally expensive, and require advanced expertise in computer programming.

Robotics construction kits, on the contrary, allow the user to build small mobile robots using bricks, sensors, and motors. A simple visual blocks programming language programs the behavior of the robot. These tools stimulate creativity and manipulation. They may be used to perform various activities and to teach various skills in children and adolescents. From our perspective, the use of robotics construction kits is one of the best ways to allow children to work with artificially extended minds.

The Empirical Evidences

A number of authors have employed robotics construction kits over the past 15 years, for improving cognitive abilities in children with special needs, and for enhancing social and cooperative dimensions and learning in a school context.

In one of our early studies, we documented significant improvements in the academic performance and metacognitive and motivational processes in a student with intellectual disability (Caci and D'Amico, 2005). Similarly, Fridin and Yaakobi (2011) show that robotics can help to improve memory and attention in children with attentional deficit and hyperactivity disorder. More recently, after a robotics lab, we observed improvements in the short-term memory and visual-spatial abilities in a student with visual-spatial difficulties. A child with an intellectual disability and attentional deficit also showed a reduction in behavioral problems (D'Amico and Guastella, 2019). In our 2013 study (Caci et al., 2013a,b), we demonstrated that robotics labs improve the visual-spatial abilities of groups of children with typical development. We also described (D'Amico and Guastella, 2018) how robot construction kits can be adopted in the field of affective computing to support the social and emotional learning of children with typical development or with special needs. In fact, in a study involving a child with autism spectrum disorder, we observed significant changes in social reciprocity and emotional expression, as well as improvements in fine motor skills, verbal and preverbal skills, and a considerable increase in the child's interest in the activity and in-play skills (Guastella et al., 2020).

Many authors have claimed that educational robotics is positively correlated with collaborative learning behaviors, social skills, and the perception of individual and collective self-efficacy (Denis and Hubert, 2001; Kanda et al., 2012). In a school context, the most frequent application of robot construction kits is as a tool for supporting learning in the STEM disciplines (science, technology, engineering, and mathematics). Williams et al. (2007) showed how cooperative learning activities associated with technological tools facilitate the learning of physics in secondary school students, but not in science. Barak and Zadok (2009) aimed to improve learning concepts in Science, Technology, and problem-solving, and showed that the use of a project-based methodology through the use of robotic kits in high school had benefits for an individual's cognitive flexibility, problem-solving and teamwork. Barker and Ansorge (2007) found that an educational robotics course focused on science teaching improved academic performance in a group of students, aged 9 to 11, compared to a control group. Whittier and Robinson (2007) showed that non-English students made significant gains in their conceptual understanding of science concepts after ER activities. Finally, Wei and Hung (2011) showed that ER provides learners with more opportunities for hands-on exercises in mathematics, deepens their perception of the learning contents and that it improved their motivation in the study.

As Toh et al. (2016) claimed, many studies using ER are based on qualitative results, and only one (Whittier and Robinson, 2007) used a quasi-experimental design and provided quantitative measures. Other studies providing quantitative measures are thus needed to confirm the results about learning in STEM disciplines. At the same time, it is crucial to determine whether educational robotics can support learning other than in the STEM disciplines.

Starting from these considerations, we carried out two studies in schools using a quasi-experimental design and employing quantitative tools to measure the impact of educational robotics for fostering the learning of physics and geography. We will first describe the study that involved students from upper secondary school and then the study that involved lower secondary school students, although both studies were carried out independently and at almost the same time by teachers that were been trained in Educational Robotics.

STUDY 1

The goal of Study 1 was an evaluation of the application of robot construction kits as tools for supporting the learning of STEM disciplines. In particular, we focused on teaching and learning the concept of physics.

Method

A quasi-experimental design was used: two classes of the same level were selected. One was assigned to the experimental condition (ER-Physics Lab, described in section The Empirical Evidences) and the second was assigned to the control condition. The class assigned to the control condition attended only one lesson about robotics (focusing on the use of sensors, actuators and processors), but then they attended traditional theoretical lessons about physics, studying the same concepts than students in experimental condition. Both classes shared the same teacher of physics. To test the efficacy of the experimental condition, before and after the whole set of lessons, students in the experimental and control classes completed two questionnaires about physics concepts. We also collected their school marks for physics.

Participants

A total of 49 students (26 males, 23 females) of about 16 years of age (mean = 16 years, standard deviation = 12 months) participated in the study. They were divided into an experimental and control group. They were from two classes attending the third year of an Italian upper secondary school. One class, including 23 students (11 males, 12 females), was assigned to the experimental condition, and the other class, including 26 students (15 males, 11 females), was assigned to the control condition. There were no students with special educational needs.

Although all the students in experimental condition attended the laboratory, some were excluded from the final analyses since they missed more than two lessons during the laboratory or because they were absent when the pre-test or the post-test was administered (this kind of dropout, unfortunately, is very frequent in studies performed at school). We were thus able to perform the final analyses on a total of 8 students for the experimental group (6 males, 2 females) and 9 students (5 males, 4 females) for the control group.

The ER-Physics Lab

The class in the experimental condition attended the ER-Physics Lab, which included activities concerning the use of ER for teaching and learning concepts of physics, and particularly the concepts of energy and motion. Lessons were conducted by the teacher, in collaboration with the experimenter from our team, in the role of observer.

Five LEGO Mindstorm EV3 robotics construction kits and five PCs were used in the school's computer lab. The students were divided into five groups so that each group had the opportunity to use one robotic construction kit. The ER activities were carried out in 6 weekly lessons lasting 2 h each, as described below.

First Lesson

During the first lesson, the students learned to use the main programming blocks related to the robot motion and functionalities of the sensors. After a brief theoretical introduction, they were free to perform two simple robot motion programs by employing sensors to avoid obstacles, or to stop when a specific condition occurred. At the end of the programming phase, each group showed the program to the other groups, describing the main problems encountered and illustrating the solutions found to solve them.

Second Lesson

The second lesson focused on the use of flow programming blocks (loop, switch and wait) and the data operations (variable definition, mathematical operations, and arrays) necessary to code the formulas about motion. The students attended a tutorial and coded a program using these functions. At the end of the programming phase, they showed their results to their classmates.

Third Lesson

During the third lesson, after studying formulas for motion calculation, and the respective inverse formulas, each group had to solve a problem with physics using robots. The first step was to estimate the robot's speed. This is a function that is not pre-programmed in EV3 construction kits. To achieve the goal, they employed a meter tool to calculate the length of a path and a stopwatch to measure the time needed by the robot to follow it. Once the students had captured data about space and time, they easily derived the robot speed. The information was then employed to predict the robot's travel time along a straight path or to measure the length of different surfaces, by applying the inverse formulas of the uniform rectilinear motion.

Fourth Lesson

At the fourth lesson, the students were asked to implement a program to compute the speed and acceleration of the robot through the radius of wheels and the number of rotations of an engine. The robot was programmed using specific variables (in this case, the time and the number of rotations) and by displaying the results on the computer programming interface or in the small display embodied in the robot. Every time the engine power and the distance traveled were modified, then the program computed and displayed the new speed and acceleration. At the end of the programming phase, each group showed their results to their classmates.

Fifth Lesson

During the fifth lesson, the students were asked to calculate the speed of a rotating rod attached to a robot wheel (whose radius was known). The students were also asked to calculate the acceleration of the rod. This request was made in order to activate the students' critical reasoning capabilities, as they had to realize that the required calculation could not be done using the available data only.

Sixth Lesson

During the final lesson, students created a program of their choice using the formulas or programming blocks employed in the previous lessons, according to the concepts of physics or programming.

They created exciting programs: Group 1 created a robot that computed its length when dragged on a surface; Group 2 created a robot able to follow a path without impacting against walls or obstacles; Group 3 worked on a variant of the robot that calculates its speed but without using the stopwatch and setting a timer, so that the robot would stop after a period; Group 4 proposed a variant of the robot that calculates distance, was able to display the results of the calculations on the embodied display and to keep the previous measurements in its memory; and finally, Group 5 created a robot able to recognize the color of a spherical object and to carry it to a predetermined position.

Measures

Physics School Marks

The average marks gained by each student in physics before and after the study activity were computed by the teacher. They were based on the student's performance in both written assignments and oral questions and, as usual in Italian schools, ranged from 0 to 10.

Questionnaires About Physics and Robotics

In the test and the re-test phase, we employed two parallel questionnaires that covered the topics of physics and robotics and were focused on the periods before and after the lab. Both questionnaires were prepared by the teacher and presented slightly different items, but with the same level of difficulty. They both comprised eight questions: two questions concerned physics problems where the students had to choose the correct formula to solve a problem. The remaining six questions covered topics related to robotics (sensors, actuators, processors, and data collection).

The questionnaires about physics and robotics are to be considered objective measures of learning since they were based on answers given by the students. They were employed to avoid teachers being influenced by the experimental hypotheses when evaluating the students.

Both questionnaires were administered in the classroom, separately for the experimental and control group, in two joint sessions before and after activities.

Results

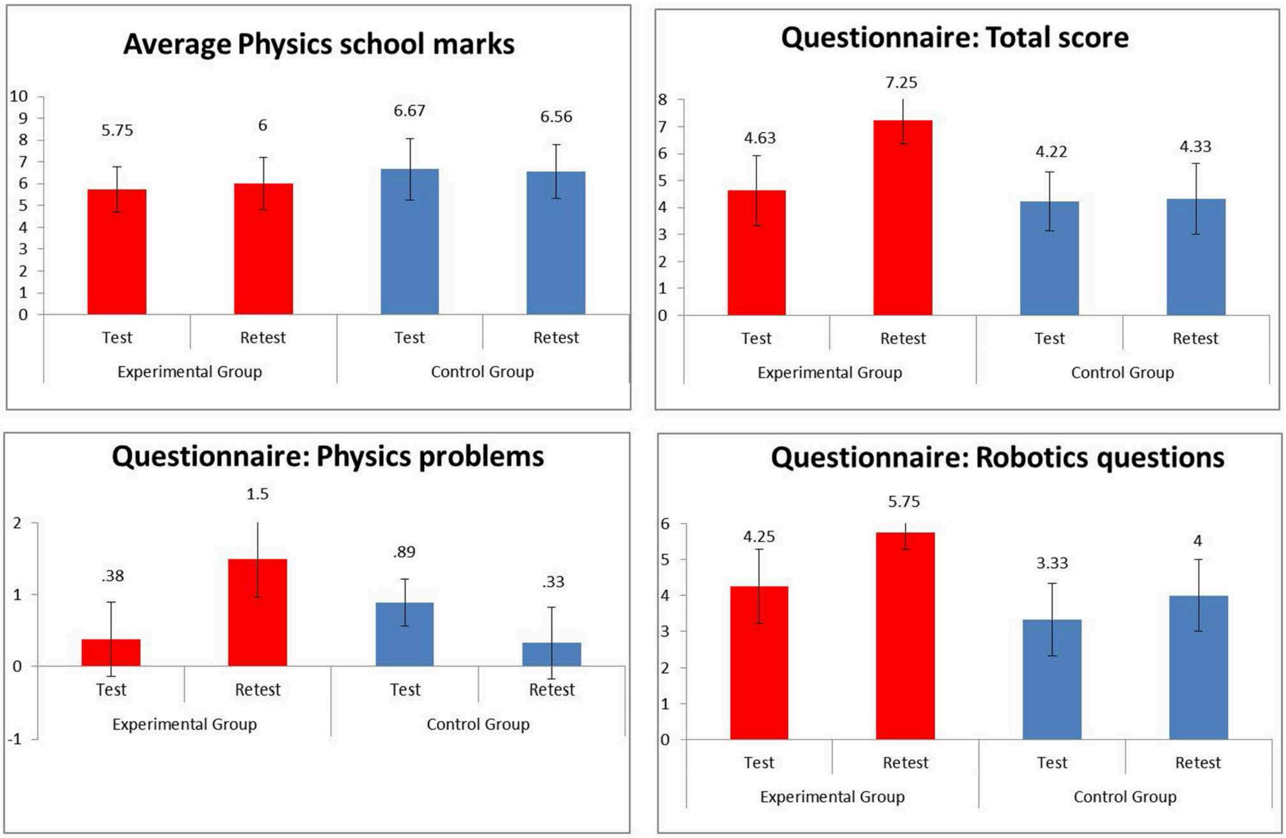
A series of 2 × 2 factorial ANOVA using group (experimental, control) × time (test, retest) were performed on data collected during test and retest phases. The analyses were performed on: (a) the average marks in physics; (b) the total score for the questionnaire (QTot); (c) the score related only to questions about physics (QPhys); (d) the score related only to questions about robotics (QRob).

The means and standard deviations of all measures collected in the experimental and control groups before and after the activities and results of statistical analyses are shown in **Table 1** and **Graph 1**.

The results of the Group x Time factorial ANOVA performed on the average marks for physics collected before and after the activities did not reveal a significant main effect for Group [$F_{(1, 15)} = 1.69$, $p = 0.202$, $\eta^2 = 0.102$], Time [$F_{(1, 15)} = 0.12$, $p = 0.736$, $\eta^2 = 0.008$), nor interaction Group × Time [$F_{(1, 15)} = 0.80$, $p = 0.386$, $\eta^2 = 0.050$]. The average marks of the students in the Experimental and Control groups for Physics were thus similar before the start of the activities, and had not substantially changed from test to retest for either the students in the experimental group, nor for those in control groups.

The ANOVA computed on the total score of the questionnaire (Qtot), conversely, showed the significant main effect of Group [$F_{(1, 15)} = 18.87$, $p = 0.001$, $\eta^2 = 0.557$], indicating that, in general, the experimental group obtained better results than the controls in the questionnaire. The significant main effect of Time [$F_{(1, 15)} = 10.62$, $p = 0.005$, $\eta^2 = 0.415$] indicated that all the participants improved their performance in Qtot from the test to the retest. A significant interaction between Group and Time [$F_{(1, 15)} = 8.97$, $p = 0.009$, $\eta^2 = 0.374$] indicated that students in the experimental group improved significantly more than students in the control group.

Significant results were obtained even when only the test scores relative to physics were taken into account (QPhys). The results of the ANOVA indeed showed a main effect of Group very close to statistical significance [$F_{(1, 15)} = 4.21$, $p = 0.058$,



GRAPH 1 | Study 1: Scores of experimental and control group for all variables measured in test and retest phases.

TABLE 1 | Mean and standard deviation of all measures collected in the experimental and control groups and results of statistical analyses.

	Experimental group				Control group				ANOVAs		
	Before		After		Before		After		Group	Time	Group × Time
	M	SD	M	SD	M	SD	M	SD	$F_{(1, 15)}$	$F_{(1, 15)}$	$F_{(1, 15)}$
Average Physics school marks	5.75	1.03	6.00	1.19	6.67	1.41	6.56	1.23	1.69	0.12	0.80
Questionnaire: Total score (QTot)	4.63	1.302	7.25	0.886	4.22	1.093	4.33	1.323	18.87**	10.62*	8.97*
Questionnaire: Physics problems (QPhys)	0.38	0.518	1.50	0.535	0.89	0.33	0.33	0.500	4.21	2.90	25.28**
Questionnaire: Robotics questions (QRob)	4.25	1.03	5.75	0.46	3.33	1.00	4.00	1.00	20.53**	10.65*	1.58

*$p < 0.05$, **$p < 0.001$.

$\eta^2 = 0.219$], indicating that, in general, the experimental group obtained better results than the controls in QPhys. The non-significant main effect of Time [$F_{(1, 15)} = 2.90$, $p = 0.109$, $\eta^2 = 0.162$] indicated that when the score of all participants was taken together, there was no improvement from the test to the retest session. A significant interaction between Group and Time [$F_{(1, 15)} = 25.28$, $p < 0.001$, $\eta^2 = 0.628$] indicated that students in the experimental group improved significantly more than students in the control group in answering the questions about physics.

The ANOVA performed on the item of the questionnaire related to questions about robotics only (QRob) demonstrated that there was a significant main effect of Group [$F_{(1, 15)} = 20.53$, $p < 0.001$, $\eta^2 = 0.578$], indicating that students in the experimental group answered better than students in the control group. The main effect of time [$F_{(1, 14)} = 10.65$, $p = 0.005$, $\eta^2 = 0.415$] indicated that both the experimental and control group improved their performance in QRob from test to retest, however, and, unexpectedly, that there was no interaction between Group and Time [$F_{(1, 15)} = 1.58$, $p = 0.229$, $\eta^2 = 0.095$], indicating that there was no significant difference in the improvement of students belonging to the experimental group when compared to students in the control group.

Discussion

The results of Study 1 demonstrated that the ER-Physics Lab failed to exert a significant improvement in the general knowledge about physics, since we did not find any difference between the experimental and the control group in their average marks for physics, however, it has to be said that none of the students (in the experimental or the control group) showed an improvement in the average physics mark from test to retest, probably because these marks were based on a range of school activities and they generally remain reasonably stable in a narrow period such as that considered in the study.

Remarkably, thanks to the ER-Physics Lab, students in the experimental group obtained a significant improvement compared to students in the control group when answering the questionnaire about physics and robotics administered before and after the activities. An in-depth analysis revealed that the group differences concerned their abilities when solving the two problems with physics. At the same time, there were no significant effects due to the ER-Physics Lab when answering the questions about robotics. This is because the questions about robotics concerned general knowledge about the use of a robot's components, which students in the control group may have acquired during the frontal lesson about robotics. Conversely, solving a problem of physics requires a more in-depth understanding of the physics concepts, and, in this sense, the results demonstrated that the ER-Physics lab allowed students to acquire significant and conceptual learning about the topics in focus during the lab.

Impressive outcomes were derived from the informal observations performed during the lab by the teacher and by the experimenter. They reported active participation by each student group with an evident stimulation of their ability to work in groups, intense collaboration, and a constant comparison among pairs. In particular, during the final lesson, where students were free to create a program of their choice, both teacher and observer reported the emergence of various skills and abilities: some groups focused on solving problems in physics by the different strategies described in sixth lesson; others worked on the programming of the robot (i.e., to follow a track avoiding obstacles or recognizing colors), and others focused on "hardware aspects," by adapting the robot body according to the task to be performed. Moreover, the teacher perceived an unexpected increase in the performance of the students who were usually less motivated to learn.

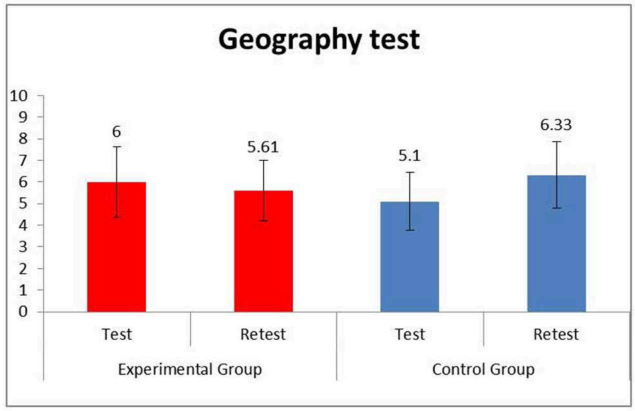
STUDY 2

Study 2 evaluated the efficacy of ER for teaching and learning concepts of geography about the regions of Italy to students attending the second class of an Italian lower secondary school.

The study has two novel aspects compared to the previous one, and also to the literature in the ER field: the first is that it is focused on geography, a subject that, to our knowledge, has never been considered in studies about educational robotics. The second is that it employs BB8, a small robot by Lego, never previously employed in experimental studies.

Method

Similarly to study 1, also in study 2 we used a quasi-experimental design: two classes of the same level were selected, and one was assigned to the experimental condition. The second was assigned to a control situation. The performance of both classes in geography was tested before and after the labs.



GRAPH 2 | Study 2: Scores of experimental and control group for geography test in test and retest phases.

Participants

A total of 25 students (11 males, 14 females) of about 12 years of age (mean = 12 years, standard deviation = 12 months) participated in the study. They belonged to two classes attending the first year of a secondary lower level school in Italy. One class, including 12 students (6 males, 6 females), was assigned to the experimental condition, and the other class, including 13 students (5 males, 8 females), was assigned to the control condition. One student with special needs was present in each class; they took part in the same activities performed by classmates.

The Geography Labs

Differently from Study 1, in Study 2 both classes undertook a laboratory in geography using an innovative method of 3 weekly lessons lasting 2 h each.

In the control condition (GeoLab), the activities described below were carried out in the students' classroom without robots. In the experimental condition (ER-GeoLab) the lessons were carried out using educational robots in the school's computer lab, supervised by an experimenter from our team. The experimenter didn't interact with the students but observed how the lessons developed, intervening only to solve technical problems related to robot functioning. In the ER-GeoLab, five BB8 Sphero toys and five Android Tablets were used. BB8 is composed of two spheres (the head and the body), and it belongs to the Star Wars film saga. It is simple to use and affordable. It may include different behaviors, when programmed using the free iOS and Android compatible apps.

Both laboratories were carried out by two teachers of Italian language, history, and geography. They worked together to avoid differences in the teaching and evaluation methods in the two classes. Before the beginning of the experiment, students in the experimental and control group had the opportunity to play with the robot and to attend a demonstrative session about the ways to

TABLE 2 | Mean and standard deviation of all measures collected in the experimental and control groups and results of statistical analyses.

	Experimental group				Control group				ANOVAs			*T*-tests
	Before		After		Before		After		Group	Time	Group × Time	
	M	SD	M	SD	M	SD	M	SD	$F_{(1, 23)}$	$F_{(1, 23)}$	$F_{(1, 23)}$	$t_{(1, 23)}$
Geography test	5.1	1.34	6.33	1.55	6	1.63	5.61	1.39	0.010	10.04*	39.50*	–
School achievement	35.00	9.62	–	–	39.54	15.61	–	–	–	–	–	0.866
Cognitive—motivational questionnaire	32.33	8.99	–	–	35.61	13.60	–	–	–	–	–	0.701

*$p < 0.05$.

assign specific movements to the robot by choosing directions, duration, and speed. The students in both groups were then divided into five groups of 2–3 children each to perform the activities described below.

First Lesson

During the first lesson, students in the GeoLab and the ER-GeoLab studied geography by preparing themselves cards for each Italian region, including information about the regional capital and other principal towns, environmental characteristics, economic aspects, and principal artistic and touristic attractions of the regional area. At the end of the activity, however, only the students in the ER-GeoLab had the opportunity to use the robots and to create a program allowing them to move along a blank Italian geography map.

Second Lesson

During the second lesson, teachers gave students in the ER-GeoLab target regions in which the robot had to stop. The students thus had a role in programming the robots' movements in order to reach each region and to describe the information about the area, using the "region cards" previously compiled.

Students in the GeoLab carried out a similar activity, but without robots: teachers gave students the target regions, and they had to reach them using a pawn, and then they had to describe the information about the area using the "region cards."

Third Lesson

To consolidate the learning of geography, during the third lesson, the students in the ER-GeoLab performed the same activity as in the second lesson, however, in this case, they could not use the cards, and instead, they had to recollect from memory and repeat the information about each region. Their classmates played a role in verifying the correctness of the given information. Again, students in the GeoLab carried out the same activity but with pawns rather than robots.

Measures

Geography Test

A geography test was given to the students in the experimental and control groups before and after the activities, to determine whether (1) there were differences between groups in geography knowledge before the labs; (2) the two groups showed a different level of geography learning after the ER-GeoLab and the GeoLab. The test consisted of placing the capitals of the Italian regions onto a blank Italian geography map. The students also had to describe the characteristics of the regions (rivers, mountains, primary, secondary, and tertiary sectors). Each student obtained a score from 1 to 10, depending on performance.

As in Study 1, the geography test was used as an objective measure of learning based on the answers given by students, and it was employed so that teachers were not affected by the experimental hypotheses when evaluating the students.

School Achievement and Cognitive-Motivational Style

Before the activities, teachers were requested to complete a school achievement grid and cognitive-motivational questionnaire for each student in both groups (as used in Caci and D'Amico, 2005; D'Amico, 2018). This was in order to measure the eventual individual differences in school achievement and cognitive-motivational styles of students in experimental and control groups, which could affect the results of the experiment.

The teachers from the team were requested to report the school marks (from 1, insufficient to 5, excellent) of each student in each of the 13 disciplines taught at school on the school achievement grid. A total school achievement score, ranging from 13 to 65, was then computed for each student.

The cognitive-motivational questionnaire was composed of 18 items that explored the cognitive and metacognitive level (e.g., recognition of their limits, focused attention while performing a task) and motivation (e.g., commitment, curiosity) for each student. For each item, the teachers had to assign a score using a Likert scale from 0 (not at all) to 3 (very much) based on how often they observed each behavioral style in their pupils. A total score, ranging from 0 to 72, was computed for each student.

The Pleasantness of the ER-GeoLab Activities

At the end of the activities, only students in the ER-GeoLab group compiled a short questionnaire (see **Table 3**) aimed at measuring the pleasantness of the activities performed with the BB8 robot.

Results

The mean and standard deviations of all the measures collected in the experimental and control groups and the results of statistical analyses are reported in **Table 2** and **Graph 2**.

To measure differences between the ER-GeoLab and the GeoLab groups in geography knowledge, a 2 × 2 factorial ANOVA using Group (ER-GeoLab, GeoLab) × Time (test, retest) was conducted before and after the labs, on the total geography test scores collected before and after the labs. The results showed that there was no main effect of Group [$F_{(1, 23)} = 0.01$, $p = 0.922$, $\eta^2 = 0.00$], indicating that there was no group difference in the general mean, including the performance of students before and after the labs. There was a significant effect of time [$F_{(1, 23)} = 10.04$, $p = 0.004$, $\eta^2 = 0.30$], indicating that both groups of students significantly increased their competence in geography knowledge after the labs. Finally, there was a significant interaction between Groups and Time [$F_{(1, 23)} = 39.50$, $p < 0.001$, $\eta^2 = 0.632$), indicating that after the lab, students in the ER-GeoLab group improved their knowledge about geography. At the same time, students in the GeoLab group experienced a slight decrease in their performance.

To evaluate whether there were significant differences among groups in school achievement that could have affected these results, a t-test was computed on the total scores they had obtained, however, there were no statistically significant group differences [$T_{(1, 23)} = 0.866$, $p = 0.396$], although students in the GeoLab group performed slightly better than students in the ER-GeoLab group (see **Table 2**). The same result was obtained for the scores of the questionnaire about cognitive and motivational styles: students in the GeoLab group achieved a higher score than students in the ER-GeoLab group (see **Table 1**), but again the group difference was not statistically significant [$T_{(1, 23)} = 0.701$, $p = 0.490$].

The results for the questionnaire about the pleasantness of the activities performed by students of the ER-GeoLab group with the BB8 robot showed that a high percentage of students evaluated the experience positively (see **Table 3**).

TABLE 3 | Rating by students about the level of the pleasantness of the ER-GeoLab activities.

Questions:	Not at all %	A little %	Enough %	Very much %
Have the activities been fun?	0	16.66	41.66	41.68
In your opinion, has the use of BB-8 facilitated your learning?	0	0	33.33	66.67
How much did you participate in the activities?	0	0	33.33	66.67

Discussion

The results of this study demonstrated that the use of ER in a geography lab has significant effects on learning geography. although both experimental and control groups performed similar activities in order to learn information about the Italian regions, only students in the experimental group obtained significant improvements in geography knowledge after the lab. Moreover, it has to be stressed that these results were obtained even though the experimental group reported slightly lower scores in general school achievement and in the questionnaire about cognitive-motivational styles during the school year.

ER stimulated participation and fun in students of the experimental group. They considered the activity facilitated learning, as documented by answers to a questionnaire about the pleasantness of activities.

GENERAL DISCUSSION AND CONCLUSIONS

In conclusion, the results of the two reported studies showed that the use of robotics at school improved the learning of the students involved in the activities. Although Study 1 and Study 2 are not directly comparable due to their numerous differences, the students belonging to the experimental group in both studies achieved better learning results than the students in the control groups.

In Study 1, even though there were no differences between the groups in their physics marks before the activities, students in the experimental group obtained higher scores than the controls in the school-type problems of physics included in the questionnaire. Thus, as already claimed, they generalized the acquired ability to solve the problem of Physics using robots (learning by doing) to the solution of other types of paper-pencil problems in Physics.

Conversely, and quite surprisingly, students in the experimental group obtained results similar to the controls in the items of the questionnaire that concerned robotics. All considered, it seems that it is enough to address the topic "robots" to improve learning, given its level of interest: indeed, the score on the questions about robotics was high for both groups. This also suggests that if, on the one hand, ER helps students to learn better about traditional subjects such as physics, on the other hand, robotics is such an innovative and attractive topic that students learn basic concepts about it even when a conventional teaching method is used.

In Study 2, we also demonstrated the efficacy of ER for teaching subjects in the humanities and not only in STEM subjects. Even when students in the experimental group had lower scores than the controls in school achievement and cognitive-motivational styles, they improved their knowledge of geography. In contrast, the controls showed a slight decrease in performance from the test to the retest. Moreover, as already claimed, both experimental and control groups realized a non-conventional learning lab, through group activities and playful teaching methodologies, and the only difference between the experimental and control group was the use of robots. We can thus confirm that the results obtained do not depend in general on the use of an innovative educational method, but the use of ER was also considered fun and useful by the students.

Why is robotics so valuable and attractive? There may be many reasons, as we have described in the introduction. From our perspective, the main reason is definitely that educational robotics combines physical and mental experience: according to the constructionist and EC theories, learning and cognitive functioning are affected by the physical and mental experience of interaction with the environment and with the tools it contains. It allows students to learn by doing, to manipulate concepts, and to embody cognition. In ER sessions, students have the opportunity to approach an idea from both an abstract and a concrete point of view. This leads to the creation of different forms of memory, such as semantic memory (i.e., memorizing the role and the function of each component) or the procedural memory (i.e., learning how each part works and how it has to be managed) to create accurate and complete episodic learning.

The second reason, as claimed by many authors, is that robotics may increase motivation for learning in situations that are generally seen by children as passive and not very stimulating. In Study 2, all children participating in ER-GeoLab declared that the activities were fun and useful for learning, and we already know that the so-called "digital native" generation (Prensky, 2001) is undoubtedly more interested in digital and multimedia technologies than previous generations. Children spend more time playing with technological tools than with "non-technological" tools. Beran et al. (2011) and Bullen and Morgan (2011) also demonstrated that this is not only limited to digital natives but also the so-called digital immigrants, people who started to use technologies as adults, and who are sometimes more passionate than children about technologies.

We also stress that the ER is not only a facilitator for students but also teachers. All teachers expressed great interest toward the use of robotics in teaching since it has positive effects on the school climate and may contribute significantly to creating a stimulating learning environment, both for students and teachers. The greater involvement of students in the topics dealt with using the robots makes it easier to keep the class's attention in the long-term and makes the teacher's work more rewarding. During and after the laboratory, the teachers involved in Studies 1 and 2 reported that the activities improved their feeling about the effectiveness of their teaching method, making not only students but also themselves more motivated. They not only consider robotics an excellent tool for supporting teaching, but they also see it as a significant change to the monotony of traditional education. In conclusion, the perception of their work status was improved.

LIMITATIONS AND FUTURE DIRECTIONS

Although the results obtained are encouraging and reinforce the idea that robotics can be considered a valuable teaching tool, there are many limitations in these studies that have to be considered and hopefully overcome in future studies. The first involves the experimental setting chosen for the activities. The school context presents many constraints: (1) it is not possible to realize a very innovative design by assigning single participants randomly to experimental or control conditions. Each class in a school, may participate or not as a whole, mainly due to problems involving the organization of school activities and timetables; (2) activities in groups could be challenging to organize. For instance, the contrasting leadership of two members can result in a group that does not collaborate very well, and the absence of one of the group members may affect the work of the other members; (3) there could be significant differences between two classes, due to the students that belong to them, but also due to the team of teachers, who can have different levels of competence or use different systems of evaluation or didactic methods; (4) In our study we chose classes with the same teachers to reduce these differences in teaching, however, using the same teacher in both experimental and control class, doesn't allow for "blind" testing. The teacher knows which is the experimental and the control class, and they are involved as the first person in the experience. Their increased enthusiasm for the new robot-mediated method may affect the teachers' behavior; (5) more studies involving longer interventions are needed. In particular, it will be necessary in the future to design longer-lasting interventions with different follow-up assessments to clarify whether, behind the initial motivational boost that robotics exerts in digital natives and teachers, robotics-based teaching leaves longer-lasting memory traces in students and allows a more in-depth and meta-cognitive comprehension of the studied topics than traditional methods.

Despite these limitations, it seems to us that educational robotics can have a significant impact on the school.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

All authors equally contributed to manuscript preparation and revision, read, and approved the submitted version.

ACKNOWLEDGMENTS

The studies were conducted in the frame of the PROMISE project, funded by MIUR (Italian Ministry of Education, University and Research), and carried out in Italy from 2017 to 2019 thanks to the collaboration of many public and private entities such as Scuola Superiore Sant'Anna (lead institution), the University of Milan-Bicocca, CNR (Institute of Educational Technologies of Palermo), University of Siena, University of Palermo, MetaIntelligenze Onlus, PalermoScienza, and the National Network of Schools for Robocup Junior Italy. The project aimed to create a national Italian network of entities working on educational robotics (schools, museums, and companies). The specific goal of the project was to train 36 teachers and 24 experts in the use of educational robotics in learning contexts. Teachers and experts attended three training courses (in the north, center, and south of Italy) consisting of 20 h of theoretical lessons and 36 h of internship in order to apply the skills acquired in schools and museums. We especially thank to the teacher Silvia Busacca, for the activities performed in Study 1 at the higher secondary school S. Cannizzaro in Palermo, and the teachers Vittoria Ruvituso and Anna Rollo for the realization of the activities reported in Study 2 at the lower secondary school Silvio Boccone in Palermo.

REFERENCES

Barak, M., and Zadok, Y. (2009). Robotics projects and learning concepts in science, technology and problem solving. *Int. J. Technol. Design Educ.* 19, 289–307. doi: 10.1007/s10798-007-9043-3

Barker, B. S., and Ansorge, J. (2007). Robotics as means to increase achievement scores in an informal learning environment. *J. Res. Technol. Educ.* 39, 229–243. doi: 10.1080/15391523.2007.10782481

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: a review. *Sci. Robotics* 3, 1–9. doi: 10.1126/scirobotics.aat5954

Beran, T. N., Ramirez-Serrano, A., Kuzyk, R., Fior, M., and Nugent, S. (2011). Understanding how children understand robots: perceived animism in child–robot interaction. *Int. J. Hum. Comp. Studies* 69, 539–550. doi: 10.1016/j.ijhcs.2011.04.003

Bullen, M., and Morgan, T. (2011). *Digital Learners Not Digital Natives*. La Cuestión Universitaria. 7, 60–68.

Caci, B., Chiazzese, G., and D'Amico, A. (2013b). Robotic and virtual world programming labs to stimulate reasoning and visual-spatial abilities. *Proc. Soc. Behav. Sci.* 93, 1493–1497 doi: 10.1016/j.sbspro.2013.10.070

Caci, B., and D'Amico, A. (2005). "Robotics: a new tool for education of subjects with cognitive diseases," in *Methods and Technologies for Learning, WIT Transaction on Information and Communication Technologies*, eds G. Chiazzese, M. Allegra, A. Chifari, and S. Ottaviano (Southampton: WIT Press), 563–567.

Caci, B., D'Amico, A., and Chiazzese, G. (2013a). "Robotics and virtual worlds: an experiential learning lab," in *Biologically Inspired Cognitive Architectures 2012* (Berlin, Heidelberg: Springer), 83–87.

D'Amico, A. (2018). The use of technology in the promotion of children's emotional intelligence: the multimedia program "Developing Emotional Intelligence". *Int. J. Emot. Educ.* 10, 47–67. Available online at: www.um.edu.mt/ijee

D'Amico, A., and Guastella, D. (2018). Robotics construction kits: from "objects to think with" to "objects to think and to emote with". *Fut. Internet* 10, 1–6. doi: 10.3390/fi10020021

D'Amico, A., and Guastella, D. (2019). The robotic construction kit as a tool for cognitive stimulation in children and adolescents: the RE4BES Protocol. *Robotics* 8, 1–10. doi: 10.3390/robotics8010008

Denis, B., and Hubert, S. (2001). Collaborative learning in an educational robotics environment. *Comput. Hum. Behav.* 17, 465–480. doi: 10.1016/S0747-5632(01)00018-8

Fong, T., Nourbakhsh, I., and Dautenhahn, K. (2003). A survey of socially interactive robots. *Robot. Autonom. Syst.* 42, 143–166. doi: 10.1016/S0921-8890(02)00372-X

Fridin, M., and Yaakobi, Y. (2011). "Educational robot for children with ADHD/ADD, architecture design," *Paper Presented at the International Conference on Computational Vision and Robotics* (ICCVR) (Bhubaneswar), 13–14.

Gibbs, R. W. Jr. (2005). *Embodiment and Cognitive Science*. Cambridge: Cambridge University Press.

Guastella, D., D'Amico, A., and Paci, G. (2020). *Protocollo RE4BES per il Potenziamento delle Abilità Cognitive, Emotive e Sociali di un Bambino con Disturbo Dello Spettro Autistico.* Sistemi Intelligenti, XXXII, 1.

Herrington, J., and Kervin, L. (2007). Authentic learning supported by technology: ten suggestions and cases of integration in classrooms. *EMI. Educ. Media Int.* 44, 219–236. doi: 10.1080/09523980701491666

Kanda, T., Shimada, M., and Koizumi, S. (2012). "Children learning with a social robot," in *Proceedings of the 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Boston, MA), 351–358.

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., et al. (2013). Social robots as embedded reinforcers of social behavior in children with autism. *J. Autism Dev. Disord.* 43, 1038–1049. doi: 10.1007/s10803-012-1645-2

Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. New York, NY: Basic Books, Inc. Publishers.

Papert, S. (1993). *The Children's Machine: Rethinking School in the Age of the Computer*. New York, NY: Basic Books, Inc. Publishers.

Papert, S., and Harel, I. (1991). Situating constructionism. *Constructionism* 36, 1–11.

Piaget, J., and Inhelder, B. (1966). *La Psychologie de l'Enfant*. Paris: Presses universitaires de France.

Prensky, M. (2001). Digital natives, digital immigrants part 1. *Horizon* 9, 1–6. doi: 10.1108/10748120110424816

Raskin, J. D. (2002). Constructivism in psychology: personal construct psychology, radical constructivism, and social constructionism. *Am. Commun. J.* 5, 1–25. Available online at: https://www.researchgate.net/publication/235930929_Constructivism_in_Psychology_Personal_Construct_Psychology_Radical_Constructivism_and_Social_Constructionism

Shamsuddin, S., Yussof, H., Ismail, L., Hanapiah, F. A., Mohamed, S., Piah, H. A., et al. (2012). "Initial response of autistic children in human-robot interaction therapy with humanoid robot NAO," in *Signal Processing and Its Applications (CSPA), 2012 IEEE 8th International Colloquium* (Melaka), 188–193.

Shapiro, L. (2010). *Embodied Cognition*. London: Routledge.

Toh, E., Poh, L., Causo, A., Tzuo, P. W., Chen, I., and Yeo, S. H. (2016). A review on the use of robots in education and young children. *J. Educ. Technol. Soc.* 19, 148–163. Available online at: https://www.researchgate.net/publication/301284225_A_Review_on_the_Use_of_Robots_in_Education_and_Young_Children (accessed March 02, 2020).

Turkle, S., and Papert, S. (1992). Epistemological pluralism and the revaluation of the concrete. *J. Math. Behav.* 11, 3–33.

Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge, MA: Harvard University Press.

Wei, C. W., and Hung, I. (2011). A joyful classroom learning system with robot learning companion for children to learn mathematics multiplication. *Turkish Online J. Educ. Technol.* 10, 11–23. Available online at: https://www.researchgate.net/publication/239443461_A_joyful_classroom_learning_system_with_robot_learning_companion_for_children_to_learn_mathematics_multiplication

Whittier, L. E., and Robinson, M. (2007). Teaching evolution to non-English proficient students by using LEGO robotics. *Am. Secondary Educ.* 35, 19–28. Available online at: http://www.jstor.org/stable/41406087

Williams, D. C., Ma, Y., Prejean, L., Ford, M. J., and Lai, G. (2007). Acquisition of physics content knowledge and scientific inquiry skills in a robotics summer camp. *J. Res. Technol. Educ.* 40, 201–216. doi: 10.1080/15391523.2007.10782505

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

BRIEF RESEARCH REPORT
published: 13 March 2020
doi: 10.3389/frobt.2020.00021

Identification of the Students Learning Process During Education Robotics Activities

David Scaradozzi[1,2]*, *Lorenzo Cesaretti*[1,3], *Laura Screpanti*[1] *and Eleni Mangina*[4]

[1] *Dipartimento di Ingegneria dell'Informazione (DII), Università Politecnica delle Marche, Ancona, Italy,* [2] *LSIS - umr CNRS 6168, Laboratoire des Sciences de l'Information et des Systèmes, Equipe I&M (ESIL), Marseille, France,* [3] *TALENT srl, Osimo, Italy,* [4] *School of Computer Science, University College Dublin, Dublin, Ireland*

This paper presents the design of an assessment process and its outcomes to investigate the impact of Educational Robotics activities on students' learning. Through data analytics techniques, the authors will explore the activities' output from a pedagogical and quantitative point of view. Sensors are utilized in the context of an Educational Robotics activity to obtain a more effective robot–environment interaction. Pupils work on specific exercises to make their robot smarter and to carry out more complex and inspirational projects: the integration of sensors on a robotic prototype is crucial, and learners have to comprehend how to use them. In the presented study, the potential of Educational Data Mining is used to investigate how a group of primary and secondary school students, using visual programming (Lego Mindstorms EV3 Education software), design programming sequences while they are solving an exercise related to an ultrasonic sensor mounted on their robotic artifact. For this purpose, a tracking system has been designed so that every programming attempt performed by students' teams is registered on a log file and stored in an SD card installed in the Lego Mindstorms EV3 brick. These log files are then analyzed using machine learning techniques (k-means clustering) in order to extract different patterns in the creation of the sequences and extract various problem-solving pathways performed by students. The difference between problem-solving pathways with respect to an indicator of early achievement is studied.

Keywords: educational robotics, educational data mining, learning analytics, STEM activities assessment, learning process identification

OPEN ACCESS

Edited by:
Dirk Ifenthaler,
Universität Mannheim, Germany

Reviewed by:
Robertas Damasevicius,
Kaunas University of Technology, Lithuania
Michael Vallance,
Future University Hakodate, Japan

Correspondence:
David Scaradozzi
d.scaradozzi@univpm.it

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *29 September 2019*
Accepted: *06 February 2020*
Published: *13 March 2020*

Citation:
Scaradozzi D, Cesaretti L, Screpanti L and Mangina E (2020) Identification of the Students Learning Process During Education Robotics Activities. Front. Robot. AI 7:21. doi: 10.3389/frobt.2020.00021

INTRODUCTION

Educational Robotics (ER) has been widely used to support integrative STEM education because of its power to realize engaging multidisciplinary activities about science, technology, engineering, and mathematics, but also arts, language, and humanities (Mubin et al., 2013; Scaradozzi et al., 2019a,b). Furthermore, ER can also support inclusive education (Daniela and Lytras, 2019) and computer science and robotics literacy at all ages (Burbaite et al., 2013; Štuikys et al., 2013; Berry et al., 2016; Damaševicius et al., 2017; Vega and Cañas, 2019). Even if many studies explored ER to motivate students to learn, not all of them reported an evaluation of activities; those who focused on the evaluation of ER activities adopted qualitative (Denis and Hubert, 2001; Liu, 2010; Elkin et al., 2014), quantitative (Atmatzidou and Demetriadis, 2016; Kandlhofer and Steinbauer, 2016; Cesaretti et al., 2017; Scaradozzi et al., 2019c), or mixed methods approaches (Kim et al., 2015; Chalmers, 2018). In fact, in an ER activity (a lesson characterized by one or more ER exercises), students design, build, program, debug, and share their robotic artifacts; ER is based on the constructionist approach proposed by Papert (1980): when pupils create personal and meaningful products, they "build" knowledge in their mind. This kind of educational activity is characterized by a workflow, modeled by Martinez and Stager (2013) with the "Think Make Improve" (TMI) cycle, where three different phases are repeated cyclically:

- At the beginning, the educator proposes a problem to solve so students usually start thinking and designing their solution ("Think" phase);
- Then learners build their product: in an ER activity, it could be a hardware (a prototype) or software (a sequence of instructions) creation ("Make" phase);
- At the end of the construction phase, students start the robot, observe and analyze its behavior, debugging errors or trying to optimize the performance of the artifact (Improve phase): pupils have to examine carefully the feedback of the robot in order to decide the next designing or programming steps (so the cycle starts again with the Think phase).

The evaluation of a product created during a constructionist activity can be a challenging and time-consuming activity (Berland et al., 2014). Moreover, it is often based on the final product and not on the process underlying the designed task (Blikstein, 2011). However, new data mining and machine learning technologies allow researchers to capture detailed data related to problem-solving and programming trajectory of a large number of learners (Blikstein et al., 2014).

Recent studies (Berland et al., 2013; Blikstein et al., 2014; Chao, 2016; Wang et al., 2017; Bey et al., 2019; Filvà et al., 2019) have mostly applied machine learning techniques to data gathered from students during programming activities without the presence of physical robots, obtaining good results in the identification of different patterns in specific coding tasks (**Table 1** summarizes machine learning techniques and features selected in these studies). Berland et al. (2013) and Chao (2016) used a k-means algorithm to discover patterns in the programming activity of novice programmers; the first study identified three general patterns (Tinkering, Exploring, and Refining) and presented a positive correlation between the quality of the programming sequences designed by the students and two of the emerged patterns (Tinkering and Refine). The second study represented the students' programming activity using five indicators and identified four clusters (sequent approach, selective approach, repetitious approach, and trial approach); the study showed that the performance was lower for learners in the trial approach cluster compared to the sequent and repetitious approach clusters. Blikstein et al. (2014) proposed two experiments using different machine learning techniques, trying to discover patterns in data collected from 370 undergraduate students and to predict their midterm and final exam grades. They obtained best results modeling students' programming trajectories using hidden Markov models and demonstrated that the group in which a student was clustered into was predictive of his or her midterm grade. Wang et al. (2017) used log data from Code.org[1] and applied a long short-term memory recurrent neural network to predict students' future performance, obtaining good results in terms of accuracy and recall. Bey et al. (2019) identified three clusters in a dataset created collecting programs from 100 students registered on a 3-week course on the essential of Shell programming; they applied unsupervised clustering techniques (Hopkins statistic methods) for automatically identifying learners' programming behavior. Filvà et al. (2019) used the k-means technique on data generated by students' clicks in Scratch (and not on handpicked features), with the objective of categorizing learners' behavior in programming activities: they identified three different patterns and a strong correlation between these behaviors and the evaluation given by some teachers involved in the research project, using a rubric for programming assessment.

However, only one research study (to the best of our knowledge) applied machine learning to data collected during ER activities (Jormanainen and Sutinen, 2012); they did not collect data related to the programming sequences designed by the students but related to the pupils' interactions with the essential elements of the visual programming environment. Their system, using trees algorithm (J48 implementation), classified the students' activities into four classes, differentiating the observed students' group's progress with the purpose of identifying pupils with difficulties during the robot programming task. Ahmed et al. (2018) presented an interesting system that gives feedback to pupils in real time while they are programming the Lego Mindstorms EV3 robot[2]; they implemented a system (ROBIN) so that the Lego Mindstorms EV3 robot provided reflective feedback to pupils, transforming it into a learning companion: using ROBIN, students obtained advices based on the sequences created on the programming environments and based on the exercise proposed by the educator. But in this research project, the researchers did not train their system using machine learning techniques but using deterministic rules. The promising results obtained using machine learning techniques on data gathered from students during programming activities, and the lack of this type of study in the field of ER (Scaradozzi et al., 2019b), have prompted the research described in this paper. Thanks to an upgrade of the Lego Mindstorms EV3 programming blocks (implemented by the authors), it was possible to register some log files containing the programming sequences created by 353 Italian primary and secondary school students (organized in 85 teams) during the resolution of a robotics exercise related to the ultrasonic sensor. Integrating sensors allowed learners to obtain an interaction between the robot and the environment, but to effectively use these devices, they had to understand some key concepts about robotics and computer science, such as how to acquire and store data, how to cyclically repeat an acquisition (using loops), and how to create algorithms to obtain different robot's behaviors depending on the values detected by sensors (using conditional statements). The authors inputted the collected log files into k-means algorithms, with the purpose of verifying if there are different problem-solving patterns emerging from this dataset and of examining the interrelationships between the different problem-solving

[1]https://code.org

[2]https://education.lego.com/en-us/middle-school/intro/mindstorms-ev3

TABLE 1 | Features and machine learning techniques of recent studies carried out in constructionist environments.

Paper	Features selected in the experimentation	Machine learning techniques
Blikstein et al. (2014)	[1st experiment] Code update differential, characterized by: number of lines added, lines deleted, lines modified, characters added, characters removed, characters modified.	[1st experiment] Simple regression between exam grades and average size of the code updates per student.
	[2nd experiment] Code update differential	[2nd experiment] X-means clustering algorithm.
	[3rd experiment] Code update differential	[3rd experiment] X-means clustering algorithm.
	[4th experiment] Code update curves (combination of frequency and size in changes made by students).	[4th experiment] Dynamic time warping and scaled dynamic time warping distance (to calculate the difference between two given code update curves).
	[5th experiment] Modeling of a student's trajectories as a hidden Markov model (HMM).	[5th experiment] k-medioid and hierarchical agglomerative clustering (to compute the different states of the HMM). Expectation maximization algorithm to compute both the transition and emission probabilities in the state diagram.
Berland et al. (2013)	Measures of individual program states (measures calculated for each program state) considering five features: action, logic, unique primitives, length, coverage.	X-Means clustering algorithm
Jormanainen and Sutinen (2012)	Six events: add statement, add command to code, remove line, upload program to robot, compiling errors, sum of all these events.	Decision trees, decision tables, Bayesian networks, and multilayer perceptrons to predict the students' progress. To measure the accuracy of the tested algorithms, they used the 10-fold cross-validation method.
Chao (2016)	Related to computational practice (five measures): sequence, selection, simple iteration, nested iteration, testing.	Ward's minimum variance method (to identify number of clusters), followed by the k-mean cluster analysis (on the identified cluster number).
Wang et al. (2017)	[1st experiment] A student's trajectory consists of all the program submissions, which are represented as ASTs (that contain all the information about a program and can be mapped back into a program). These ASTs are converted into program embeddings using a recursive neural network.	[1st experiment] Long Short-Term Memory (LSTM) Recurrent Neural Network (RNN).
	[2nd experiment] Same features of their 1st experiment, from ASTs they calculated program embeddings.	[2nd experiment] LSTM RNN
Bey et al. (2019)	Number of submissions, average time between two submissions, average number of changes, percentage of syntactical errors, time standard deviation (the standard deviation of the average time between two submissions), code standard deviation (the standard deviation of the average number of changes).	Mixture Gaussian Clustering algorithm
Filvà et al. (2019)	Clickstream	K-means cluster analysis

patterns and a performance indicator showing the students' team capability to reach a working program solution.

METHODS

Procedure

At the first stage of this research project, authors implemented a *software modification* to the Lego Mindstorms EV3 Education Software blocks; thanks to this software development, every time that students tested their program on the robot Lego Mindstorms EV3, a "track" of the coding sequence was written in a log file stored in the SD card mounted on the robot. Fourteen schools participated in the experimentation, and the same protocol was performed for each of them. Firstly, an educator of TALENT srl (an Italian innovative startup involved in the research project) installed on the computers of the school the official Lego Mindstorms EV3 Education software and the update designed by the authors. An "Introduction to Robotics" course was then realized, taught in collaboration with TALENT; Constructionism (Papert, 1980) and problem-based learning (Savery, 2006) were the pedagogical approaches underlying the proposed course: during each lesson, students designed and created programming solutions to problems related to the robot. After a first part dedicated to the robot's actuators, the ultrasonic sensor was explored. An exercise was proposed by the instructor to the students: learners had to program the robot so that it stopped at a given distance from the wall, trying to be as precise as possible; they also had to consider a constraint: the maximum available time to design and test their coding solution (20 min for higher secondary school classes; 30 min for lower secondary and primary school classes). There are some elements that make this exercise quite tricky for novice students in robotics: they

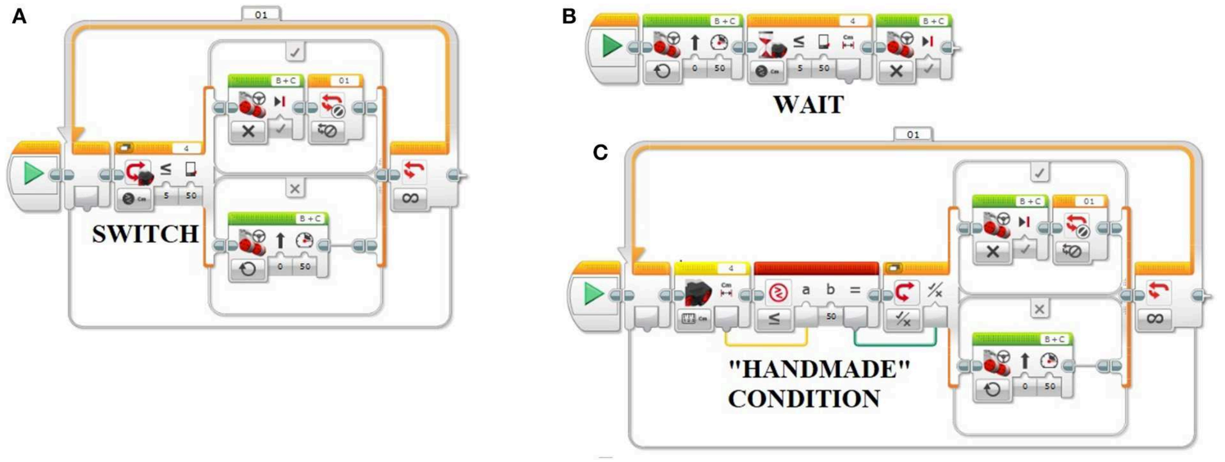


FIGURE 1 | Three possible solutions for the exercise designed using the Lego Mindstorms EV3 software blocks: "Switch case" **(A)**, "Wait" **(B)**, and "Handmade condition" **(C)**.

have to think about how to set the condition related to the ultrasonic sensor, how to use the iteration (loop block), and how to compensate the braking distance (the robot does not stop immediately when the EV3 brick sends the command "turn off" to the motors). **Figure 1** shows three possible solutions to this problem: the simplest sequence (B) in terms of the number of blocks contains the Wait block; it makes the program wait for a condition becoming "true" before continuing to the next block in the sequence. The intermediate solution (A) contains the Switch block; this block is a container that can comprehend two or more sequences of programming blocks; a test at the beginning of the Switch determines which Case will run, and in this case, the test is designed on the ultrasonic sensor. The most complex solution (C) contains a "handmade" condition, created using a Sensor block (the yellow one) and a Compare block (the red one). Students' teams involved in the experimentation were free to design and test their programming solution (usually close to one of the sequences presented in **Figure 1**): the educator only explained the general meaning and the parameters of the useful blocks, and then the pupils started to work on their program.

At the end of the exercise, all the log files generated by the tracking system were downloaded from the SD card by the TALENT's educator and stored in the cloud storage.

The authors fed the collected log files (transformed into vectors thanks to a parsing system developed in Python) into a k-means algorithm, whose results provided clusters that represent different types of sequences designed by the students to solve the exercise. Then, for each team of the students involved in the experimentation, the number of sequences belonging to each cluster was calculated in order to get new features that characterized the students' programming activity (all the programming actions carried out by the participants with the intention to obtain the desired robot's behavior). These new features were used again as input data for a k-means algorithm, and different problem-solving behaviors emerged from this last step. An expert robotics educator defined for each log file the first working sequence created by the students' team, which allowed the educator to observe in which stage of the problem-solving process learners created their first working sequence. A working sequence is a program that can solve the exercise previously presented, and the conditions to be met are: correct conditional statement on the ultrasonic sensor and motors turned on using the right modality. Then, applying the formula:

$$\textit{Indicator of early achievement} = \frac{n^{\circ} \textit{ of the first working sequence}}{\textit{total tests number}}$$

Finally, a one-way non-parametric ANOVA (Kruskal–Wallis) test was conducted to examine the differences in the indicator according to the different problem-solving behaviors, which emerged from the machine learning technique. Moreover, the *post-hoc* Dunn test (Dunn, 1964), appropriate for groups with unequal numbers of observations (Zar, 2010), was employed to examine the significance of all possible pairwise comparisons among clusters.

Participants

From March 2018 to September 2019, a total of 353 students from 14 Italian primary and lower/higher secondary schools (located in the Emilia Romagna and Marche regions) were involved in this study. Sixty-two students divided into 19 teams [Average Age (AA) = 17.29, Standard Deviation (SD) = 0.55] from school 1 were involved. School 2 had 22 students involved, divided into six teams (AA = 11.45, SD = 0.50). School 3 had 24 students involved, divided into six teams, but valid data were collected only from two of them (AA = 10.08, SD = 0.65). School 4 had 21 students involved, divided into five teams (AA = 11.70, SD = 0.47). School 5 had 19 students involved, divided into seven teams (AA = 11.63, SD = 0.83). School 6 had 25 students involved, divided into five teams (AA = 15.92, SD =0.28). School 7 had 24 students involved, divided into six teams, but valid data were collected only from three of them (AA = 12.00, SD = 0.46). School 8 had 23 students involved, divided into five teams (AA = 12.43, SD = 0.94). School 9 had 30 students, divided into six teams (AA = 9.63, SD = 0.53). School 10 had 26 students involved, divided into six teams (AA = 12.54, SD =0.51). School 11 had 19 students involved, divided into five teams (AA = 10.21, SD = 0.98). School 12 had nine students involved, divided into three teams (they were from lower secondary school, but no personal data were available). School 13 had 23 students involved, divided into six teams (AA = 11.87, SD = 1.29). School 14 had 26 students involved, divided into eight teams (AA = 10.24, SD = 0.83).

Data Preparation

Students' teams designed 3,292 programming sequences to solve the robotics exercise previously described. Some technical steps are performed to transform these sequences into matrices; after this transformation, the following 12 indicators are calculated for each programming sequence. A function designed in Python realizes the parsing of the log file to calculate these 12 values.

- **Motors**: how many Motor blocks are contained in the sequence;
- **Loops**: how many Loop blocks are contained in the sequence;
- **Conditionals**: how many Conditional and Sensors blocks are contained in the sequence;
- **Others**: how many blocks are contained in the sequence belonging to different categories than Motors, Loops, and Conditionals;
- **Added**: how many blocks have been added, compared to the previous sequence;
- **Deleted**: how many blocks have been deleted, compared to the previous sequence;
- **Changed**: how many blocks have been changed, compared to the previous sequence;
- **Equal**: how many blocks have remained unchanged, compared to the previous sequence;
- **Delta Motors**: the amount of change in Motor blocks parameters, compared to the previous sequence (calculated only for blocks of the "Changed" category);
- **Delta Loops**: the amount of change in Loop blocks parameters, compared to the previous sequence;
- **Delta Conditionals**: the amount of change in Conditional blocks parameters, compared to the previous sequence;
- **Delta Others**: the amount of change in Other blocks parameters, compared to the previous sequence.

The authors decided to calculate the first four indicators (Motors, Loops, Conditionals, Others) because they represent the features of a sequence designed using the Lego Mindstorms EV3 software; moreover, they are key concepts in ER and computational curricula (Grover and Pea, 2013; Scaradozzi et al., 2015, 2019b; Allsop, 2019). Furthermore, as previously stated, an ER activity is characterized by a cyclical procedure for improving the programming sequence: for this reason, it is essential to calculate the differences between two contiguous sequences, represented by the last eight parameters (Added, Deleted, Changed, Equal, Delta Motors, Delta Loops, Delta Conditionals, and Delta Others). Each programming sequence designed by the learners is thus represented using these 12 indicators, and it can be considered as a point in the problem-solving trajectory (Berland et al., 2013) carried out by the students' team.

RESULTS

Clusters resulting from the application of k-means algorithm on programming sequences designed by the students' teams are shown in **Table 2**. Fourteen clusters were identified applying the Elbow Method (Kodinariya and Makwana, 2013), and their relation to teams' behavior is briefly reported.

Cluster 1: the team tested the same programming sequence several times (characterized by four blocks, similar to solution B in **Figure 1**); 32.99% of the sequences are categorized in this cluster.

Cluster 2: the team changed the condition and the threshold value for the ultrasonic sensor throughout the programming attempts (programming sequence similar to solution B in **Figure 1**); 3.25% of the sequences are categorized in this cluster.

Cluster 3: the team heavily changed the condition and the threshold value for the ultrasonic sensor (i programming sequence similar to solution B in **Figure 1**); 2.13% of the sequences are categorized in this cluster.

Cluster 4: the team refined the threshold value for the ultrasonic sensor and some parameters in a Motors block at the same time (in a programming sequence similar to solution A in **Figure 1**); 6.71% of the sequences are categorized in this cluster.

Cluster 5: the team refined both some parameters in a Motors block and some parameters in Others blocks (programming sequence similar to solution A in **Figure 1**); 0.18% of the sequences are categorized in this cluster.

Cluster 6: the team modified some parameters in a Loops block (in a programming sequence similar to solution A in **Figure 1**); 0.03% of the sequences are categorized in this cluster.

Cluster 7: the team heavily modified some parameters in a Motors block and refined the threshold value for the ultrasonic sensor (in a programming sequence similar to solution A or B in **Figure 1**); 1.64% of the sequences are categorized in this cluster.

Cluster 8: the team tested the same programming sequence (characterized by 11–12 blocks, similar to solution A or C in **Figure 1** with the addition of Others block); 4.19% of the sequences are categorized in this cluster.

Cluster 9: the team tested the same programming sequence (characterized by eight to nine blocks, similar to solution A or B in **Figure 1**); 24.14% of the sequences are categorized in this cluster.

Cluster 10: the team refined the threshold designed for the ultrasonic sensor (programming sequence like solution A in **Figure 1**) and added two blocks; 4.04% of the sequences are categorized in this cluster.

Cluster 11: the team refined both the threshold for the ultrasonic sensor and some parameters in a Motors block and deleted two blocks (in a programming sequence similar to solution A in **Figure 1**); 4.04% of the sequences are categorized in this cluster.

Cluster 12: the team refined both the threshold designed for the ultrasonic sensor and some parameters in a Motors block

TABLE 2 | Mean values and standard deviation values of the 12 indicators (reported in the section Data Preparation) calculated for each cluster, as presented in the section Results (M(SD)).

	Same	Modified	Added	Deleted	DeltaMotors	DeltaLoops	DeltaConditionals	DeltaOthers	Motors	Loops	Conditionals	Others
Cluster 1	3.80 (1.23)	0.00 (0.00)	0.04 (0.19)	0.03 (0.18)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.89 (0.31)	0.21 (0.62)	1.99 (0.33)	1.03 (0.16)
Cluster 2	4.11 (1.54)	1.09 (0.29)	0.50 (0.71)	0.20 (0.40)	0.80 (3.87)	0.00 (0.00)	41.77 (8.82)	0.00 (0.00)	1.37 (0.61)	1.05 (1.18)	2.17 (0.40)	1.06 (0.23)
Cluster 3	4.39 (1.68)	1.14 (0.39)	0.63 (0.85)	0.30 (0.67)	0.93 (6.21)	0.00 (0.00)	101.92 (41.18)	0.00 (0.00)	1.49 (0.53)	1.20 (1.29)	2.36 (0.48)	1.11 (0.32)
Cluster 4	7.13 (1.03)	1.97 (0.36)	0.06 (0.24)	0.04 (0.19)	1.37 (5.00)	0.00 (0.00)	4.48 (6.26)	0.00 (0.00)	1.98 (0.13)	2.83 (0.60)	3.00 (0.15)	1.33 (0.60)
Cluster 5	7.50 (1.05)	2.83 (0.75)	0.00 (0.00)	0.00 (0.00)	6.67 (8.76)	0.00 (0.00)	0.60 (0.67)	20.00 (3.85)	2.00 (0.00)	3.00 (0.00)	3.00 (0.00)	2.33 (0.52)
Cluster 6	5.00 (0.00)	1.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	20.00 (0.00)	0.00 (0.00)	0.00 (0.00)	1.00 (0.00)	2.00 (0.00)	2.00 (0.00)	1.00 (0.00)
Cluster 7	3.78 (1.99)	1.44 (0.57)	0.31 (0.84)	0.11 (0.42)	58.78 (24.34)	0.00 (0.00)	14.51 (21.81)	0.02 (0.14)	1.37 (0.49)	0.76 (1.08)	2.13 (0.48)	1.06 (0.30)
Cluster 8	11.67 (1.91)	0.11 (0.38)	0.23 (0.58)	0.09 (0.32)	0.47 (2.33)	0.00 (0.00)	0.14 (0.91)	0.01 (0.12)	1.99 (0.12)	3.11 (0.79)	2.98 (0.19)	4.11 (1.07)
Cluster 9	8.51 (0.82)	0.03 (0.17)	0.04 (0.20)	0.04 (0.19)	0.06 (0.98)	0.00 (0.00)	0.03 (0.23)	0.00 (0.00)	1.92 (0.28)	2.67 (0.51)	2.64 (0.50)	1.16 (0.37)
Cluster 10	5.50 (1.80)	0.25 (0.50)	2.64 (1.13)	0.24 (0.51)	0.88 (4.72)	0.00 (0.00)	3.86 (11.23)	0.00 (0.00)	1.74 (0.56)	2.17 (0.96)	2.81 (1.02)	1.24 (0.66)
Cluster 11	5.96 (2.04)	0.37 (0.55)	0.32 (0.61)	2.30 (0.87)	1.67 (8.11)	0.00 (0.00)	6.10 (13.76)	0.00 (0.09)	1.08 (0.46)	1.99 (0.85)	2.05 (0.74)	1.33 (0.80)
Cluster 12	3.04 (0.66)	1.03 (0.18)	0.05 (0.21)	0.06 (0.23)	2.01 (6.03)	0.00 (0.00)	4.70 (6.47)	0.00 (0.00)	1.02 (0.14)	0.08 (0.40)	1.98 (0.25)	1.00 (0.00)
Cluster 13	4.64 (1.05)	0.17 (0.38)	0.05 (0.22)	0.03 (0.17)	0.05 (0.64)	0.00 (0.00)	0.56 (1.91)	0.00 (0.00)	2.02 (0.13)	0.13 (0.52)	2.03 (0.18)	1.00 (0.06)
Cluster 14	11.4 (1.79)	0.53 (0.90)	0.33 (0.84)	0.20 (0.55)	1.13 (6.21)	0.00 (0.00)	1.87 (7.34)	0.00 (0.00)	2.33 (0.48)	2.2 (0.66)	6.37 (0.61)	1.37 (0.49)

(programming sequence similar to solution B in **Figure 1**, but with only one Motors block); 6.35% of the sequences are categorized in this cluster.

Cluster 13: the team refined both the threshold for the ultrasonic sensor and some parameters in a Motors block (in a programming sequence similar to solution B in **Figure 1**, with two Motors blocks); 9.39% of the sequences are categorized in this cluster.

Cluster 14: the team refined both the threshold designed for the ultrasonic sensor and some parameters in a Motors block (in this case, the sequence is extremely complex, characterized by 11–12 blocks, similar to solution C in **Figure 1**); 0.91% of the sequences belong to this cluster.

Figure 2 presents the silhouette scores (Rousseeuw, 1987) for the 14 clusters identified by the k-means algorithm. **Table 3** shows the Pearson correlation between these clusters and the indicator of early achievement: only cluster 3 shows a statistically significant positive correlation (Pearson coefficient correlation = 0.411, $p < 0.0001$); so teams that heavily changed the condition and the threshold value for the ultrasonic sensor did not obtain a working sequence in the first part of their work.

As previously stated in the section *Procedure*, after having clustered the students' programming sequences, the percentage of sequences belonging to each cluster was calculated for each group. Thus, the problem-solving process for each team was represented using a vector with 14 elements, the percentage of coding sequences in cluster 1, the percentage of coding sequences in cluster 2, etc. A matrix (size: 85 × 14) created considering these 14 features calculated for the 85 teams was then used as inputs for a k-means algorithm, with the aim of grouping teams with similar behavior. Applying again the Elbow Method (Kodinariya and Makwana, 2013), 10 different problem-solving pathways emerged (**Table 4**):

- Pathway 1: Prevalence of sequences belonging to cluster 9 and cluster 4; these teams designed a complex sequence (type A or C in **Figure 1**), generally refining the parameters, with a very low percentage of large changes in the condition or in the threshold for the ultrasonic sensor and implementing a quite high number of trials (18 teams in this cluster, 21.18%).
- Pathway 2: 17% of sequences belonging to clusters 3 and 4; these teams applied high changes in the condition and in the threshold designed for the ultrasonic sensor (eight teams in this cluster, 9.41%).
- Pathway 3: Prevalence of sequences belonging to cluster 13; these teams designed a compact sequence (type B in **Figure 1**) generally refining the threshold designed for the ultrasonic sensor and some parameters in a Motors block (eight teams in this cluster, 9.41%).
- Pathway 4: Prevalence of sequences belonging to cluster 1; these teams designed a compact sequence (type B in **Figure 1**) sometimes (14%) refining the threshold designed for the ultrasonic sensor and some parameters in a Motors block, sometimes (8%) applying high changes to the condition or to the threshold related to the ultrasonic sensor or to the Motors' parameters; a very high number of trials characterized this cluster (22 teams in this cluster, 25.88%).

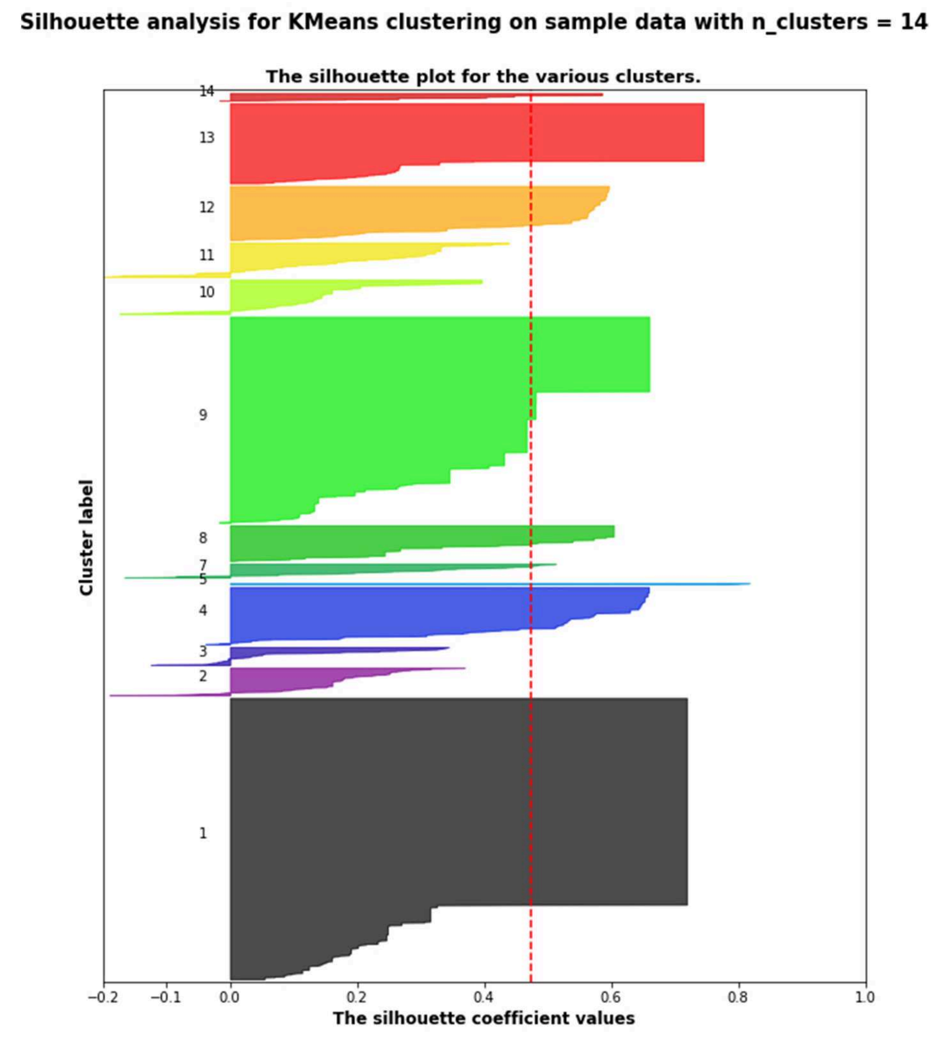


FIGURE 2 | Silhouette scores for the 14 clusters (presented in the section Results) identified by the k-means algorithm.

- Pathway 5: Prevalence of sequences belonging to clusters 8 and 4; these teams designed a very complex sequence (type A or C in **Figure 1**), using also Others blocks and generally refining the parameters of the programming blocks (three teams in this cluster, 3.53%).
- Pathway 6: Relevant percentage (32%) of sequences belonging to clusters 10 and 11; these teams repeatedly deleted and added blocks to their sequence (similar to type A or B in **Figure 1**); a low number of trials characterized this cluster (13 teams in this cluster, 15.29%).
- Pathway 7: Prevalence of sequences belonging to clusters 8 and 4; the team designed a complex sequence using also four Others blocks (type A in **Figure 1**), generally refining the parameters, without any sequence with large changes in the condition or in the threshold for the ultrasonic sensor; this team also experimented some simple sequences (cluster 1, type B in **Figure 1**) (one team in this cluster, 1.18%).
- Pathway 8: Prevalence of sequences belonging to cluster 14; these teams designed the most complex sequences (type C in **Figure 1**), generally refining the parameters, without any sequence with large changes in the condition or in the threshold for the ultrasonic sensor (two teams in this cluster, 2.35%).
- Pathway 9: Relevant percentage (32%) of sequences belonging to clusters 10 and 11; the team repeatedly deleted and added blocks to their sequence (similar to type A in **Figure 1**) and repeatedly changed parameters in the programming blocks (36% of sequences in cluster 4); a low number of trials characterized this cluster (one team in this cluster, 1.18%).
- Pathway 10: the lowest number of trials (18) and a relevant percentage (11%) of sequences in cluster 7 (high changes in Motors parameters) characterized these teams (nine teams in this cluster, 10.59%).

TABLE 3 | Pearson correlation coefficient between clusters (see **Table 2**) and the indicator of early achievement.

Cluster	Correlation	*P*-value
Cluster 3	0.411	< 0.0001
Cluster 13	0.124	0.259
Cluster 2	0.122	0.266
Cluster 11	0.014	0.897
Cluster 1	−0.006	0.956
Cluster 12	−0.009	0.936
Cluster 6	−0.024	0.826
Cluster 5	−0.024	0.826
Cluster 14	−0.037	0.735
Cluster 8	−0.049	0.654
Cluster 9	−0.057	0.602
Cluster 10	−0.082	0.453
Cluster 4	−0.088	0.421
Cluster 7	−0.122	0.267
Trials	−0.128	0.243

Figure 3 shows the silhouette scores for the 10 pathways presented above; **Figure 4** is obtained after applying a two-dimensional principal component analysis (PCA): it presents the distribution of the identified pathways implemented by the students' teams along two principal components, calculated according to the PCA approach.

Figure 5 presents the age-related differences between the students' teams involved in the experimentation, within these 10 pathways: the majority of the higher school students adopted pathways 1 (a complex sequence with some refinements of the programming parameters) and 6 (a complex sequence with considerable variation to the condition set for the ultrasonic sensor); the majority of the lower school students adopted pathways 3 (a compact sequence with a refinement of the programming parameters) and 4 (a compact sequence with considerable variation to the condition set for the ultrasonic sensor); the majority of the primary school students adopted pathway 4 (a compact sequence with considerable variation to the condition set for the ultrasonic sensor).

Figure 6 shows the distributions of the indicator of early achievement in the 10 selected cluster. Excluding those problem-solving behaviors that were shown by less than three groups (styles 7, 8, 9) from the analysis, significant differences (chi-square $= 25.54$, $p = 0.0002711$, $d_f = 6$) were found among the seven different clusters of group behavior. Pairwise comparisons using Dunn's test for multiple comparisons of independent samples with Bonferroni's *P*-value adjustment method showed that significant differences could be found between clusters 1 and 4 ($p = 0.014$), 2 and 5 ($p = 0.035$), and 4 and 5 ($p = 0.016$).

DISCUSSION

This brief research report presents an innovative application of machine learning techniques in the field of ER, for the

TABLE 4 | mean (M) and standard deviation (SD) of the 10 pathways (P) resulting from the second clustering algorithm (M(SD)).

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11	Cluster 12	Cluster 13	Cluster 14	Trials
P1	0.04 (0.05)	0.02 (0.03)	0.02 (0.02)	0.24 (0.17)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.56 (0.20)	0.05 (0.03)	0.05 (0.04)	0.00 (0.01)	0.01 (0.02)	0.01 (0.03)	45.67 (25.53)
P2	0.25 (0.27)	0.07 (0.05)	0.1 (0.03)	0.04 (0.05)	0.00 (0.00)	0.00 (0.00)	0.00 (0.01)	0.00 (0.00)	0.27 (0.23)	0.05 (0.05)	0.05 (0.04)	0.01 (0.02)	0.16 (0.24)	0.00 (0.00)	33.38 (17.94)
P3	0.07 (0.07)	0.08 (0.11)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.01 (0.04)	0.00 (0.00)	0.05 (0.12)	0.00 (0.01)	0.01 (0.03)	0.01 (0.02)	0.76 (0.23)	0.00 (0.00)	31.75 (19.75)
P4	0.72 (0.16)	0.04 (0.03)	0.02 (0.01)	0.01 (0.03)	0.00 (0.00)	0.00 (0.00)	0.02 (0.03)	0.00 (0.00)	0.02 (0.05)	0.01 (0.02)	0.01 (0.02)	0.14 (0.15)	0.02 (0.05)	0.00 (0.00)	56.95 (30.75)
P5	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.09 (0.15)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.81 (0.16)	0.01 (0.02)	0.03 (0.03)	0.06 (0.02)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	42.67 (36.14)
P6	0.17 (0.11)	0.02 (0.03)	0.01 (0.02)	0.02 (0.06)	0.00 (0.00)	0.00 (0.00)	0.01 (0.02)	0.00 (0.00)	0.42 (0.12)	0.18 (0.07)	0.14 (0.06)	0.01 (0.02)	0.03 (0.09))	0.00 (0.00)	22.15 (15.02)
P7	0.05 (0.00)	0.00 (0.00)	0.00 (0.00)	0.13 (0.00)	0.11 (0.00)	0.00 (0.00)	0.02 (0.00)	0.47 (0.00)	0.16 (0.00)	0.04 (0.00)	0.02 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	55.00 (0.00)
P8	0.05 (0.07)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.03 (0.04)	0.11 (0.06)	0.09 (0.05)	0.02 (0.04)	0.03 (0.04)	0.67 (0.02)	18.00 (2.83)
P9	0.12 (0.00)	0.00 (0.00)	0.00 (0.00)	0.36 (0.00)	0.00 (0.00)	0.04 (0.00)	0.00 (0.00)	0.00 (0.00)	0.16 (0.00)	0.12 (0.00)	0.2 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	25.00 (0.00)
P10	0.20 (0.17)	0.01 (0.02)	0.00 (0.01)	0.02 (0.04)	0.00 (0.00)	0.00 (0.00)	0.11 (0.04)	0.00 (0.01)	0.27 (0.29)	0.08 (0.03)	0.05 (0.05)	0.04 (0.05)	0.21 (0.25)	0.00 (0.00)	18.22 (8.42)

Columns report clusters obtained from the first clustering technique (as illustrated in the section Results). Trials represents the mean number of tests carried out by the team of students.

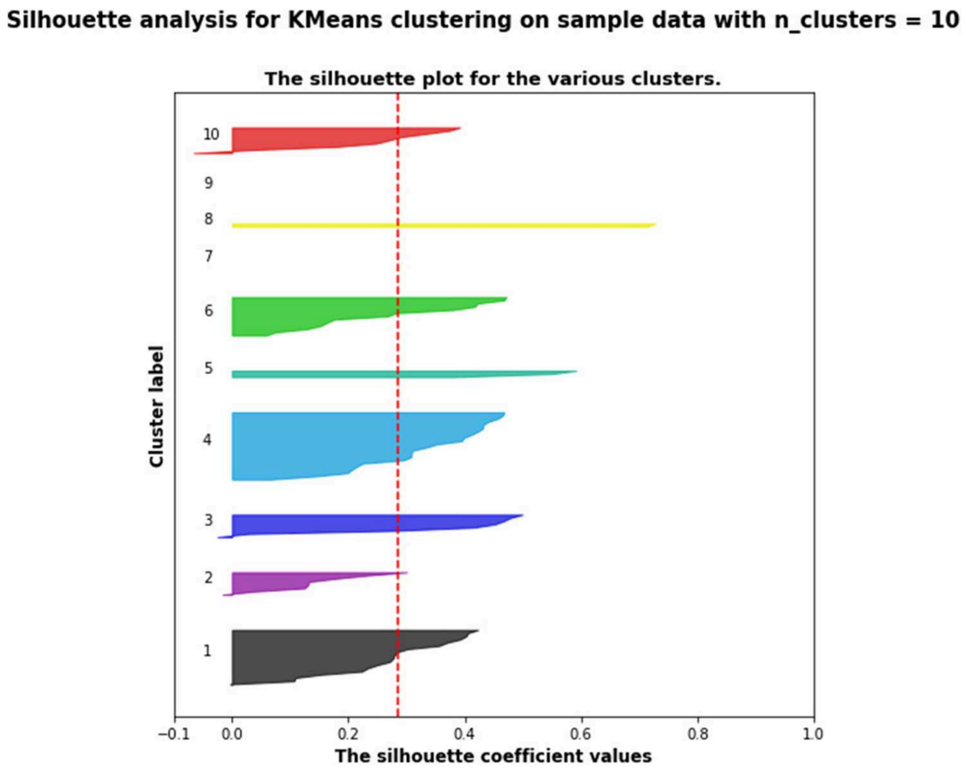


FIGURE 3 | Silhouette scores for the 10 pathways (presented in the section Results) identified by the k-means algorithm.

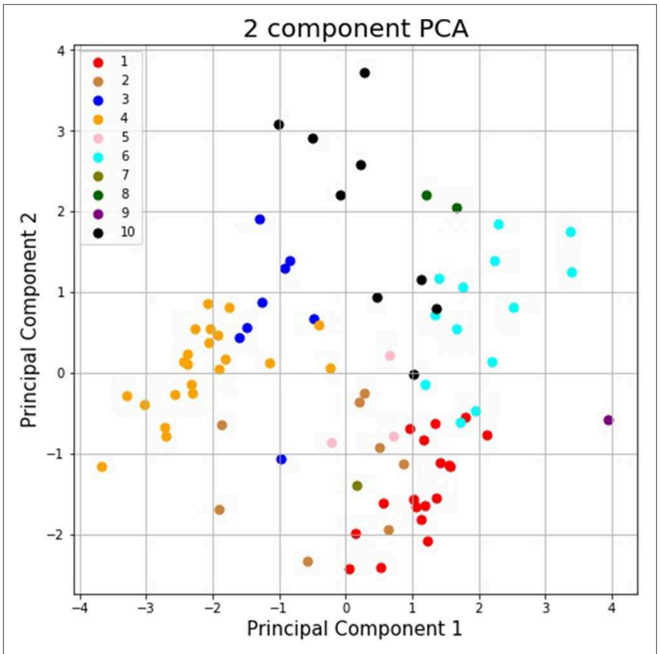


FIGURE 4 | Representation of the identified pathways implemented by the students' teams along two principal components, calculated according to the PCA approach.

identification of different problem-solving pathways and the analysis of how students learn to utilize sensors during an ER activity. The k-means algorithm identified 10 "pathways" that marked the students' teams' programming activity, during the resolution of specific exercise related to the ultrasonic sensor. Analyzing the pathways presented in the previous section, two main approaches to programming emerged: some teams modify the blocks' parameters implementing small changes, moving toward their objective by "small steps" (pathways 1, 3, 5); other teams design high changes (frequently modifying the symbol in the condition for the ultrasonic sensor, applying considerable variation to the threshold set for the ultrasonic sensor, repeatedly deleting/adding blocks, etc.) to their programming blocks from one test to another (pathways 2, 4, 6). The majority of the groups showing the first incremental approach (pathways 1, 3, 5) reached a working sequence during their first testing stage (an indicator of early achievement <0.4), unlike the teams with the "high changes approach" (pathways 2, 4, 6). This is a similar result compared to Blikstein et al. (2014), who identified that a "steadier incremental steps" strategy of programming correlated to a better performance in the resolution of the exercise. Pathway 4, with the highest number of trials (57) (**Table 2**), contains teams that did not obtain a working sequence in their first part of their work, and this result is similar to Chao (2016) but opposed to Filvà et al. (2019).

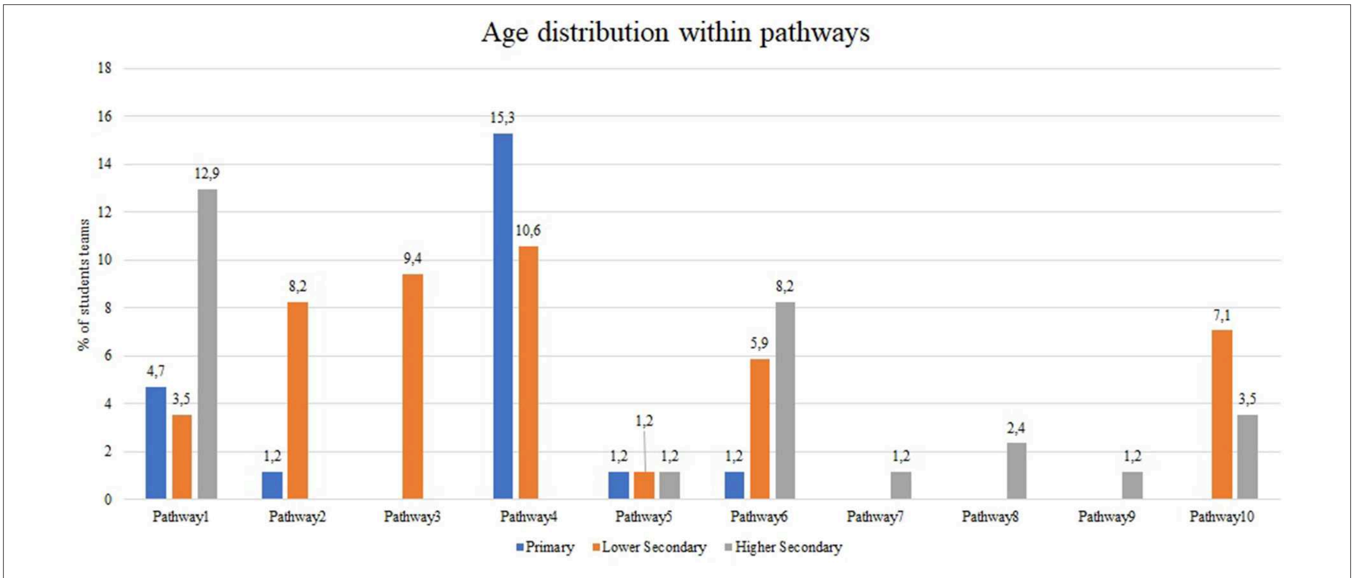


FIGURE 5 | The age-related differences between the students' teams involved in the experimentation, within the 10 problem-solving pathways.

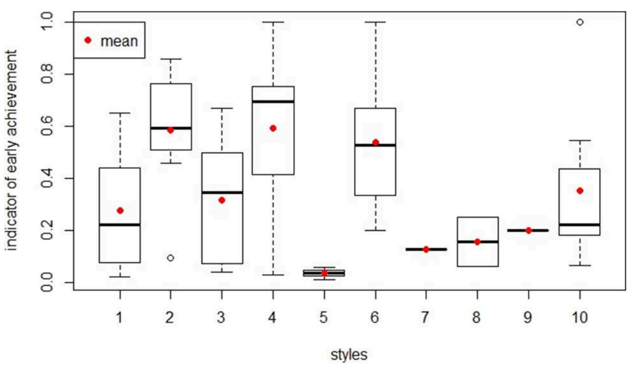


FIGURE 6 | Distributions of the indicator of early achievement in the 10 selected pathways.

Future work of this study includes the analysis of more extensive set of challenges in order to obtain more general results. The dataset considered in this paper is quite small (in particular, for pathways 7, 8, and 9): ER is an approach characterized by teamwork, so despite having involved 353 primary and secondary school students in the experimentation, we obtained valid data from 85 teams (participants were divided into teams of three to four members who worked together to design software solutions). The promising results of this preliminary study have encouraged the authors to involve new classes in the experimentation in order to continue the validation of the approach. The authors intend also to utilize a recurrent neural network, in particular, the long short-term memory autoencoders (a structure specifically designed to support sequences of input data Hochreiter and Schmidhuber, 1997), in order to translate the programming sequences created by the students into fixed-length vectors (compress representation of the input data), maintaining a high level of information content. As a result, these vectors obtained from the autoencoders compression will be used as input features for supervised and/or unsupervised algorithms. Another possible approach that the authors intend to use for the same task (dimensionality reduction) is the PCA.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent was obtained from the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

DS coordinated the research project and revised the paper. LC conceived the presented idea, designed the tracking system, implemented the machine learning analysis of the data, and wrote the paper, in collaboration with LS. LS developed the statistical analysis. EM supervised the machine learning analysis and revised the paper.

FUNDING

This material was based upon the work supported by the European Union, with the Erasmus+ KA1 Programme for students mobility and Erasmus+ KA201 ROBOPISCES (innovative educational ROBOtics strategies for PrImary SChool ExperienceS) project (2019-1-IT02-KA201-063073), by the Italian Ministry of Education (Ministero dell'Istruzione, dell'Università e della Ricerca-MIUR) with programme Fondi Strutturali Europei–Programma Operativo Nazionale Per la scuola, competenze e ambienti per l'apprendimento 2014–2020 Asse I–Istruzione–Fondo Sociale Europeo (FSE)–Obiettivo Specifico 10.1–Azione 10.1.1 Sottoazione 10.1.1A–PON Inclusione sociale e lotta al disagio, and by the Italian Department for Gender Equality (Dipartimento per le Pari Opportunità della Presidenza del Consiglio dei Ministri–DPO) with Programme In Estate si Imparano le STEM.

ACKNOWLEDGMENTS

The authors would like to show their gratitude to all the schools involved in the research project and to the educators of the Italian startup TALENT srl involved in the project (Luca Balestra, Federico Camilletti, Susanna Ortini, Roberto Faloppi, Michele Storti, and Elisa Mazzieri): their work was fundamental for the data collection during the ER courses.

REFERENCES

Ahmed, I., Lubold, N., and Walker, E. (2018). "ROBIN: using a programmable robot to provide feedback and encouragement on programming tasks," in *Artificial Intelligence in Education. AIED 2018. Lecture Notes in Computer Science*, Vol. 10948, eds C. P. Rosé, R. Martínez-Maldonado, H. Ulrich Hoppe, R. Luckin, M. Mavrikis, K. Porayska-Pomsta, B. McLaren, B. du Boulay (Cham: Springer), 9–13.

Allsop, Y. (2019). Assessing computational thinking process using a multiple evaluation approach. *Int. J. Child. Comput. Interact.* 19, 30–55. doi: 10.1016/j.ijcci.2018.10.004

Atmatzidou, S., and Demetriadis, S. (2016). Advancing students' computational thinking skills through educational robotics: a study on age and gender relevant differences. *Rob. Auton. Syst.* 75, 661–670. doi: 10.1016/j.robot.2015.10.008

Berland, M., Baker, R. S., and Blikstein, P. (2014). Educational data mining and learning analytics: applications to constructionist research. *Technol. Knowledge Learn.* 19, 205–220. doi: 10.1007/s10758-014-9223-7

Berland, M., Martin, T., Benton, T., Petrick Smith, C., and Davis, D. (2013). Using learning analytics to understand the learning pathways of novice programmers. *J. Learn. Sci.* 22, 564–599. doi: 10.1080/10508406.2013.836655

Berry, C. A., Remy, S. L., and Rogers, T. E. (2016). Robotics for all ages: a standard robotics curriculum for K-16. *IEEE Robot. Autom. Mag.* 23, 40–46. doi: 10.1109/MRA.2016.2534240

Bey, A., Pérez-Sanagustín, M., and Broisin, J. (2019). "Unsupervised automatic detection of learners' programming behavior," in *Transforming Learning With Meaningful Technologies. EC-TEL 2019. Lecture Notes in Computer Science*, Vol. 11722, eds M. J. Scheffel Broisin, V. Pammer-Schindler, A. Ioannou, and J. Schneider (Cham: Springer), 69–82.

Blikstein, P. (2011). "Using learning analytics to assess students' behavior in open-ended programming tasks," in *Proceedings of the 1st International Conference on Learning Analytics and Knowledge (ACM)* (Banff, AB), 110–116.

Blikstein, P., Worsley, M., Piech, C., Sahami, M., Cooper, S., and Koller, D. (2014). Programming pluralism: using learning analytics to detect patterns in the learning of computer programming. *J. Learn. Sci.* 23, 561–599. doi: 10.1080/10508406.2014.954750

Burbaite, R., Stuikys, V., and Damasevicius R. (2013). "Educational robots as collaborative learning objects for teaching computer science," in *proceedings of 2013 IEEE International Conference on System Science and Engineering (ICSSE)* (Budapest), 211–216.

Cesaretti, L., Storti, M., Mazzieri, E., Screpanti, L., Paesani, A., and Scaradozzi, D. (2017). An innovative approach to school-work turnover programme with educational robotics. *Mondo Digitale* 16, 2017–2015. Available online at: http://mondodigitale.aicanet.net/2017-5/best_papers_didamatica_2017/MD72_03_paper_64.pdf

Chalmers, C. (2018). Robotics and computational thinking in primary school. *Int. J. Child. Comput. Interact.* 17, 93–100. doi: 10.1016/j.ijcci.2018.06.005

Chao, P. Y. (2016). Exploring students' computational practice, design and performance of problem-solving through a visual programming environment. *Comput. Educ.* 95, 202–215. doi: 10.1016/j.compedu.2016.01.010

Damaševicius, R., Narbutaite, L., Plauska, I., and BlaŽauskas, T. (2017). Advances in the use of educational robots in project-based teaching. *TEM J.* 6:342. doi: 10.18421/TEM62-20

Daniela, L., and Lytras, M. D. (2019). Educational robotics for inclusive education. *Technol. Knowledge Learn.* 24, 219–225. doi: 10.1007/s10758-018-9397-5

Denis, B., and Hubert, S. (2001). Collaborative learning in an educational robotics environment. *Comput. Hum. Behav.* 17, 465–480. doi: 10.1016/S0747-5632(01)00018-8

Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics* 6, 241–252. doi: 10.1080/00401706.1964.10490181

Elkin, M., Sullivan, A., and Bers, M. U. (2014). Implementing a robotics curriculum in an early childhood Montessori classroom. *J. Inf. Technol. Educ. Innov. Prct.* 13, 153–169. doi: 10.28945/2094

Filvà, D. A., Forment, M. A., García-Peñalvo, F. J., Escudero, D. F., and Casañ, M. J. (2019). Clickstream for learning analytics to assess students' behavior with scratch. *Future Generation Comput. Syst.* 93, 673–686. doi: 10.1016/j.future.2018.10.057

Grover, S., and Pea, R. (2013). Computational thinking in K—12: a review of the state of the field. *Educ. Res.* 42, 38–43. doi: 10.3102/0013189X12463051

Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. *Neural Comput.* 9, 1735–1780. doi: 10.1162/neco.1997.9.8.1735

Jormanainen, I., and Sutinen, E. (2012). "Using data mining to support teacher's intervention in a robotics class," in *Proceedings of the 2012 IEEE Fourth International Conference on Digital Game and Intelligent Toy Enhanced Learning* (Washington, DC), 39–46.

Kandlhofer, M., and Steinbauer, G. (2016). Evaluating the impact of educational robotics on pupils' technical-and social-skills and science related attitudes. *Rob. Auton. Syst.* 75, 679–685. doi: 10.1016/j.robot.2015.09.007

Kim, C., Kim, D., Yuan, J., Hill, R. B., Doshi, P., and Thai, C. N. (2015). Robotics to promote elementary education pre-service teachers' STEM engagement, learning, and teaching. *Comput. Educ.* 91, 14–31. doi: 10.1016/j.compedu.2015.08.005

Kodinariya, T. M., and Makwana, P. R. (2013). Review on determining number of Cluster in K-means clustering. *Int. J. Adv. Res. Comput. Sci. Manage. Stud.* 1, 90–95. Available online at: https://pdfs.semanticscholar.org/948d/9d70bde0911606f621c620b8acf3d3d8cd42.pdf?_ga=2.55819371.1972312037.1583319245-1442875807.1569345736

Liu, E. Z. F. (2010). Early adolescents' perceptions of educational robots and learning of robotics. *B. J. Educ Technol.* 41, E44–E47. doi: 10.1111/j.1467-8535.2009.00944.x

Martinez, S. L., and Stager, G. (2013). *Invent to Learn: Making, Tinkering, and Engineering in the Classroom*. Torrance, CA: Constructing Modern Knowledge Press.

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., and Dong, J. J. (2013). A review of the applicability of robots in education. *J. Technol. Educ. Learn.* 1:13. doi: 10.2316/Journal.209.2013.1.209-0015

Papert, S. (1980). *Mindstorms: Children, Computers, and Powerful Ideas*. New York, NY: Basic Books.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65.

Savery, J. R. (2006). Overview of problem-based learning: definitions and distinctions. *Interdiscip. J. Probl. Based Learn.* 1, 9–20. doi: 10.1016/0377-0427(87)90125-7

Scaradozzi, D., Cesaretti, L., Screpanti, L., Costa, D., Zingaretti, S., and Valzano, M. (2019a). "Innovative tools for teaching marine robotics, IoT and control strategies since the primary school," in *Smart Learning With Educational Robotics*, ed L. Daniela (Cham: Springer), 199–227.

Scaradozzi, D., Screpanti, L., and Cesaretti, L. (2019b). "Towards a definition of educational robotics: a classification of tools, experiences and assessments," in *Smart Learning With Educational Robotics*, ed L. Daniela (Cham: Springer), 63–92.

Scaradozzi, D., Screpanti, L., Cesaretti, L., Storti, M., and Mazzieri, E. (2019c). Implementation and assessment methodologies of teachers' training courses for STEM activities. *Technol. Knowledge Learn.* 24, 247–268. doi: 10.1007/s10758-018-9356-1

Scaradozzi, D., Sorbi, L., Pedale, A., Valzano, M., and Vergine, C. (2015). Teaching robotics at the primary school: an innovative approach. *Procedia Soc. Behav. Sci.* 174, 3838–3846. doi: 10.1016/j.sbspro.2015.01.1122

Štuikys, V., Burbaite, R., and Damaševicius, R. (2013). Teaching of computer science topics using meta-programming-based GLOs and LEGO robots. *Inf. Educ.* 12, 125–142. Available online at: https://files.eric.ed.gov/fulltext/EJ1064312.pdf

Vega, J., and Cañas, J. M. (2019). PyBoKids: an innovative python-based educational framework using real and simulated arduino robots. *Electronics* 8:899. doi: 10.3390/electronics8080899

Wang, L., Sy, A., Liu, L., and Piech, C. (2017). "Learning to represent student knowledge on programming exercises using deep learning," in *Proceedings of the 10th International Conference on Educational Data Mining* (Wuhan), 324–329.

Zar, J. H. (2010). *Biostatistical Analysis*, 5th ed. Upper Saddle River, NJ: Pearson Prentice Hall.

Conflict of Interest: LC was employed by the company TALENT srl.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 31 March 2020
doi: 10.3389/frobt.2020.00037

Collaborative Research Project: Developing and Testing a Robot-Assisted Intervention for Children With Autism

Viviane Kostrubiec[1,2]* and *Jeanne Kruck*[1]

[1] Centre d'Etudes et de Recherches en Psychopathologie et Psychologie de la Santé (CERPPS), Université de Toulouse, UT2J, Toulouse, France, [2] Université de Toulouse, UT3, Toulouse, France

OPEN ACCESS

Edited by:
Agnese Augello,
Institute for High Performance Computing and Networking (ICAR), Italy

Reviewed by:
Erik A. Billing,
University of Skövde, Sweden
Wing Chee So,
The Chinese University of Hong Kong, China

****Correspondence:***
Viviane Kostrubiec
viviane.kostrubiec@univ-tlse3.fr

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *21 September 2019*
Accepted: *03 March 2020*
Published: *31 March 2020*

Citation:
Kostrubiec V and Kruck J (2020) Collaborative Research Project: Developing and Testing a Robot-Assisted Intervention for Children With Autism. Front. Robot. AI 7:37. doi: 10.3389/frobt.2020.00037

The present work is a collaborative research aimed at testing the effectiveness of the robot-assisted intervention administered in real clinical settings by real educators. Social robots dedicated to assisting persons with autism spectrum disorder (ASD) are rarely used in clinics. In a collaborative effort to bridge the gap between innovation in research and clinical practice, a team of engineers, clinicians and researchers working in the field of psychology developed and tested a robot-assisted educational intervention for children with low-functioning ASD ($N = 20$) A total of 14 lessons targeting requesting and turn-taking were elaborated, based on the Pivotal Training Method and principles of Applied Analysis of Behavior. Results showed that sensory rewards provided by the robot elicited more positive reactions than verbal praises from humans. The robot was of greatest benefit to children with a low level of disability. The educators were quite enthusiastic about children's progress in learning basic psychosocial skills from interactions with the robot. The robot nonetheless failed to act as a social mediator, as more prosocial behaviors were observed in the control condition, where instead of interacting with the robot children played with a ball. We discuss how to program robots to the distinct needs of individuals with ASD, how to harness robots' likability in order to enhance social skill learning, and how to arrive at a consensus about the standards of excellence that need to be met in interdisciplinary co-creation research. Our intuition is that robotic assistance, obviously judged as to be positive by educators, may contribute to the dissemination of innovative evidence-based practice for individuals with ASD.

Keywords: social robotics, social skills, evidence-based practices, robot acceptance, applied analysis of behavior

1. INTRODUCTION

There is a growing recognition of the innovation-to-practice gap arisen in social robotics (Fernaeus et al., 2010; Pennisi et al., 2016; Walters, 2018; Ismail et al., 2019), a field dedicated to developing robots to assist persons with special needs. To date, few social robots have gone beyond the prototype stage, or else are only deployed for research purposes (Wagenmakers, 2016). Their sale volume is still low (6,423 units in 2017), compared with that of domestic help robots (6.1 million units in 2017) (IFR, 2018). Kim et al. (2012) (see also Cabibihan et al., 2013; Pennisi et al., 2016) ascribed these difficulties to the lack of collaboration between researchers and end-users. Too long, research effort focused on the technological features of newly engineered robots (e.g., Kozima et al., 2007; Robins et al., 2009), not taking into account the specific needs of end-users. End-users do not evaluate a technical innovation, however outstanding it may be (Payne, 2015). Rather, they evaluate its added value relative to existing alternatives and its accordance to work routines (Joachim et al., 2018).

The hard earned lesson now is that to overcome the innovation-to-practice gap, close collaboration between engineers, researchers, caregivers and management team is needed. The collaboration may take the form of a participatory, pragmatic, or *collaborative approach*, where all the stakeholders work hand in hand to co-create tools best fitting the needs of end-users (Schwartz and Lellouch, 1967; March et al., 2005; Zwarenstein et al., 2008; Marchand et al., 2011; Forman et al., 2013; Bauer et al., 2015). In this emerging framework, having recently gained impetus from the paper by Balas and Boren (2000), researcher does not solely ask whether a new tool works when used in optimal laboratory conditions. Rather, he evaluates whether the tool works when used in real-life clinical settings, without highly-qualified staff, a homogenous group of patients, or tight experimental control (Cargo and Mercer, 2008; Zwarenstein et al., 2008; Brownson et al., 2012). The tool's acceptance is assessed by a questionnaire and implementation failures and context reported as a result on its own (Stahmer et al., 2015). We exploit here the collaborative approach to co-create and test socially assistive robot during an educational intervention dedicated to children with *autism spectrum disorder* (ASD).

1.1. Robots and ASD

ASD is an early-onset, pervasive developmental disorder that manifests itself in anomalies in social communication and interaction, together with abnormal restricted and/or repetitive patterns of behavior and interests (Lord et al., 1994; DSM 5, 2013). For instance, children with ASD avoid physical contact, do not orient toward humans, do not point to communicate, do not express enjoyment or interest, and may spend hours at lining up toys or flipping objects (Rutter et al., 2003). As ASD is incurable, some persons with this disorder require costly and intensive lifetime care, support and treatment, motivating the development of social robots to assist them and their caregivers.

The arising of social robots dedicated to ASD can be traced back to the seminal study by Emanuel and Weir (1976) (see also Howe, 1983), where a computer-controlled electrotechnical device, a turtle-like robot (LOGO) moving on wheels around the floor, was used as a remedial tool for an ASD boy. It was not until the late 1990s that multiple laboratories adopted this topic for research (see Werry and Dautenhahn, 1999; Diehl et al., 2012; Begum et al., 2016; Ismail et al., 2019; for reviews).

To date, nearly 30 robots were tested as remedial tools for ASD [e.g., : Labo-1 (Werry et al., 2001); Muu (Miyamoto et al., 2005), Robota (Billard et al., 2007), FACE (Pioggia et al., 2007), Keepon (Kozima et al., 2007), Aibo (Francois et al., 2009), IROMEC (Iacono et al., 2011), Charlie (Boccanfuso and O'Kane, 2011), NAO (Shamsuddin et al., 2012), Flobi (Damm et al., 2013); GIPY-1 (Giannopulu, 2013), Pleo (Kim et al., 2013), KASPAR (Wainer et al., 2014), Darwin-OP (Peng et al., 2014), Pabi (Dickstein-Fischer and Fischer, 2014), Zeno (Salvador et al., 2015), Jibo (Guizzo, 2015), Probo (Simut et al., 2016), Maria (Valadao et al., 2016), Sphero (Golestan et al., 2017), CARO (Yun et al., 2017), KiliRo (Bharatharaj et al., 2018), MINA (Ghorbandaei Pour et al., 2018), QTrobot (Costa et al., 2018), Milo (Chalmers, 2018), Leo (She et al., 2018), Daisy (Pliasa and Fachantidis, 2019), SAM (Lebersfeld et al., 2019), SPRITE (Clabaugh et al., 2019), Actroid-F (Yoshikawa et al., 2019) etc.].

The key hypothesis behind this endeavor states that social robots can maybe overcome some of the motivational and sensory barriers encountered by individuals with ASD when they interact with humans partners (Dautenhahn, 1999). In contrast to their typically developing peers, for whom social interactions are inherently rewarding, children with ASD exhibit only weak activation of the brain's reward system in response to social reinforcement (Chevallier et al., 2012; Delmonte et al., 2012; Watson et al., 2015). *Social Motivation Theory of ASD*, Chevallier et al. (2012) argued that ASD children neither seek out nor seek to maintain relations with human partners, showing instead a preference for nonhuman and often mechanic stimuli (Watson et al., 2015).

In addition to these motivational issues, sensory processing of persons with ASD is abnormal: they are often intolerant of complex multimodal stimuli (Bogdashina, 2010, 2012), display detail-focused perception (Happé and Frith, 2006), and sensory sensitivities or aversions (Bogdashina, 2010), with intense social anxiety (Spain et al., 2018). According to the *Weak Central Coherence theory* (Happé et al., 2001) and *Enhanced Perceptual Functioning model* (Mottron et al., 2006), the perceptual processing of ASD persons is biased toward local features: these children are incapable of integrating the variety of individual pieces of information into global patterns. *Intense World Theory of Autism* (Markram, 2007) sugested that these persons suffer from excessive neuronal information processing causing informational overload and abnormal levels of anxiety, which they seek to reduce with stereotypical and repetitive behaviors (Rodgers et al., 2012).

Given these characteristics of ASD, it seems useful to examine whether a social robot, with its motivational appeal, behavioral repetitiveness, simplified appearance and lack of social judgment, may be more appealing to individuals with ASD than real humans. Therefore, in line with Social Motivation Theory of ASD (Chevallier et al., 2012) our first working hypothesis (Hypothesis 1) is that children with ASD should positively react to sensory rewards delivered by a robot, by manifesting their interest and satisfaction when these stimuli are provided. In line with Intense World Theory of Autism (Markram, 2007), we also expect a reduction of anxiety-related undesirable behaviors (e.g., stereotypes, screams, auto-aggressions, etc.) in the presence of the robot (Hypothesis 2).

Yet, the key hope behind social robotics for ASD is that robots act as *social mediators*: they mediate, that is, promote or "catalyze" a cascade of so-called prosocial behaviors directed toward humans: eye or head orienting, physical contact, pointing to shared interest etc. (Dautenhahn, 2003; Feil Seifer and Mataric, 2009; Diehl et al., 2012). Our third working hypothesis (Hypothesis 3) is that in robot-assisted experimental conditions the child produces prosocial behaviors not only toward the robot but also toward humans. For the sake of clarity, a behavior is coined below as "prosocial" only in case it is dedicated to human, not to robot.

1.2. Building Up Robot Acceptance

In order to fulfill acceptance criteria of end-users, robot-assisted interventions should meet the efficiency standards of health services, tasked with assessing the level of experimental evidence supporting the added value of newly created tools (Burns et al., 2011), and providing recommendations to practitioners (GRADE Working Group, 2004). To accumulate such supportive evidence, multiple experiences should show that interventions for ASD work better when assisted by robots than in control condition, without the help of electromechanical devices.

To date, such evidence is scarce (Miguel Cruz et al., 2017). Of the 758 studies on robot-assisted interventions for ASD listed by Pennisi et al. (2016), only 29 (0.04%!) were selected as meeting clinical concerns. Publications still too often take the form of pilot studies (e.g., Werry et al., 2001; Miyamoto et al., 2005; Duquette et al., 2008; Robins et al., 2009; Costa et al., 2011; Dickstein-Fischer and Fischer, 2014) without control conditions, inferential statistics, diagnosis methods and inclusion/exclusion criteria (see Pennisi et al., 2016; Ismail et al., 2019 for critical reviews). Although necessary as a starting point, these preliminary studies are unable to establish the effectiveness of robotic tools in clinical samples, according to the rules of clinical methodology (Kazdin, 1998). The best-established effect is the "likability" of robots (Begum et al., 2016): children with ASD show enthusiasm for robotic devices and willingly participate in games assisted by these devices (Pliasa and Fachantidis, 2019)

To fit the needs of special needs educators, a collaborative approach was adopted. The idea of the robot in this project was born in 2011 in France when a father asked a team of young engineers from School of Industrial Biology at Cergy Pontoise to create games for his child with ASD. In 2014, a newly created French start-up created a low-cost, remotely controlled robot ball, that moves by rolling, vibrates and illuminates its transparent cover with different colors. Similar to spherical GIPY-1 (Giannopulu, 2013), Roball (Michaud et al., 2005), or SPRK+ Sphero (Golestan et al., 2017) the robot belongs thus to nonhumanoid devices.

The management team controlling the workflow enrolled the special educators and the children with ASD, and only then tasked researchers who could identify the educational goals and develop the procedure for the robot-assisted psychosocial skills training intervention. Children enrolled displayed low-functioning ASD, that is, intellectual quotient lower than 70 (i.e., intellectual dysfunction). Note however that the focus lies here on the effectiveness of the robot-assisted intervention, not on the specific functioning of these low-functioning children. At the end of our mission, we administered an acceptance questionnaire to analyze whether and how special educators accepted the robot-assisted intervention. We hoped that the intervention is judged as useful and fitting work routines (Hypothesis 4).

1.3. Intervention

We proposed an educational intervention targeting social skills and evaluated how efficient the robot is, as compared to the intervention without robotic help. In order to teach the social skills, we designed two sets of lessons to be taught using the *Applied Analysis of Behavior* (ABA) (Cooper et al., 2019) educational method recommended by health services. The key idea of ABA is to increase the probability of desirable behaviors by providing reinforcers in the form of rewards (Skinner, 1981). For the purpose of the present study, we chose the two general social skills that are most often targeted by educational interventions in ASD: requesting and turn-taking (Still et al., 2014; Huijnen et al., 2017). Requesting allows children to initiate a social interaction, express their needs and seek help, and leads to greater independence. Turn-taking is involved in the regulation of any social interaction. In order to exploit the added value of robots, compared with computer-mediated therapy, we administered tasks requiring body displacement in space, in particular during turn-taking lessons.

In line with ABA, the principles of the *Pivotal Training Method* (PTM) (Koegel et al., 1999, 2001) proposes that the learning of general skills (here: turn-taking and requesting) should bring about collateral improvements in a variety of nontrained prosocial behaviors in interpersonal interaction. In the present study, we thus focused on these expected collateral improvements, hoping that nontrained prosocial behaviors (here: orienting toward human, physical contact with human, pointing to communicate enjoyment and interest etc.) are more frequent in the robot-assisted than in control condition (viz. Hypothesis 3).

To sum up, the goal of these analyses was twofold. (1) First, we assessed the efficiency of the robot as a reward deliverer (viz. Hypothesis. 1), as an undesirable behavior reducer (viz. Hypothesis 2) and as social mediator (viz. Hypothesis 3). We expected that positive reactions to reward and nontrained prosocial behaviors are more frequent and that undesirable behaviors are less frequent in the robot, as compared to the control condition. (2) Second, we evaluated the acceptance of robot-assisted intervention by special educators (viz. Hypothesis 4). As in collaborative research interventions are administered by real caregivers, we anticipated that they could derail from the experimental procedure dictated by experimenters (viz. Hypothesis 5).

According to the suggestions of collaborative approach (Dingfelder and Mandell, 2011; Marchand et al., 2011), we conducted our study in two steps. After designing the first set of lessons devoted to requesting, we made successive modifications to the experimental protocol as problems emerged. Only then was the second turn-taking set of lessons administered and used for further analyses.

2. METHODS

2.1. Participants

The teamwork coordinator enrolled 20 children with ASD and 15 special educators in the study. They came from five special-needs schools and centers in France (APEAI Ouest Herault in Béziers, ADAPEI Papillons Blancs in Dunkirk, ADAPEI Papillons Blancs d'Alsace in Mulhouse, Ar'Roch in Rennes, DASCA Adéle de Glaubitz in Strasbourg, ADAPEI 44 in Nantes, APPARTE) where children receive care for their behavioral disorders.

As these centers correspond to small structures taking care of children with various mental disorders, only 1–2 individuals in each center fitted our inclusion/exclusion criteria: (1) 60–122 months of age at enrollment; (2) developmental age of 18–30 months assessed by verbal and preverbal cognition subtest from Psychoeducational Profile (PEP-3) (Schopler et al., 2004) (see below); (2) a diagnostic of ASD made by expert psychologists from Regional Autism Resources Center, and reconfirmed here by *Social communication Questionnaire* (SCQ, Rutter et al., 2003, see below); (3) no identifiable neurological disease or major neurological treatment. The ratio between developmental and maturational age was 0.28 ($SD = 0.09$), qualifying the children as low-functioning (i.e., severe intellectual deficit). Further psychological characteristics of our sample are provided in **Table 1**. The female-male ratio was 3/17.

As the robot had a low level of autonomy (Level 2; see Parasuraman et al., 2000), in each experimental session, in addition to the special educator interacting with the child, another person controlled the robot. Fifteen educators who cared for the children applied the experimental protocol: 56% were special needs monitors and 31% were special needs professionals, 67% had more than 10 years of experience, and 93% were women. At least one special educator in each center reported having already undergone a short ABA familiarization course. Just under half (47%) stated that they had never used new technologies, and just over half (53%) that they used them occasionally. The interventionist and the families of all the children received a letter explaining the goals, experimental procedure and rights of parents and children, and provided their written informed consent, in accordance with the Declaration of Helsinki. Each parent completed a form provided by the University of Toulouse informing them about their rights and predictable risks in comparison with foreseeable benefits. An ethics and scientific committee of the consulting company in the role of intermediary between the start-up, researchers and investors approved the experimental protocol; the committee members were also present during the first meeting. A declaration of ethical collection and storage of data was also made to the French Data Protection Authority (CNIL; ref.: 7e42415863j).

2.2. Material

2.2.1. Robot

We used a white, spherical prototype, measuring 18 cm in diameter and weighing 900 grams that was enclosed in a transparent plexiglass sphere resistant to shocks and pressure. Designed with a smiling face, equipped with actuators (LEDs, motors) and sensors (IMU 6-Axis, RFID), the prototype could light up or blink in different colors, and moved on two wheels in contact with the sphere. The robot was powered by AAA batteries and had autonomy of 3 to 4 h. Its behavior was remotely controlled by a touch pad (iPad iOS 10 or 11) with which it communicated through Bluetooth Low Energy over a distance of about 20 m.

In view of the intervention, three key functions were programmed in the robot. It acted as reward deliver, displaying colored lights and spinning movements, and also as cue provider: it offered specific lights and displacements prescribing required behavior of the child (e.g., "Touch the robot if it is your turn and if the robot is lit up in blue," see **Table 5**). Finally, it acted as lessons organizer, as explained below.

Two sets of seven lessons were developed. The application on the graphic tablet allowed the interventionist to consult the child's profile, which contained his/her experimental history and preferred sensory rewards, select a lesson, and display the lesson description and lesson control panel. The control panel featured various icons to launch the robot's cue, record the child's response, and provide rewards. Four types of responses from the child could be recorded: failure, success with total prompt, success with partial prompt, and success without prompt.

The control programs were developed on C++ for the robot and on Swift for the tablet. We were not allowed by investors to provide more technological details or the name of the device, never described in the literature and not commercialized to date.

In addition to the robot, a shoulder strap was provided to hold the graphic tablet. For the purpose of the experiment, a GoPro camera (Hero), a tripod (Fotopro), and a memory card (microSDHC SanDisk Extreme 32) were given to each center. A child's chair, and hoops were also required for the intervention. Because of the spherical design of the robot, balls were used in control condition.

2.2.2. Tutorials

Three tutorials were offered to the educators: (1) a brief introduction to ABA; (2) a technical description of the robot (see section 2.2.1), together with a detailed presentation of each set of lessons (see section 2.2.3); and (3) a description of the experimental design underlying the intervention (see section 2.3.2).

In the description of ABA, we recalled that in line with the principle of selection by consequences (Skinner, 1981), the educators would have to manage the sequence of events controlling each child's behavior (antecedent, behavior, consequence). According to the Discrete Trial Training method (Smith, 2001) learning should take the form of trials, each sequence involving an antecedent cue anticipating the appropriate behavior (e.g., "Touch the robot in turn"), a prompt wherein the educator assists the child (e.g., demonstration of required gesture, hand-over-hand assistance, pointing, nodding etc.), the child behavior (e.g., touching the robot in turn), the environmental consequence (e.g., verbal praise), and the intertrial interval (see **Figure 1**). We explained that providing a reward immediately after the to-be-learned target response reinforces the latter, increasing the probability of the target response being produced in the future.

Before each session, given their knowledge of the child's abilities and needs, the educators were asked to anticipate the required level of prompting, to avoid delays between the instruction and the prompt. They were told they should not hesitate to start with all prompts to facilitate learning. Prompts should be gradually faded out as learning proceeded, or increased in the case of a child failing (Leaf et al., 2016). We explained how instructions and rewards should be efficiently applied (e.g., brief,

TABLE 1 | Psychological tests used in the present experiment.

Test	No. items	Item scoring	Score interpretations	Child score
SCQ	40	0–1	score>15: possible ASD	21.43 (3.06)
V-listening	20	0–2	score <70: delay in receptive communication	16.81 (9.62)
V-speaking	32	0–2	Score <70: delay in expressive communication	22.15 (7.95)
V-autonomy	27	0–2	score <70: delay in personal autonomy	39.96 (15.46)
V-socialization	26	0–2	score <70: delay in socialization	22.08 (7.09)
V-adaptation	30	0–2	score <70: delay in social adaptation	9.56 (6.85)
PEP-3: AEs	11	0–2	Higher score: better affective expression	9.93 (4.43)
PEP3: SR	11	0–2	Higher score: better social reciprocity	11.14 (4.02)
PEP3: CVPV	34	0–2	Provides developmental age	26.44 (7.37)
SPCR	85	0–1	Higher score: more sensory abnormalities	26.86 (5.64)
ESES	13	1–9	Higher score: higher self-efficacy belief	85.86 (10.31)

For each test, the number of items, score range, score interpretation, mean score and standard deviation (SD) are provided for the children with ASD. SCQ, Social Communication Questionnaire; PEP, Psychoeducational Profile; SPCR, Sensory Profile Checklist Revised; AE, affective expressions; SR, social reciprocity; V, Vineland.

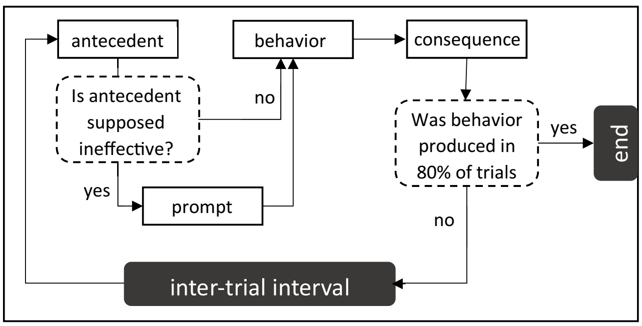


FIGURE 1 | ABA-based procedure of learning.

TABLE 2 | Required responses (R) for requesting set of lessons.

Set 1	Requesting
R1.	Look at the robot
R2.	Get closer to the robot
R3.	Touch the robot
R4.	Get closer to and touch the robot
R5.	Hold inactive robot to the adult
R6.	Hold inactive robot to the adult, who then activates it
R7.	Spontaneously hold inactive robot to the adult, who activates it

TABLE 3 | Required responses (R) for turn-taking set of lessons.

Set 2	Turn-taking
R1.	Touch the robot in turn
R2.	Touch the robot if it is your turn and if the robot is lit up in blue
R3.	Get closer to and touch the robot, in turn
R4.	If it is your turn and if the robot is lit up in blue, get closer and touch the robot
R5.	If it is your turn and if the robot is lit up in blue, imitate the adult who followed the robot along a short distance
R6.	Wait until the robot has reached the end of a short pathway and, if it is your turn and if the robot is lit up in blue, follow the path and touch the robot
R7.	Touch the robot to select the color controlling the turn-taking; wait until the robot has reached the end of a short pathway and, if the robot is lit up in blue, follow the path and touch the robot

clear, short and consistent instructions, provided when the target response was not being produced; reward applied immediately after the target response). Undesirable behaviors had to be gently and briefly interrupted, and the child immediately prompted to provide the target response (Cividini-Motta et al., 2019).

The descriptions of each lesson given to educators interacting with the child contained the learning goal (e.g., "Touch the robot in turn"), corresponding verbal instruction (e.g., "It's your turn"), required material (e.g., child's chair), preparation procedure (e.g., place the robot in the center of the room), a step-by-step procedure for learning, and a validation criterion (see below).

2.2.3. Sets of Lessons

Given that the volume of the tutorial depicting the lessons was 30 pages long, we provide below an abbreviate illustration of its content. Each set comprised a learning procedure that was ultimately aimed at enabling children to produce spontaneously and appropriately the general social skill targeted by the intervention: requesting (Set 1) and turn-taking (Set 2). Each set was composed of seven lessons, each with a learning goal, corresponding to a *required response* to be acquired by the child (e.g., "Look ate the robot," see **Table 2**, or "Touch the robot in turn", see **Table 3**). Required responses progressed from simple to complex, from prompted by the educator to initiated spontaneously by the child, from centered on the toy (robot or ball) to centered on the interaction with the educator (see **Tables 2**, **3** for the sequence of lessons in each set).

In each lesson, a step-by-step procedure described the elementary actions required from the robot (e.g., light up in blue), the interventionist (e.g., say "It's my turn"), and the child (e.g., "Touch the robot in turn"). Each lesson entailed five discrete learning trials (e.g., five turn-takings) where the interventionist

TABLE 4 | Step by step procedure for the first lesson in the turn-taking set.

Required response	Touch the robot in turn
Antecedent	1. Sit facing the child, and place the robot between you. The robot is inactive.
	2. Touch the robot on the top: it will light up in blue for a moment.
	3. Then encourage the child to do the same. Each time, say "It's my turn / It's your turn."
Behavior	4. If the child respects his/her turn, the robot will light up in blue.
	5. If not, the interventionist will blocks him/her, saying "No, it's my turn".
	6. If the child does not attempt to touch the robot, the educator selects a guidance specific to the child.
Consequence	7. After an errorless sequence of six turn-takings, the robot provides a sensory reward (specific to each child) and the interventionist gives verbal praise.
Validation Criterion	8. Repeat the sequence of turn-takings 5 times in a row (30 trials in all).
	9. Go to the next lesson if the child has produced a correct turn-taking sequence four times out of five.

TABLE 5 | Step by step procedure for the second lesson in the turn-taking set.

Required response	Touch the robot if it is your turn and if the robot is lit up in blue
Antecedent	1. Sit facing the child, and place the robot between you. The robot is active and lit up either in blue or red.
	2. If the robot's light is blue say "The robot is blue! Touch it!.
	3. If the robot's light is red say "The robot is red! Don't touch it!".
Behavior	4. If the robot's light is red and the child reaches to touch it, the educator will block the gesture, saying "The robot is red! Don't touch it!"
	5. If the robot's light is blue and the child does not attempt to touch it, the educator selects a guidance specific to the child (ex. The light is blue, you can touch it).
	6. If the robot lights up in blue and the child touches it, the robot light up in white for a moment.
Consequence	6. After an errorless sequence of six turn-takings, the robot provides a sensory reward (specific to each child) and the educator gives verbal praise.
Validation Criterion	7. Repeat the sequence of turn-takings 5 times in a row (30 trials in all)
	Go to the next lesson if the child has produced a correct turn-taking sequence four times out of five.

attempted to elicit the required response. In accordance with ABA criteria, a required response was deemed to be acquired if it was produced in 80% of these trials, without or with partial prompting (see **Figure 1**). If, after the five repetitions of the same trial, the child failed to meet this criterion, the educator stopped the whole experimental protocol. The step-by-step procedure for the first lesson in the turn-taking set appears in **Table 4**, for the second lesson in **Table 5**.

2.2.4. Workbooks

Information about the children and their caregivers was collected in two workbooks. The first workbook collected general information about the child (i.e., age, diagnostic tools used, developmental age) and provided five psychological tests for psychometric assessment: *Social Communication Questionnaire* (SCQ) (Rutter et al., 2003), *Vineland II* (Sparrow et al., 2012), *Psychoeducational Profile* (PEP-3) (Schopler et al., 2004), *Sensory Profile Checklist Revised* (SPCR) (Bogdashina, 2012), and *Educators' Sense of Efficacity Scale* (ESES), adapted from Teachers' Sense of Efficacity Scale (Tschannen-Moran and Hoy, 2001). These tools are described in **Appendix 1**; their key features and interpretation in **Table 1**. The second workbook included Educators' Sense of Efficacity Scale and the acceptance questionnaire.

2.2.5. Post-intervention Acceptance Questionnaire

To assess acceptance of the intervention, we developed a questionnaire for the educators targeting several issues: (1) for what kind of children is a robot-assisted intervention best suited? (2) what is its added value, advantages and disadvantages? (3) what is its effect on workload, educational intervention, and children's learning? and (4) what training is required to use the robot in educative intervention?

2.3. Procedure

2.3.1. Collaboration Procedure

In the present work, the stakeholders first met in order to discuss ethical issues, methodological requirements, and acceptance of the intervention by the children and educators. Two training meetings were organized for them. In the first training meeting, held before the start of experimentation, researchers described the experimental goals and procedure, simulated learning sessions, and described how to manage challenging behaviors. The second training meeting took place during the administration of the first set of lessons: the experimenters provided feedback to the educators, using videos of previous learning sessions. Half a day each week, a hotline was manned by JK to answer the educators' questions. The final meeting took place after the experimentation, in order to present the results and discuss the strengths and weaknesses of the robot-assisted intervention. Each family received a brief summary of their child's progress.

2.3.2. Intervention Procedure

After the educators had taken notice of ABA principles, of lessons content, and of the experimental design, described in the tutorials (see section 2.2.2), they completed the psychological tests from the first workbook (see **Table 1**). Then, the children underwent a familiarization session, where they were merely put in the presence of an inactive robot. The following week, the lessons started: requesting (see **Table 2**) followed by turn-taking (see **Table 3**), according to the step-by-step procedure as described in the tutorials (see **Tables 4**, **5**). Each child was administered each lesson in two conditions, in random order: with the robot and

FIGURE 2 | Intervention conditions: with robot **(left)** and with ball **(right)**.

with the ball (see **Figure 2**). At least one session with the robot and one with the ball was administered for each lesson. Each set of lessons was taught over 12 weeks. The entire intervention took place over 24 weeks. After the intervention, a second workbook was provided, including the ESES and acceptance questionnaire.

2.4. Data Reduction and Analysis

2.4.1. Observation Grid

After the end of interventions, the method of direct observation from videos was used (Hops et al., 1995). Video recordings of all the experimental sessions were analyzed by two trained coders (psychology undergraduates), who were familiar with ABA and blind to the purpose of the experiment. They used an observation grid listing 16 categories of responses (e.g., proximal pointing, head/gaze oriented toward human, stereotypies, see **Table 6**, right column), organized in four global classes: positive reactions to reward, prosocial behaviors, undesirable behaviors, and orientations (see **Table 6**, left column). To assess child autonomy, coders were required to record the prompts initiated by educators. To assess implementation quality, they were also asked to record the educators' implementation errors. Cohen's kappa was calculated to measure interrater agreement ($k = 0.92$).

2.4.2. Dependent Variables

All dependent variables were measured after the end of interventions. For each child and each experimental condition (robot, ball), we recorded the number of times each response category (e.g., proximal pointing) occurred, resulting in 16 summed scores (see **Table 6**, right column). These scores were then combined to four dependent variables corresponding to above-mentioned global classes (i.e., positive reactions to reward, prosocial behaviors, undesirable behaviors, and orientations, see **Table 6**, left column).

To take a deeper look into the effect of robot-assisted intervention, we computed the proportion of prosocial and undesirable behaviors produced in robot condition. The proportion was then normalized (from 1 to −1):

$$Normalized.Proportion = 2 \times \left(\frac{x_{robot}}{x_{robot} + x_{ball}}\right) - 1 \qquad (1)$$

TABLE 6 | Dependent variables and to-be-observed response catégories.

Dependent variables	Response categories	Label
Positive reactions to reward	To reward delivered by human	(PRH)
	To reward delivered by robot	(PRR)
Prosocial behaviors	Proximal pointing	(PP)
	Distal pointing	(DP)
	Joint gazing	(JG)
	Physical contact with human	(CH)
	Head/gaze oriented toward human	(OH)
	Social smiles	(SS)
	Desirable vocalizations	(DV)
Orientations	Head/gaze targeting human	(OTH)
	Head/gaze targeting toy: ball or robot	(OTT)
Undesirable behaviors	Inappropriate behaviors	(IA)
	Stereotypies	(S)
	Undesirable vocalizations	(UV)
	Lack of interest	(LI)
	Attentional dropout	(AD)

Each dependent variable in the left hand column is a combination of responses categories shown in the middle column. Left hand colums displays response category labels.

The normalized proportion takes a positive value when most of these behaviors were produced in robot condition, and inversely:

$$\begin{cases} 1 & if \quad x_{robot} > x_{ball} \\ 0 & if \quad x_{robot} = x_{ball} \\ -1 & if \quad x_{robot} < x_{ball} \end{cases} \qquad (2)$$

In the formula, x_{robot} and x_{ball} refer to the number of behaviors produced in robot and ball condition, respectively.

2.4.3. Statistical Analyses

To capture the characteristics of the children for whom the intervention was stopped and those who passed from lesson to lesson, one-tailed t-tests were carried out on all psychological test scores. Three groups were compared: the group who stopped the first set of lessons (i.e., Requesting), the group who started the second set (i.e., Turn-taking), and the group who completed the second lesson of the second set.

For further analysis, four experimental factors were envisioned: Condition (robot, ball), Reaction target (human, toy), Orientation Target (human, toy), and Prompt (with, without). Note, for Reaction target and Orientation target, the toy refers to robot in robot condition and to ball in ball condition.

To assess the efficacy of the robot-assisted intervention, we ran three statistical analyses. A 2 (Reaction Target = human in robot condition, human in ball condition, robot in robot condition) ANOVA was performed on positive reaction to reward and a 2 (Orientation Target = human, toy) × 2 (Condition = robot, ball) ANOVA was on orientations. A 2 (Condition = robot, ball) × 2 (Prompt = with, without) ANOVA was also run on prosocial behaviors and on undesirable behaviors to check whether the robot improved the children's social skills.

In all the ANOVAs, repeated measures were used on all dependent variables. Because each experimental factor (Condition, Reaction Target, Orientation Target and Prompt) had two levels, the assumptions of sphericity and of homogeneity of variances were always met. The distributions of dependent variables did not diverge from normal, as indicated by Lilliefors test for normality ($D = 0.1052$, $p = 0.5939$; $D = 0.0636$, $p = 0.99$; $D = 0.1145$, $p = 0.2722$; $D = 0.0443$, $p = 0.9901$, for reactions to reward, prosocial behaviors, orientations and undesirable behaviors, respectively).

If required, the ANOVAs were followed by appropriate two-tailed t-tests. The sign of normalized proportion was tested using one-sample t-test with 0 as comparison value. Finally, a matrix of correlation indices (r) was computed using all scores from the psychological tests and categories of responses. For all the above-mentioned analyses, the significance level was set at p <0.05, with the corresponding estimates of the effect size (η^2).

2.4.4. Statistical Analyses for Single Participant

Single-participant analyses were then performed on one of the children with ASD who successfully completed the whole intervention protocol. For this dataset, Bayesian statistics for single cases (de Vries and Morey, 2013; de Vries et al., 2015) were used. The posterior distribution for the standardized mean differences and Bayes factors were computed using the JZS+AR model with 10,000 Gibbs sampler iterations (de Vries et al., 2015). The Bayes factor quantifies evidence in the data for the null hypothesis against the alternative one: an inverse Bayes factor (1/BF) greater than 1 supports the alternative hypothesis. All 16 categories of responses, together with prosocial behaviors and undesirable behaviors, were submitted to this analysis.

2.4.5. Descriptive Statistics

To provide a glimpse into implementation fidelity, that is, the degree to which the educators strayed from the procedure specified by the experimenters, the coders were required to record any implementation error. The frequency of the failures was computed as a ratio of the number of failures to the number of videos. Finally, responses to the acceptance questionnaire were scored as percentages.

3. RESULTS

3.1. Child Sample Results

Children's Vineland-II and PEP-3 scores in our sample were low (see **Table 1**), indicating severe delays in social adaptive behavior, as well as in AE and SR skills. On average, sensory abnormalities were moderate. Of the 20 children with ASD who were initially enrolled, 15 reached the second set of lessons. The five participants who had to stop the first set had lower Vineland scores on listening, speaking and autonomy than the remaining participants, $t_{(18)} = 3.20$, p <0.007; $t_{(18)} = 3.04$, p <0.007; and $t_{(18)} = 2.29$, p <0.032. Of the 15 children who started the second set of lessons, only eight completed it. These eight children had higher Vineland listening scores than those who failed to complete the first and second sets of lessons, $t_{(13)} = 2.23$, p <0.044.

3.2. Robot-Assisted Intervention Results

3.2.1. Reward Deliver

A 2 (Reaction Target = human in robot condition, human in ball condition, robot in robot condition) ANOVA on positive reactions showed a main effect of Reaction Target, $F_{(2,14)} = 4.06$, $p = 0.04$, $\eta^2 = 0.546$. Corrected pairwise comparisons (see **Figure 3A**) showed that there was more positive reactions to the reward when it was delivered by the robot rather than by the human in robot and in ball conditions [$t_{(7)} = 2.37$, $p = 0.049$; $t_{(7)} = 2.50$ $p = 0.04$].

3.2.2. Undesirable Behavior Reducer

A 2 (Condition) ×2 (Prompt) ANOVA on undesirable behaviors revealed no statistically reliable effects.

3.2.3. Social Mediator

A 2 (Orientation Target) × 2 (Condition) ANOVA on orientation indicated an important main effect of Orientation Target, $F_{(1, 7)} = 23.538$, p <0.002, $\eta^2 = 0.771$. Children oriented more frequently toward the toy (i.e., ball or robot) than toward the educator. As illustrated in **Figure 3B**, there was also an Target Orientation × Condition interaction, $F_{(1, 7)} = 12.850$, p <0.009, $\eta^2 = 0.647$. When the children played with the robot, they oriented more often toward the robot than toward the educator, $t_{(7)} = 7.78$ p <0.0001. When they played with the ball, there was no effect of Orientation target, $t_{(7)} = 1.80$, $p = 0.1142$.

A 2 (Condition)× 2 (Prompt) on prosocial behaviors revealed a main effect for Prompt on prosocial behaviors only, $F_{(1, 7)} = 9.688$, p <0.017, $\eta^2 = 0.581$: Prosocial behaviors occurred more frequently with the prompt (20.06, $SD = 10.75$) than without it (9.50, $SD = 8.45$).

The value of the normalized proportion of prosocial behaviors was significantly negative, $t_{(6)} = 2.948$, $p = 0.026$: There were more prosocial behaviors in the ball rather than in the robot condition.

3.2.4. ASD Children Characteristics

There was a positive correlation between SCQ scores and orientations toward the ball condition ($r = 0.794$, $p = 0.033$), and a negative correlation between orientations toward the robot and auditory sensory abnormalities ($r = -0.907$, $p = 0.005$).

Further conclusions were drawn from the correlation between the normalized proportion of prosocial behaviors and SCQ score: the more severe the symptoms (i.e., the higher the SCQ value), the lower the proportion of prosocial behaviors produced in the robot as compared to the ball condition ($r = -0.813$, $p = 0.026$).

3.2.5. 6 Longitudinal Single-Participant Analysis

The child with ASD who completed all the sessions directed his gaze more often toward the robot than toward the ball (1/BF = 1.32 >1) (**Figure 4A**). He also produced more stereotypic behaviors in the robot than in the ball condition (1/BF = 2.82 >1) (**Figure 4B**).

3.3. Implementation Issues

Given that in collaborative/applied research, experimenters do not have total control of the implementation process

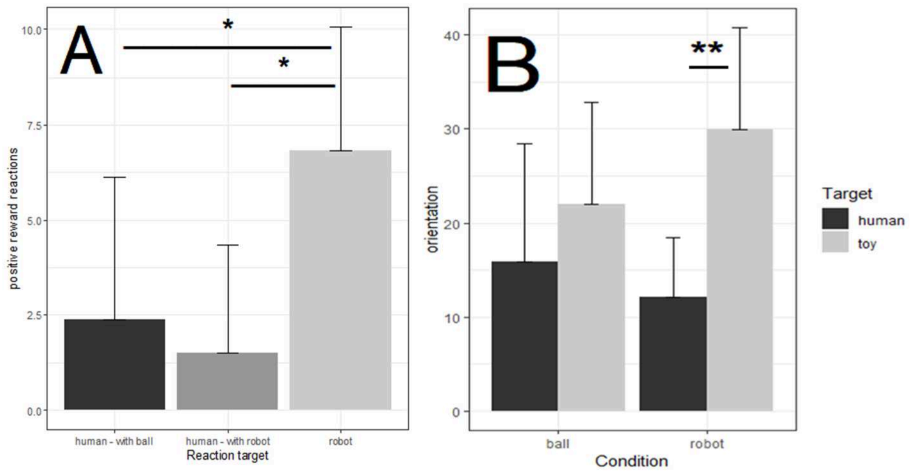


FIGURE 3 | Positive reactions **(A)** and orientations **(B)** as a function of Condition (ball, robot). *$p < 0.05$, **$p < 0.01$and ***$p < 0.001$.

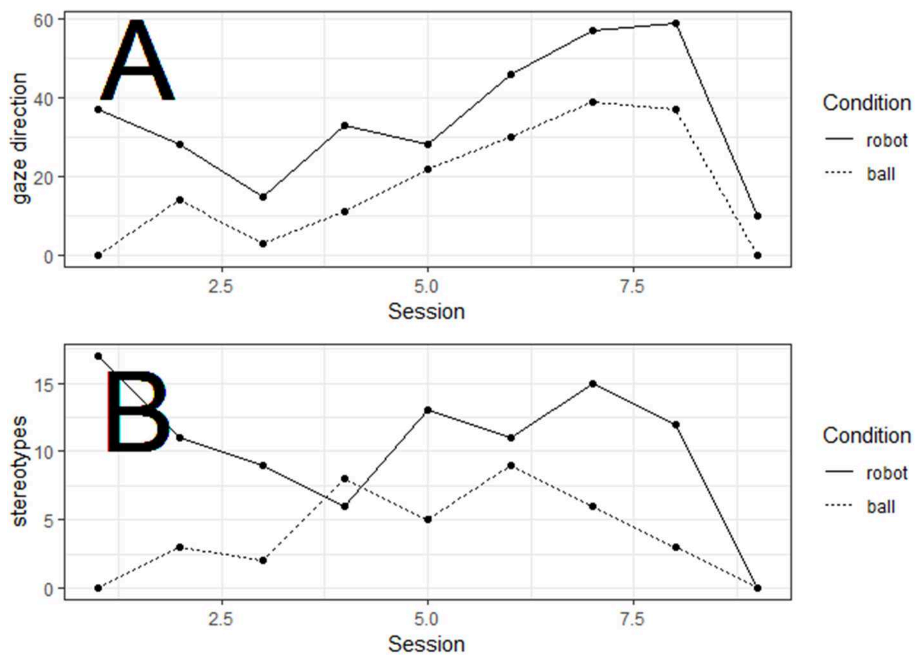


FIGURE 4 | Gaze direction **(A)** and stereotyped behaviors **(B)** as a function of Lessons and Condition (robot, ball).

and context of the experimental procedure, it is essential to describe the context delivery and the real-world difficulties encountered. This may prove to be particularly valuable in future efforts predicting, avoiding or better adapting to these socio-ecological constraints.

During the first set of lessons, the experimenters and coders identified five implementation failures where educators strayed from experimental requirements: instruction repeated too often or delivered at an inappropriate time; errors in action sequencing (i.e., instruction + prompt + interval, behavior + reward); reward omitted or delivered at an inappropriate time (e.g., before the child's behavior or after a failure); trial omission; and distractors not removed. In the set of 32 videos that were examined, 48 implementation failures were recorded, thus resulting in 1.5 failures per session. As indicated in **Table 7**, the most frequent failures were associated with reward or trial omission. However, the most severe procedural error was the omission of baseline conditions: before the intervention, the researchers had asked the educators to perform two baseline sessions: one with the robot and one with the ball, but some educators only carried out the baseline condition with the robot.

TABLE 7 | Implementation failures.

Nature of implementation failure	Failure frequency
Instruction	0.16
Action sequencing	0.13
Reward	0.69
Trial	0.47
Distractor	0.06

3.4. Robot Acceptance

3.4.1. Acceptance Questionnaire

The distributions of responses to the acceptance questionnaire showed that 87% of educators were satisfied or quite satisfied with their experience with the robot, 73% agreed that the robot brought substantial added value and transformed their practice, and 87% wanted to keep on using the robot in the future. Nevertheless, 67% of respondents confessed that they had been tempted to stop the intervention procedure. The reasons they gave included technical (80%) and organizational (33%) difficulties. The list of robot disadvantages also included substantial personal investment (60%) and increased workload (43%). In response to the questions assessing their training requirements, 40% of interventionists deemed that they need training in applying a structured educational approach.

3.4.2. Interventionists' Self-Efficacy Assessment

The interventionists' feeling of self-efficacy was initially high, and rose from 78.43 ($SD = 11.97$) before the intervention to 93.26 ($SD = 10.29$) after the intervention on a scale of 0–100, representing a significant increase, $t_{(6)} = -4.5962$, $p < 0.004$.

4. DISCUSSION

To better understand how to construct robotic tools for individuals with ASD, we conducted a collaborative study assessing the effects of a robot-assisted intervention on children with low-functioning ASD. Our intervention provided mixed results. As expected, children reacted more positive affect to rewards in robot as compared to control condition (viz. Hypothesis 1), and educators were quite enthusiastic about the robotic help in the learning task (viz. Hypothesis 4). However, contrary to our expectations, our robot was not able to act as a social mediator (viz. Hypothesis 3): when children played with the robot, they payed more attention to the toy than to the educator and the proportion of prosocial behaviors was higher in the control condition. Undesirable behaviors did not decrease (viz. Hypothesis 2). Of interest, the progression in the curriculum was IQ-specific: among the children we enrolled, those who displayed higher listening skills moved easily from lesson to lesson.

4.1. Reward Deliver

Children with ASD had more positive reactions to reward delivered by robot rather than to praises delivered by the eductor. This observation is analogous to enthusiastic reactions to robot reported in previous case studies (Dautenhahn, 1999, 2000; Kozima et al., 2007).

This enhanced reaction did not generalize to rewards delivered by the educator in robot condition though. The robot did not act as a general motivator (i.e., "motivating operation," Laraway et al., 2003; Edwards et al., 2019) enhancing the reinforcing effectiveness of any reward delivered in its presence. Rather, it acted as a preferred object: a strongly attractive object for children with ASD (DeLeon et al., 2001). In further studies, robots might be thus used to reinforce behaviors targeted by interventions, and compared to already exiting preferred toys.

4.2. Undesirable Behavior Reducer

Our robot had no consistant effect on undesirable behaviors: stereotypic behaviors even increased in one child. Ismail et al. (2012) suggested that robots may contribute to reduce the frequency of stereotypic behavior only for children with mild or no intellectual deficit. This demonstrates the need for psychometric descriptions of children in studies on robot-assisted interventions.

4.3. Social Mediator

The proportion of prosocial behaviors was higher in the control condition, rather than the robot-assisted intervention. We failed to offer support to social mediator hypothesis. Robins et al. (2005) warned that instead of social mediator, robots may sometimes take the role of social isolator. Meucci et al. (2019) suggested that the advantage of the interaction with a robot depends on the level of intellectual functioning of the children with ASD. In our data, we indeed noted that the more severe the ASD the lower the proportion of prosocial behaviors produced in the robot condition.

Note, extant information on social mediator hypothesis mostly comes from pilot studies or technical reports, without control condition, descriptive and inferential statistics (Werry et al., 2001; Robins et al., 2009; Iacono et al., 2011; Shamsuddin et al., 2012) and/or without diagnostic method, exclusion and inclusion criteria, developmental age etc. (Feil Seifer and Mataric, 2009; Valadao et al., 2016). Further studies could better comply with the requirements of clinical methodology.

Our intuition here is that using a highly attractive tool comes with the risk of turning the child with ASD away from the interpersonal social interaction skill, target of the training program. Our data indeed showed that children with ASD primarily gazed at the toy, seeing it as more attractive than the educator, in line with Social Motivation Theory of Autism (Chevallier et al., 2012; Delmonte et al., 2012). We suppose that robots would be more likely to "catalyze" prosocial behaviors if they interacted directly with the child, without any remote control, and if they endorsed a social role: that of prompter, teacher, helper in critical situations, etc. (Zubrycki and Granosik, 2016; Huijnen et al., 2017). Children with ASD would be therefore efficiently trained to produce and interpret social cues exchanged with the robot, and perhaps could generalize this learning to interpersonal interaction. In future research, robots of higher autonomy, similar to Jibo (Guizzo, 2015) or MINA (Ghorbandaei Pour et al., 2018) deserve particular attention.

4.4. Sensory Aversions and Inter-individual Heterogeneity Issue

Before the intervention, we feared that our robot, with its lighting signals and noisy functioning, might trigger anxiety among the children with ASD. The Intense World Theory of Autism (Markram, 2007) warned us indeed that children with ASD may be hypersensitive to these stimuli. This turned out to be a legitimate concern, as most of the auditory-sensitive children turned away from the robot.

This finding underscores the overlooked challenge faced by robots in the context of ASD: the inter-individual heterogeneity of children with ASD is shaping their reactions. This inter-individual heterogeneity makes it unlikely that a given robot or a given intervention will work for all children with ASD. In clinical settings, interventionists are used to adjust to each individual (Stahmer et al., 2011). They identify the sensory and cognitive particularities of each individual in order to decide which toy and which educational goal may be selected. They determine in real time how to attract the child's attention and modulate child anxiety, and which instructions, prompts, rewards and pauses should be administered. In further studies, robots should be endowed with an extensive set of educational goals and sensory options so that the administration of the educational procedure can be personalized. A first step toward this goal was recently made by Clabaugh et al. (2019) who developed a fully autonomous robot, SPRITE, able to personalize its instruction and feedback to each child's proficiency.

4.5. Collaboration Issues

One of the most often debated issues in the field of robotic assistance for children with ASD is infringement of the methodological rules of clinical research (Kim et al., 2012; Pennisi et al., 2016). This was an acute problem in our participatory study too. In the face of the understandable enthusiasm of the other stakeholders, it was difficult for the researchers to make their warnings heard. Nonexperimentalists have difficulty accepting that the violation of methodological rules inexorably means that some of the data that are collected are unusable.

Despite the obvious advantages of participatory research, it is important to acknowledge that this strategy creates huge problems in terms of coping with the priorities and constraints of different stakeholders, often working at cross purposes (Kim et al., 2012). Evidently, investors need to deliver a compelling marketable innovation capable of a sustainable commercial growth. Engineers want to promote innovative technological platforms that make existing ones obsolescent (Kim et al., 2012). Researchers are concerned with the originality and efficacy of the educational intervention, and thus need to respect to rigorous methodological criteria (Pennisi et al., 2016). The special need educators are interested in creating a user-friendly, personalizable tool that meets the specific needs of individual patients and fits in with current learning routines (Boardman et al., 2005). The company organizing the project has to factor in the time-limited and evanescent nature of the funding. There may be insufficient time and financial resources to organize meetings in order to build communication and trust between partners and work out a consensus on the standards of excellence to be met.

4.6. Implementation Fidelity Issue

As feared, the educators derailed from procedure dictated by research design (viz. Hypothesis 5). Despite workbooks, demonstrations and a hotline, educators made 1.5 implementation failures per session. In this respect, our intervention attempt was no different from others: Stahmer et al. (2015). showed that even after 28 h of intensive workshops, followed by 2 years of observation and coaching, the percentage of sessions meeting 80% implementation fidelity was just 60% for discrete trial teaching and as low as 20% for pivotal response training. Contrary to academic staff, special needs educators do not undergo years of training in administering trial-based, experiment-like procedures. Their skills imply intimate understanding of the child's difficulties and needs. Our intuition is that robots may play a non-negligible role here. If they can be designed to free educators from structuring the intervention according to the guidelines of educational protocol, they may contribute to the dissemination and application of structured educational approaches (e.g., ABA) recommended by health services.

4.7. Acceptance of the Robot-Assisted Intervention

The educators who took part in the present study were highly satisfied with their interaction with the robot. Coders noted that they seemed to take greater pleasure in interacting with the children. They had a greater feeling of self-efficiency after the experiment. Although we suspect that responses to the self-efficiency questionnaire were affected by a social desirability bias (Troye and Supphellen, 2012), leading the care staff to ignore undesirable traits such as self-doubt, it is quite possible that being supported by a robotic tool, instead of facing the child alone, engendered feelings of relief and satisfaction.

5. CONCLUSION

To better understand how to construct convincing tools for individuals with ASD, we conducted a collaborative study that assessed the effects of a robot-assisted intervention on both the prosocial and undesirable behaviors of children with low-functioning ASD. The robot attracted orienting responses from the children and the rewards it offered elicited more positive responses, but it failed to act as a social mediator: it did not motivate desired social behaviors toward humans. Robotic assistance was obviously judged to be positive by educators, thus contributing to the dissemination of evidence-based practices for individuals with ASD. In further studies, robots with higher levels of autonomy and differentiation, of richer set of educational goals and sensory response options might be tested as reinforcers of social behaviors targeted by educative intervention.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article can be found through Figshare (10.6084/m9.figshare.11994801).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Committee for Non-Invasive Procedures (CERNI) of the Université Fédérale Toulouse Midi-Pyrénées. The interventionist and the families of all the children received a letter explaining the goals, experimental procedure and rights of parents and children, and provided their written informed consent, in accordance with the Declaration of Helsinki. Each parent completed a form provided by the University of Toulouse informing them about their rights and predictable risks in comparison with foreseeable benefits. Written informed consent was obtained from the individual(s) or the individuals' legal guardian/next of kin for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

JK designed the study and supervised data collection. VK analyzed the results and wrote the manuscript. JK and VK critically reviewed and edited the manuscript for important intellectual content. All authors approved the final manuscript.

FUNDING

We thank AG2R La Mondiale for funding a preliminary study, described in a confidential activity report (Kruck and Kostrubiec, 2018) focused on the analysis of other distinct variables, and for the acknowledgment of the present work.

ACKNOWLEDGMENTS

The authors would like to thank children with ASD, their families, special need workers and engineers for their cooperation and help. We also thank Teresa Blicharska for her careful reading of the final version of the manuscript.

REFERENCES

Balas, E. A., and Boren, S. A. (2000). Managing clinical knowledge for health care improvement. *Yearb. Med. Informat*. 9, 65–70. doi: 10.1055/s-0038-1637943

Bauer, M. S., Damschroder, L., Hagedorn, H., Smith, J., and Kilbourne, A. M. (2015). An introduction to implementation science for the non-specialist. *BMC Psychol*. 3:32. doi: 10.1186/s40359-015-0089-9

Begum, M., Serna, R. W., and Yanco, H. A. (2016). Are robots ready to deliver autism interventions? A comprehensive review. *Int. J. Soc. Robot*. 8, 157–181. doi: 10.1007/s12369-016-0346-y

Bharatharaj, J., Huang, L., Krägeloh, C., Elara, M. R., and Al-Jumaily, A. (2018). Social engagement of children with autism spectrum disorder in interaction with a parrot-inspired therapeutic robot. *Proc. Comput. Sci*. 133, 368–376. doi: 10.1016/j.procs.2018.07.045

Billard, A., Robins, B., Nadel, J., and Dautenhahn, K. (2007). Building robota, a mini-humanoid robot for the rehabilitation of children with autism. *Assist. Technol*. 19, 37–49. doi: 10.1080/10400435.2007.10131864

Boardman, A. G., Argüelles, M. E., Vaughn, S., Hughes, M. T., and Klingner, J. (2005). Special education teachers' views of research-based practices. *J. Spec. Educ*. 39, 168–180. doi: 10.1177/00224669050390030401

Boccanfuso, L., and O'Kane, J. M. (2011). Charlie: an adaptive robot design with hand and face tracking for use in autism therapy. *Int. J. Soc. Robot*. 3, 337–347. doi: 10.1007/s12369-011-0110-2

Bogdashina, O. (2010). *Autism and the Edges of the Known World: Sensitivities, Language, and Constructed Reality*. London: Jessica Kingsley.

Bogdashina, O. (2012). *Questions Sensorielles et Perceptives Dans L'autisme et le Syndrome d'Asperger: Des Expériences Sensorielles Diffférentes, des Mondes Perceptifs Différents*. Grasse: AFD.

Brownson, R. C., Colditz, G. A., and Proctor, E. K. (2012). *Dissemination and Implementation Research in Health Translating Science to Practice*. Oxford: Oxford University Press. doi: 10.1093/acprof:oso/9780199751877.001.0001

Burns, P. B., Rohrich, R. J., and Chung, K. C. (2011). The levels of evidence and their role in evidence-based medicine: plastic and reconstructive surgery. *Plast Reconstr. Surg*. 128, 305–310. doi: 10.1097/PRS.0b013e318219c171

Cabibihan, J.-J., Javed, H., Ang, M., and Aljunied, S. M. (2013). Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. *Int. J. Soc. Robot*. 5, 593–618. doi: 10.1007/s12369-013-0202-2

Cargo, M., and Mercer, S. L. (2008). The value and challenges of participatory research: strengthening its practice. *Annu. Rev. Public Health* 29, 325–350. doi: 10.1146/annurev.publhealth.29.091307.083824

Chalmers, C. (2018). Robotics and computational thinking in primary school. *Int. J. Child-Comput. Interact*. 17, 93–100. doi: 10.1016/j.ijcci.2018.06.005

Chevallier, C., Kohls, G., Troiani, V., Brodkin, E. S., and Schultz, R. T. (2012). The social motivation theory of autism. *Trends Cogn. Sci*. 16, 231–239. doi: 10.1016/j.tics.2012.02.007

Cividini-Motta, C., Garcia, A. R., Livingston, C., and MacNaul, H. L. (2019). The effect of response interruption and redirection with and without a differential reinforcement of alternative behavior component on stereotypy and appropriate responses. *Behav. Intervent*. 34, 3–18. doi: 10.1002/bin.1654

Clabaugh, C., Mahajan, K., Jain, S., Pakkar, R., Becerra, D., Shi, Z., et al. (2019). Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders. *Front. Robot. AI* 6:110. doi: 10.3389/frobt.2019.00110

Cooper, J. O., Heron, T. E., and Heward, W. L. (2019). *Applied Behavior Analysis. 3rd Edn*. Hoboken, NJ: Pearson.

Costa, A. P., Charpiot, L., Lera, F. R., Ziafati, P., Nazarikhorram, A., Van Der Torre, L., et al. (2018). "More attention and less repetitive and stereotyped behaviors using a robot with children with autism," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (Nanjing: IEEE), 534–539. doi: 10.1109/ROMAN.2018.8525747

Costa, S., Soares, F., Santos, C., Ferreira, M. J., Moreira, F., Pereira, A. P., et al. (2011). "An approach to promote social and communication behaviors in children with autism spectrum disorders: robot based intervention," in *2011 RO-MAN* (Atlanta, GA: IEEE), 101–106. doi: 10.1109/ROMAN.2011.6005244

Damm, O., Malchus, K., Jaecks, P., Krach, S., Paulus, F., Naber, M., et al. (2013). "Different gaze behavior in human-robot interaction in Asperger's syndrome: an eye-tracking study," in *2013 IEEE RO-MAN* (Gyeongju: IEEE), 368–369. doi: 10.1109/ROMAN.2013.6628501

Dautenhahn (2000). "Issues of robot-human interaction dynamics in the rehabilitation of children with autism," in *From Animals To Animats, The Sixth International Conference on the Simulation of Adaptive Behavior (SAB2000)* (Paris).

Dautenhahn, K. (1999). "Robots as social actors: AURORA and the case of autism," in *Proceedings Third Cognitive Technology Conference CT99 August, 1999* (San Francisco, CA).

Dautenhahn, K. (2003). Roles and functions of robots in human society: implications from research in autism therapy. *Robotica* 21, 443–452. doi: 10.1017/S0263574703004922 de Vries, R. M., Hartogs, B. M. A., and Morey, R. D. (2015). A tutorial on computing bayes factors for single-subject designs. *Behav. Ther*. 46, 809–823. doi: 10.1016/j.beth.2014.09.013

de Vries, R. M., and Morey, R. D. (2013). Bayesian hypothesis testing for single-subject designs. *Psychol. Methods* 18, 165–185. doi: 10.1037/a0031037

DeLeon, I. G., Fisher, W. W., Rodriguez-Catter, V., Maglieri, K., Herman, K., and Marhefka, J. M. (2001). Examination of relative reinforcement effects of stimuli identified through pretreatment and daily brief preference assessments. *J. Appl. Behav. Anal*. 34, 463–473. doi: 10.1901/jaba.2001.34-463

Delmonte, S., Balsters, J. H., McGrath, J., Fitzgerald, J., Brennan, S., Fagan, A. J., et al. (2012). Social and monetary reward processing in autism spectrum disorders. *Mol. Autism* 3:7. doi: 10.1186/2040-2392-3-7

Dickstein-Fischer, L., and Fischer, G. S. (2014). "Combining psychological and engineering approaches to utilizing social robots with children with Autism," in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Chicago, IL: IEEE), 792–795. doi: 10.1109/EMBC.2014.6943710

Diehl, J. J., Schmitt, L. M., Villano, M., and Crowell, C. R. (2012). The clinical use of robots for individuals with Autism Spectrum Disorders: a critical review. *Res. Autism Spectr. Disord*. 6, 249–262. doi: 10.1016/j.rasd.2011.05.006

Dingfelder, H. E., and Mandell, D. S. (2011). Bridging the research-to-practice gap in autism intervention: an application of diffusion of innovation theory. *J. Autism Dev. Disord*. 41, 597–609. doi: 10.1007/s10803-010-1081-0

DSM 5 (2013). *American Psychiatric Association Diagnostic and Statistical Manual of Mental Disorders, in Springer Reference*. Berlin; Heidelberg: Springer-Verlag.

Duquette, A., Michaud, F., and Mercier, H. (2008). Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. *Auton. Robot*. 24, 147–157. doi: 10.1007/s10514-007-9056-5

Edwards, T. L., Lotfizadeh, A. D., and Poling, A. (2019). Motivating operations and stimulus control. *J. Exp. Anal. Behav*. 112, 1–9. doi: 10.1002/jeab.516

Emanuel, R., and Weir, S. (1976). "Using LOGO to catalyse communication in an autistic child," in *AISB'76 Proceedings of the 2nd Summer Conference on Artificial Intelligence and Simulation of Behaviour* (Edinburgh), 118–129.

Feil Seifer, D., and Mataric, M. J. (2009). "Toward socially assistive robotics for augmenting interventions for children with autism spectrum disorders," in *Experimental Robotics*, eds O. Khatib, V. Kumar, and G. J. Pappas (Berlin; Heidelberg: Springer Berlin Heidelberg), 201–210. doi: 10.1007/978-3-642-00196-3_24

Fernaeus, Y., Håkansson, M., Jacobsson, M., and Ljungblad, S. (2010). "How do you play with a robotic toy animal?: a long-term study of Pleo," in *Proceedings of the 9th International Conference on Interaction Design and Children - IDC'10* (Barcelona: ACM Press), 39. doi: 10.1145/1810543.1810549

Forman, S. G., Shapiro, E. S., Codding, R. S., Gonzales, J. E., Reddy, L. A., Rosenfield, S. A., et al. (2013). Implementation science and school psychology. *School Psychol. Q*. 28, 77–100. doi: 10.1037/spq0000019

Francois, D., Stuart, P., and Dautenhahn, K. (2009). A long-term study of children with autism playing with a robotic pet: taking inspirations from non-directive play therapy to encourage children's proactivity and initiative-taking. *IS* 10, 324–373. doi: 10.1075/is.10.3.04fra

Ghorbandaei Pour, A., Taheri, A., Alemi, M., and Meghdari, A. (2018). Human-robot facial expression reciprocal interaction platform: case studies on children with autism. *Int. J. Soc. Robot*. 10, 179–198. doi: 10.1007/s12369-017-0461-4

Giannopulu, I. (2013). Multimodal cognitive nonverbal and verbal interactions: the neurorehabilitation of autistic children via mobile toy robots. *Int. J. Adv. Life Sci*. 5, 214–222. doi: 10.1109/HRI.2013.6483534

Golestan, S., Soleiman, P., and Moradi, H. (2017). "Feasibility of using sphero in rehabilitation of children with autism in social and communication skills," in *2017 International Conference on Rehabilitation Robotics (ICORR)* (London: IEEE), 989–994. doi: 10.1109/ICORR.2017.8009378

GRADE Working Group (2004). Grading quality of evidence and strength of recommendations. *BMJ* 328:1490. doi: 10.1136/bmj.328.7454.1490

Guizzo, E. (2015). Jibo is as good as social robots get. but is that good enough? *Sci. Robot*. 3:21. doi: 10.1126/scirobotics.aat7544

Happé, F., and Frith, U. (2006). The weak coherence account: detail-focused cognitive style in autism spectrum disorders. *J. Autism Dev. Disord*. 36, 5–25. doi: 10.1007/s10803-005-0039-0

Happé, F., Frith, U., and Briskman, J. (2001). Exploring the cognitive phenotype of autism: weak "central coherence" in parents and siblings of children with autism: I. Experimental tests. *J. Child Psychol. Psychiatry* 42, 299–307. doi: 10.1111/1469-7610.00723

Hops, H., Davis, B., and Longoria, N. (1995). Methodological issues in direct observation: illustrations with the living in familial environments (LIFE) coding system. *J. Clin. Child Psychol*. 24, 193–203. doi: 10.1207/s15374424jccp2402_7

Howe, J. (1983). "Autism-Using a 'turtle' to establish communication," in *High Technology Aids for the Disabled*, ed W. J. Perkins (Amsterdam: Elsevier), 179–183. doi: 10.1016/B978-0-407-00256-2.50033-2

Huijnen, C. A., Lexis, A. S. M., and de Witte, L. P. (2017). Robots as new tools in therapy and education for children with autism. *Int. J. Neurorehabilitation* 4:278. doi: 10.4172/2376-0281.1000278

Iacono, I., Lehmann, H., Marti, P., Robins, B., and Dautenhahn, K. (2011). "Robots as social mediators for children with Autism - A preliminary analysis comparing two different robotic platforms," in *2011 IEEE International Conference on Development and Learning (ICDL)* (Frankfurt am Main: IEEE), 1–6. doi: 10.1109/DEVLRN.2011.6037322

IFR (2018). *IFR Executive Summary World Robotics 2018 Service Robots*. International Federation for Robotics (IFR). Available online at: https://ifr.org/free-downloads/ (accessed September 21, 2018).

Ismail, L. I., Shamsudin, S., Yussof, H., Hanapiah, F. A., and Zahari, N. I. (2012). Robot-based intervention program for autistic children with humanoid robot NAO: initial response in stereotyped behavior. *Proc. Eng*. 41, 1441–1447. doi: 10.1016/j.proeng.2012.07.333

Ismail, L. I., Verhoeven, T., Dambre, J., and Wyffels, F. (2019). Leveraging robotics research for children with autism: a review. *Int. J. Soc. Robot*. 11, 389–410. doi: 10.1007/s12369-018-0508-1

Joachim, V., Spieth, P., and Heidenreich, S. (2018). Active innovation resistance: an empirical study on functional and psychological barriers to innovation adoption in different contexts. *Indust. Market. Manage*. 71, 95–107. doi: 10.1016/j.indmarman.2017.12.011

Kazdin, A. E. (1998). *Research Design in Clinical Psychology. 3rd Edn*. Boston, MA: Allyn and Bacon.

Kim, E., Paul, R., Shic, F., and Scassellati, B. (2012). Bridging the research gap: making HRI useful to individuals with autism. *J. Hum. Robot Interact*. 1, 26–54. doi: 10.5898/JHRI.1.1.Kim

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., et al. (2013). Social robots as embedded reinforcers of social behavior in children with autism. *J. Autism Dev. Disord*. 43, 1038–1049. doi: 10.1007/s10803-012-1645-2

Koegel, L. K., Koegel, R. L., Shoshan, Y., and McNerney, E. (1999). Pivotal response intervention II: preliminary long-term outcome data. *Res. Pract. Pers. Severe Disabil*. 24, 186–198. doi: 10.2511/rpsd.24.3.186

Koegel, R. L., Koegel, L. K., and McNerney, E. K. (2001). Pivotal areas in intervention for autism. *J. Clin. Child Adolesc. Psychol*. 30, 19–32. doi: 10.1207/S15374424JCCP3001_4

Kozima, H., Nakagawa, C., and Yasuda, Y. (2007). "Children-robot interaction: a pilot study in autism therapy," in *Progress in Brain Research, Vol. 164*, 385–400. doi: 10.1016/S0079-6123(07)64021-7

Kruck, J., and Kostrubiec, V. (2018). *Experimentation Scientifique: Rapport Final*. Paris: AG2R La Mondiale.

Laraway, S., Snycerski, S., Michael, J., and Poling, A. (2003). Motivating operations and terms to describe them: some further refinements. *J. Appl. Behav. Anal*. 36, 407–414. doi: 10.1901/jaba.2003.36-407

Leaf, J. B., Leaf, J. A., Alcalay, A., Kassardjian, A., Tsuji, K., Dale, S., et al. (2016). Comparison of most-to-least prompting to flexible prompt fading for children with autism spectrum disorder. *Exceptionality* 24, 109–122. doi: 10.1080/09362835.2015.1064419

Lebersfeld, J. B., Brasher, C., Biasini, F., Hopkins, M. (2019). Characteristics associated with improvement following the SAM robot intervention for children with autism spectrum disorder. *Int. J. Pediatr. Neonat. Care*. 5:9. doi: 10.15344/2455-2364/2019/151

Lord, C., Rutter, M., and Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J. Autism Dev. Disord*. 24, 659–685. doi: 10.1007/BF02172145

March, J. S., Silva, S. G., Compton, S., Shapiro, M., Califf, R., and Krishnan, R. (2005). The case for practical clinical trials in psychiatry. *AJP* 162, 836–846. doi: 10.1176/appi.ajp.162.5.836

Marchand, E., Stice, E., Rohde, P., and Becker, C. B. (2011). Moving from efficacy to effectiveness trials in prevention research. *Behav. Res. Therapy* 49, 32–41. doi: 10.1016/j.brat.2010.10.008

Markram, H. (2007). The intense world syndrome-an alternative hypothesis for autism. *Front. Neurosci.* 1, 77–96. doi: 10.3389/neuro.01.1.1.006.2007

Meucci, P., Zampini, L., Giovannetti, A. M., Quadraroli, A., D'Arrigo, A. V., Riva, D., et al. (2019). The challenge of studying interaction in children with autism spectrum disorder during play activity with a robotic platform. *J. Dev. Phys. Disabil.* 32, 113–129. doi: 10.1007/s10882-019-09687-z

Michaud, F., Laplante, J.-F., Larouche, H., Duquette, A., Caron, S., Letourneau, D., et al. (2005). Autonomous spherical mobile robot for child-development studies. *IEEE Trans. Syst. Man Cybern. A* 35, 471–480. doi: 10.1109/TSMCA.2005.850596

Miguel Cruz, A., Ríos Rincón, A. M., Rodríguez Dueñas, W. R., Quiroga Torres, D. A., and Bohórquez-Heredia, A. F. (2017). What does the literature say about using robots on children with disabilities? *Disabil. Rehabil. Assist. Technol.* 12, 429–440. doi: 10.1080/17483107.2017.1318308

Miyamoto, E., Lee, M., Fujii, H., and Okada, M. (2005). "How can robots facilitate social interaction of children with autism? Possible implications for educational environments," in *Proceedings of the 5th International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems* (Lund), 145–146.

Mottron, L., Dawson, M., Soulières, I., Hubert, B., and Burack, J. (2006). Enhanced perceptual functioning in autism: an update, and eight principles of autistic perception. *J. Autism Dev. Disord.* 36, 27–43. doi: 10.1007/s10803-005-0040-7

Parasuraman, R., Sheridan, T. B., and Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Trans. Syst. Man Cybern. A Syst. Hum.* 30, 286–297. doi: 10.1109/3468.844354

Payne, M. (2015). *How to Kill a Unicorn and Build the Bold Ideas That Make It to Market, Drive Growth and Transform Industries*. London: Nicholas Brealey Publishing.

Peng, Y.-H., Lin, C.-W., Mayer, N. M., and Wang, M.-L. (2014). "Using a humanoid robot for music therapy with autistic children," in *2014 CACS International Automatic Control Conference (CACS 2014)* (Kaohsiung: IEEE), 156–160. doi: 10.1109/CACS.2014.7097180

Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S., et al. (2016). Autism and social robotics: a systematic review: autism and social robotics. *Autism Res.* 9, 165–183. doi: 10.1002/aur.1527

Pioggia, G., Sica, M. L., Ferro, M., Igliozzi, R., Muratori, F., Ahluwalia, A., et al. (2007). "Human-robot interaction in autism: FACE, an android-based social therapy," in *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication* (Jeju: IEEE), 605–612. doi: 10.1109/ROMAN.2007.4415156

Pliasa, S., and Fachantidis, N. (2019). "Can a robot be an efficient mediator in promoting dyadic activities among children with Autism Spectrum Disorders and children of Typical Development?," in *Proceedings of the 9th Balkan Conference on Informatics - BCI'19* (Sofia: ACM Press), 1–6. doi: 10.1145/3351556.3351592

Robins, B., Dautenhahn, K., and Dickerson, P. (2009). "From isolation to communication: a case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot," in *2009 Second International Conferences on Advances in Computer-Human Interactions* (Cancun: IEEE), 205–211. doi: 10.1109/ACHI.2009.32

Robins, B., Dautenhahn, K., and Dubowski, J. (2005). "Robots as Isolators or Mediators for Children with Autism A Cautionary Tale," in *Procs of the AISB 05 Symposium on Robot Companions?: Hard Problems and Open Challenges in Robot-Human Interaction, AISB*. Available online at: http://hdl.handle.net/2299/6759 (accessed December 15, 2018).

Rodgers, J., Glod, M., Connolly, B., and McConachie, H. (2012). The relationship between anxiety and repetitive behaviours in autism spectrum disorder. *J. Autism Dev. Disord.* 42, 2404–2409. doi: 10.1007/s10803-012-1531-y

Rutter, M., Anthony, B., and Lord, C. (2003). *The Social Communication Questionnaire Manual*. Los Angeles, CA: Western Psychological Services.

Salvador, M. J., Silver, S., and Mahoor, M. H. (2015). "An emotion recognition comparative study of autistic and typically-developing children using the zeno robot," in *2015 IEEE International Conference on Robotics and Automation (ICRA)* (Seattle, WA: IEEE), 6128–6133. doi: 10.1109/ICRA.2015.7140059

Schopler, E., Lansing, M. D., Reichler, R. J., and Marcus, L. M. (2004). *Psychoeducational Profile Third Edition (PEP-3)*. Austin, TX: Pro-Ed.

Schwartz, D., and Lellouch, J. (1967). Explanatory and pragmatic attitudes in therapeutical trials. *J. Chronic Dis.* 20, 637–648. doi: 10.1016/0021-9681(67)90041-0

Shamsuddin, S., Yussof, H., Ismail, L. I., Mohamed, S., Hanapiah, F. A., and Zahari, N. I. (2012). Initial response in HRI- a case study on evaluation of child with autism spectrum disorders interacting with a humanoid robot NAO. *Proc. Eng.* 41, 1448–1455. doi: 10.1016/j.proeng.2012.07.334

She, T., Kang, X., Nishide, S., and Ren, F. (2018). "Improving LEO robot conversational ability via deep learning algorithms for children with autism," in *2018 5th IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)* (Nanjing: IEEE), 416–420. doi: 10.1109/CCIS.2018.8691393

Simut, R. E., Vanderfaeillie, J., Peca, A., Van de Perre, G., and Vanderborght, B. (2016). Children with autism spectrum disorders make a fruit salad with probo, the social robot: an interaction study. *J. Autism Dev. Disord.* 46, 113–126. doi: 10.1007/s10803-015-2556-9

Skinner, B. F. (1981). Selection by consequences. *Science* 213, 501–504. doi: 10.1126/science.7244649

Smith, T. (2001). Discrete trial training in the treatment of autism. *Focus Autism Other Dev. Disabl.* 16, 86–92. doi: 10.1177/108835760101600204

Spain, D., Sin, J., Linder, K. B., McMahon, J., and Happé, F. (2018). Social anxiety in autism spectrum disorder: a systematic review. *Res. Autism Spectr. Disord.* 52, 51–68. doi: 10.1016/j.rasd.2018.04.007

Sparrow, S. S., Cicchetti, D., and Balla, D. A. (2012). *Vineland Adaptive Behavior Scales, 2nd Edn*. Minneapolis, MN: Pearson Assessments. doi: 10.1007/978-0-387-79948-3_1602

Stahmer, A. C., Rieth, S., Lee, E., Reisinger, E. M., Mandell, D. S., and Connell, J. E. (2015). Training teachers to use evidence-based practices for autism: examining procedural implementation fidelity: training teachers in autism practices. *Psychol. Schs.* 52, 181–195. doi: 10.1002/pits.21815

Stahmer, A. C., Schreibman, L., and Cunningham, A. B. (2011). Toward a technology of treatment individualization for young children with autism spectrum disorders. *Brain Res.* 1380, 229–239. doi: 10.1016/j.brainres.2010.09.043

Still, K., Rehfeldt, R. A., Whelan, R., May, R., and Dymond, S. (2014). Facilitating requesting skills using high-tech augmentative and alternative communication devices with individuals with autism spectrum disorders: a systematic review. *Res. Autism Spectr. Disord.* 8, 1184–1199. doi: 10.1016/j.rasd.2014.06.003

Troye, S. V., and Supphellen, M. (2012). Consumer participation in coproduction: "I Made it Myself" effects on consumers' sensory perceptions and evaluations of outcome and input product. *J. Market.* 76, 33–46. doi: 10.1509/jm.10.0205

Tschannen-Moran, M., and Hoy, A. W. (2001). Teacher efficacy: capturing an elusive construct. *Teach. Teacher Educ.* 17, 783–805. doi: 10.1016/S0742-051X(01)00036-1

Valadao, C. T., Goulart, C., Rivera, H., Caldeira, E., Bastos Filho, T. F., Frizera-Neto, A., et al. (2016). Analysis of the use of a robot to improve social skills in children with autism spectrum disorder. *Res. Biomed. Eng.* 32, 161–175. doi: 10.1590/2446-4740.01316

Wagenmakers, R. (2016). *Social Robots*. Available online at: https://assets.kpmg.com/content/dam/kpmg/pdf/2016/06/social-robots.pdf (accessed May 23, 2019).

Wainer, J., Dautenhahn, K., Robins, B., and Amirabdollahian, F. (2014). A pilot study with a novel setup for collaborative play of the humanoid robot KASPAR with children with autism. *Int. J. Soc. Robot.* 6, 45–65. doi: 10.1007/s12369-013-0195-x

Walters, R. (2018). *Dead robots raise questions on how far home technology has come Rogers*. Financial Times. Available online at: https://www.ft.com/content/e92c5b9a-9c59-11e8-9702-5946bae86e6d (accessed: February 10, 2020).

Watson, K. K., Miller, S., Hannah, E., Kovac, M., Damiano, C. R., Sabatino-DiCrisco, A., et al. (2015). Increased reward value of non-social stimuli in children and adolescents with autism. *Front. Psychol.* 6:1026. doi: 10.3389/fpsyg.2015.01026

Werry, I., Dautenhahn, K., Ogden, B., and Harwin, W. (2001). "Can social interaction skills be taught by a social agent? The role of a robotic mediator in autism therapy," in *CT'01 Proceedings of the 4th International Conference* *on Cognitive Technology: Instruments of Mind* (London: Springer-Verlag). doi: 10.1007/3-540-44617-6_6

Werry, I. P., and Dautenhahn, K. (1999). "Applying mobile robot technology to the rehabilitation of autistic children," in *Proceedings of SIRS99, 7th Symp on Intelligent Robotic Systems* (Coimbra), 265–272.

Yoshikawa, Y., Kumazaki, H., Matsumoto, Y., Miyao, M., Kikuchi, M., and Ishiguro, H. (2019). Relaxing gaze aversion of adolescents with autism spectrum disorder in consecutive conversations with human and android robot-a preliminary study. *Front. Psychiatry* 10:370. doi: 10.3389/fpsyt.2019.00370

Yun, S.-S., Choi, J., Park, S.-K., Bong, G.-Y., and Yoo, H. (2017). Social skills training for children with autism spectrum disorder using a robotic behavioral intervention system: social skills training for children with ASD. *Autism Res.* 10, 1306–1323. doi: 10.1002/aur.1778

Zubrycki, I., and Granosik, G. (2016). Understanding therapists' needs and attitudes towards robotic support. The roboterapia project. *Int. J. Soc. Robot.* 8, 553–563. doi: 10.1007/s12369-016-0372-9

Zwarenstein, M., Treweek, S., Gagnier, J. J., Altman, D. G., Tunis, S., Haynes, B., et al. (2008). Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 337:a2390. doi: 10.1136/bmj.a2390

Conflict of Interest: The authors declare that this study received funding from AG2R La Mondiale. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication.

APPENDIX 1

List of Psychological Tests Used

The *SCQ* is a screening tool based on the DSM-IV criteria for autism and the *ADI-R* algorithm (Rutter et al., 2003). It takes the form of a standardized parent questionnaire to assist in autism diagnosis by capturing key autistic symptoms (e.g., "did he/she ever show you things that interested him/her to engage your attention?").

The *Vineland-II* is a structured interview administered to primary caregiver(s) to assess a child's daily living skills (e.g., "looks at the caregiver when s/he hears his voice"). Using this tool, we evaluated three domains (communication, daily living skills, socialization), thus obtaining an overall adaptive behavior evaluation.

The *PEP-3* identifies learning strengths and facilitates the selection of educational programs for children with ASD. In the present study, we scrutinized affective expressions (AE; e.g., "manifests an appropriate level of fear") and social reciprocity (SR; e.g., "initiates social interactions").

The *SPCR* assesses unusual sensory experiences of individuals with ASD (e.g., "covers ears when hears certain sounds").

The *ESES* comprises 13 items evaluating interventionists' beliefs about their efficiency in controlling children (e.g., "I am able to copy with disruptive behavior in a teaching session").

The *SCQ* and *Vineland-II* yield standardized scores, while the others yield raw scores. A high *SCQ* score indicates a severe form of ASD-like symptoms. For the remaining tests, low scores indicate a severe functional impairment (see **Table 1**).

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 09 April 2020
doi: 10.3389/frobt.2020.00046

Trouble and Repair in Child–Robot Interaction: A Study of Complex Interactions With a Robot Tutee in a Primary School Classroom

Sofia Serholt[1], Lena Pareto[2], Sara Ekström[2] and Sara Ljungblad[3]*

[1] *Division of Learning, Communication and IT, Department of Applied IT, University of Gothenburg, Gothenburg, Sweden,* [2] *Division of Media and Design, School of Business, Economics and IT, University West, Trollhättan, Sweden,* [3] *Division of Interaction Design, Department of Computer Science and Engineering, University of Gothenburg and Chalmers University of Technology, Gothenburg, Sweden*

OPEN ACCESS

Edited by:
Agnese Augello, Institute for High Performance Computing and Networking (ICAR), Italy

Reviewed by:
Paul Vogt, Tilburg University, Netherlands
Luke Jai Wood, University of Hertfordshire, United Kingdom

***Correspondence:**
Lena Pareto
lena.pareto@hv.se

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *18 November 2019*
Accepted: *17 March 2020*
Published: *09 April 2020*

Citation:
Serholt S, Pareto L, Ekström S and Ljungblad S (2020) Trouble and Repair in Child–Robot Interaction: A Study of Complex Interactions With a Robot Tutee in a Primary School Classroom. Front. Robot. AI 7:46. doi: 10.3389/frobt.2020.00046

Today, robots are studied and expected to be used in a range of social roles within classrooms. Yet, due to a number of limitations in social robots, robot interactions should be expected to occasionally suffer from troublesome situations and breakdowns. In this paper, we explore this issue by studying how children handle interaction trouble with a robot tutee in a classroom setting. The findings have implications not only for the design of robots, but also for evaluating their benefit in, and for, educational contexts. In this study, we conducted video analysis of children's group interactions with a robot tutee in a classroom setting, in order to explore the nature of these troubles in the wild. Within each group, children took turns acting as the primary interaction partner for the robot within the context of a mathematics game. Specifically, we examined what types of situations constitute trouble in these child–robot interactions, the strategies that individual children employ to cope with this trouble, as well as the strategies employed by other actors witnessing the trouble. By means of Interaction Analysis, we studied the video recordings of nine group interaction sessions ($n = 33$ children) in primary school grades 2 and 4. We found that sources of trouble related to the robot's social norm violations, which could be either active or passive. In terms of strategies, the children either persisted in their attempts at interacting with the robot by adapting their behavior in different ways, distanced themselves from the robot, or sought the help of present adults (i.e., a researcher in a teacher role, or an experimenter) or their peers (i.e., the child's classmates in each group). In terms of the witnessing actors, they addressed the trouble by providing guidance directed at the child interacting with the robot, or by intervening in the interaction. These findings reveal the unspoken rules by which children orient toward social robots, the complexities of child–robot interaction in the wild, and provide insights on children's perspectives and expectations of social robots in classroom contexts.

Keywords: child–robot interaction, education, social robotics, interaction trouble and repair, group interaction, robot tutee, in the wild, classroom study

INTRODUCTION

Over the past decades, research has explored the possibility of using social robots in a range of educational roles, including as teachers and tutors, peers, and novices (Belpaeme et al., 2018). For instance, the EMOTE project developed a robot with empathic qualities, which could tutor primary school students on tasks related to geography and sustainable development (Serholt and Barendregt, 2016; Obaid et al., 2018; Alves-Oliveira et al., 2019), the L2TOR project developed a robot that could tutor preschool children on second language learning (Vogt et al., 2019), whereas the CoWriter project developed a robot in the role of a novice, which children could teach handwriting skills to (El-Hamamsy et al., 2019). The motivations behind these efforts range from explorations of robots as technologies for supporting children's learning [e.g., language learning (Kory-Westlund and Breazeal, 2019)] and the development of targeted skills [e.g., self-regulated learning (Jones and Castellano, 2018)], to a conception of robots as solutions to various educational challenges, such as teachers' workload (Movellan et al., 2005) and a global teacher shortage (Edwards and Cheok, 2018). While social robots may not be used on a regular basis in education at present (Selwyn, 2019), researchers and developers continue to design novel applications for robots that aim to support education in various ways.

One caveat to this kind of research, and by extension to the benefit and usefulness of implementing social robots in education, lies in the fact that current robot solutions are expensive, have limited functionality, and are prone to breakdowns of both a social and technical nature. Ros et al. (2011) noted these difficulties during their extensive studies of Child–Robot Interaction (CRI) in a hospital setting; accordingly, they argued for the need to plan such studies appropriately by asserting that the robot used is mechanically robust, and by accounting for unpredictability in children's behavior. However, recent research suggests that these challenges are still prevalent (Belpaeme et al., 2018; Serholt, 2018), and this is partly related to the difficulty in predicting social behavior. As Honig and Oron-Gilad (2018) put it: "While substantial effort has been invested in making robots more reliable, experience demonstrates that robots operating in unstructured environments are often challenged by frequent failures" (p. 2). In CRI scenarios, social or technical breakdowns can lead to children's disappointment, loss of engagement (Ros et al., 2011), or even emotional distress (Serholt, 2018). An extended follow-up study showed that children also tend to remember such breakdown situations, even after 3 years (Serholt, 2019). However, little is known about the nature of these issues in CRI, and how children work to address or mitigate them in interaction. The lack of research on this topic provides a false presupposition that CRI is more frictionless than it actually is. By identifying and understanding the situations where robots fail in social interaction, it is possible to critically reflect on how to handle such situations from an educational and design perspective, while also furthering our understanding of how children interact with robots.

From the perspective of Interaction Analysis, breakdowns are usually preceded by what is known as "trouble" in interaction (Jordan and Henderson, 1995). Specifically, trouble in interaction becomes evident when it breaks the rhythmicity of an otherwise stable routine or interaction script, which is the given design in most CRI scenarios. When trouble occurs, people resort to repair strategies in order to handle the problem and avoid the occurrence of breakdowns (Jordan and Henderson, 1995). This process of trouble and repair probably becomes especially complicated when children deal with robots, since there is likely a mismatch between the children's and the robot's rules of interpretation—rules that are typically assumed to be somewhat aligned in everyday social interaction among people (Jordan and Henderson, 1995). In their systematic analysis of video data from five different Human–Robot Interaction (HRI) studies, Giuliani et al. (2015) found a number of differences in people's social responses to error situations in their interactions with robots. For instance, people displayed significantly more non-verbal social signals and spoke more when in a group or when an experimenter was present, vs. when they were alone. They also behaved differently depending on the type of failure, whether it was a social norm violation, i.e., "a deviation from the social script or the usage of the wrong social signals" (p. 3), or a technical failure. Giuliani et al. (2015) argue that evaluators of HRI systems should not discard data containing error situations, since it may contain valuable results.

As argued by Jordan and Henderson (1995), careful analysis of trouble in interaction "can often reveal the unspoken rules by which people organize their lives" and it is "one of the best methods for coming to an understanding of what the world looks like from somebody else's point of view" (p. 69). Hence, Interaction Analysis lends itself to exploring particular challenges related to designing robots for children, the expectations that children may have of interactions with robots, along with an understanding of the repair strategies children employ when their social expectations do not align with the social script of the robot. However, little is known about the detailed, sequential mechanisms by which interactions between children and robots play out in naturalistic settings such as classrooms, and the strategies that children employ in the face of trouble. In their recent literature review study, Honig and Oron-Gilad (2018) explored robot failures in HRI, including how people perceive and resolve these failures. However, the authors found that most such studies have been conducted in controlled, single-person environments, and that they therefore lack in ecological validity. Moreover, very few studies have considered children as the explicit target group. One exception is a previous experimental study where pairs of children aged 4–5 played a game with a robot that feigned getting lost, disobeyed the children's instructions, or made a mistake and recovered (Lemaignan et al., 2015). The authors were unable to affirm whether the children could perceive the difference between what they intended to be understood as a technical malfunction (i.e., the robot getting lost), and intentional social behavior (i.e., the robot disobeying the children, or making a mistake and recovering). The authors recommended that similar studies be replicated with older children. Another exception is an earlier study of interaction breakdowns between children and a robot tutor conducted by one of the authors of the current paper, where breakdowns were caused by both technical malfunctions, as well as social and pedagogical norm violations (Serholt, 2018).

Against this background, we present a qualitative analysis of video data obtained from a CRI field trial in a primary school classroom. As suggested by, e.g., Honig and Oron-Gilad (2018), the trial was designed to have high ecological validity, i.e., it took place in a familiar environment (the children's ordinary classrooms), and it included a variety of actors and artifacts. These actors and artifacts consisted of a social robot tutee seeking to learn arithmetic from the children, an interactive whiteboard displaying a mathematics game, groups of children in which one individual could interact directly with the robot at a time, a researcher in a teacher role, and an experimenter.

The initial aim of this field trial was to observe children's interactions with robots in naturalistic settings, in order to derive design recommendations for robot tutees. Yet, as we *familiarized ourselves with our data* (Braun and Clarke, 2006) through our qualitative, inductive approach, it became evident that the videos contained rich data regarding interaction trouble and repair strategies. Thus, the aim of the current paper is to explore trouble and repair in CRI. These findings do not only hold implications for the design of social robots for classrooms, but they also reveal the unspoken rules and/or silent expectations that children may have of robots in educational settings. The following research questions guide this study:

RQ1: What situations and/or behaviors constitute trouble in the child–robot interaction situation?
RQ2: What strategies do children employ when trouble occurs in the child–robot interaction situation?
RQ3: What strategies do the other actors (e.g., peer group members, researcher as teacher, and experimenter) employ when witnessing trouble in the child–robot interaction situation?

MATERIALS AND METHODS

We conducted field trials with a robot tutee under development in our research project Student Tutor and Robot Tutee (START), and an accompanying mathematics game at two primary schools in Sweden. The field trial constituted a first test of the children's interactions with the setup in a complex classroom setting with multiple actors. The trial took two full days at each school. The students participated in the trial in groups of four, scheduled by their teachers, and as part of their regular school activities.

Apparatus

The technical setup consisted of the humanoid robot Pepper from Softbank Robotics[1] and a digital mathematics game displayed on a wall-mounted screen. The mathematics game was adopted from a previously developed game called the Graphical Arithmetic Game, stemming from research on game-based learning and teachable virtual agents (Pareto, 2014). In our research project, the game has been updated and augmented to include a physical robot acting as a tutee and a co-player in the game (Pareto, 2017; Pareto et al., 2019). The use of teachable agents or robot tutees draws on learning-by-teaching and peer-assisted learning approaches, where children are engaged in the activity of teaching a novice or peer in order to further their own learning of a specific topic (see e.g., the CoWriter project: El-Hamamsy et al., 2019).

The mathematics game selected for this study constitutes a collaborative mini-game in the Graphical Arithmetic Game called 10-buddies. The game is a simple 2 player addition game, with the goal to add to ten by taking turns and choosing cards from the two players' respective card hands. Card values are graphically represented through colored blocks, ranging from values 1–9. In this case, a child and a fully autonomous robot tutee constitute the active players. However, as long as the robot tutee is at a novice stage, it does not actively play its own cards. Instead, the robot observes the child's choices and utilizes the existing question-and-answer repertoire of the earlier entirely text-based virtual agent, while it also exhibits socially interactive behaviors. This includes the display of some pre-programmed movement, gestures, and gazing behaviors, along with the implementation of a text-to-speech module in Swedish, in order to support verbal communication. The robot connects to the game through a local wireless network, and the game steers its behavior based on the child's actions in the game. In terms of the robot's verbal repertoire, there is a progression in what kind of questions the robot asks, depending on how well the children play and how well they manage to answer these questions. Typical questions in the beginning concern the overall game idea: how to score points and what the objects on the display mean. Then, the robot progresses to inquire about which cards will yield points, and finally, which cards are strategically smart to play considering future turns and possibilities. Hence, the robot consistently features as the children's inquisitive tutee whose goal is to learn how to play the game, and to improve its skills pertaining to mathematics. In the current study, each student group began with a novice robot tutee to teach, so all groups played both player turns, and answered the same type of inquisitive questions; the robot was programmed to ask such questions whenever the child selected a card from its hand. In order to facilitate progression in the interaction, the robot was designed to move on to the next step in the interaction if it did not receive, or was unable to perceive, any input from the game-playing child (such as a verbal response to its question). This occurred after a waiting period of 45 s, and was indicated by a verbal utterance, such as "Let's move on." The robot could also return to an earlier question by asking, "Have you come up with the answer to this question: [*question*]?"

The game and robot were placed in school spaces designated by the staff, meaning that the game had to be displayed on the schools' available equipment. In one school (School A), the trial was conducted in an empty classroom with a projector and wall-mounted canvas; in the other school (School B), a room for after-school activities with an interactive whiteboard was used. The game was displayed on the screen or the whiteboard with the robot standing in front of the display together with

[1] https://www.softbankrobotics.com/emea/en/pepper

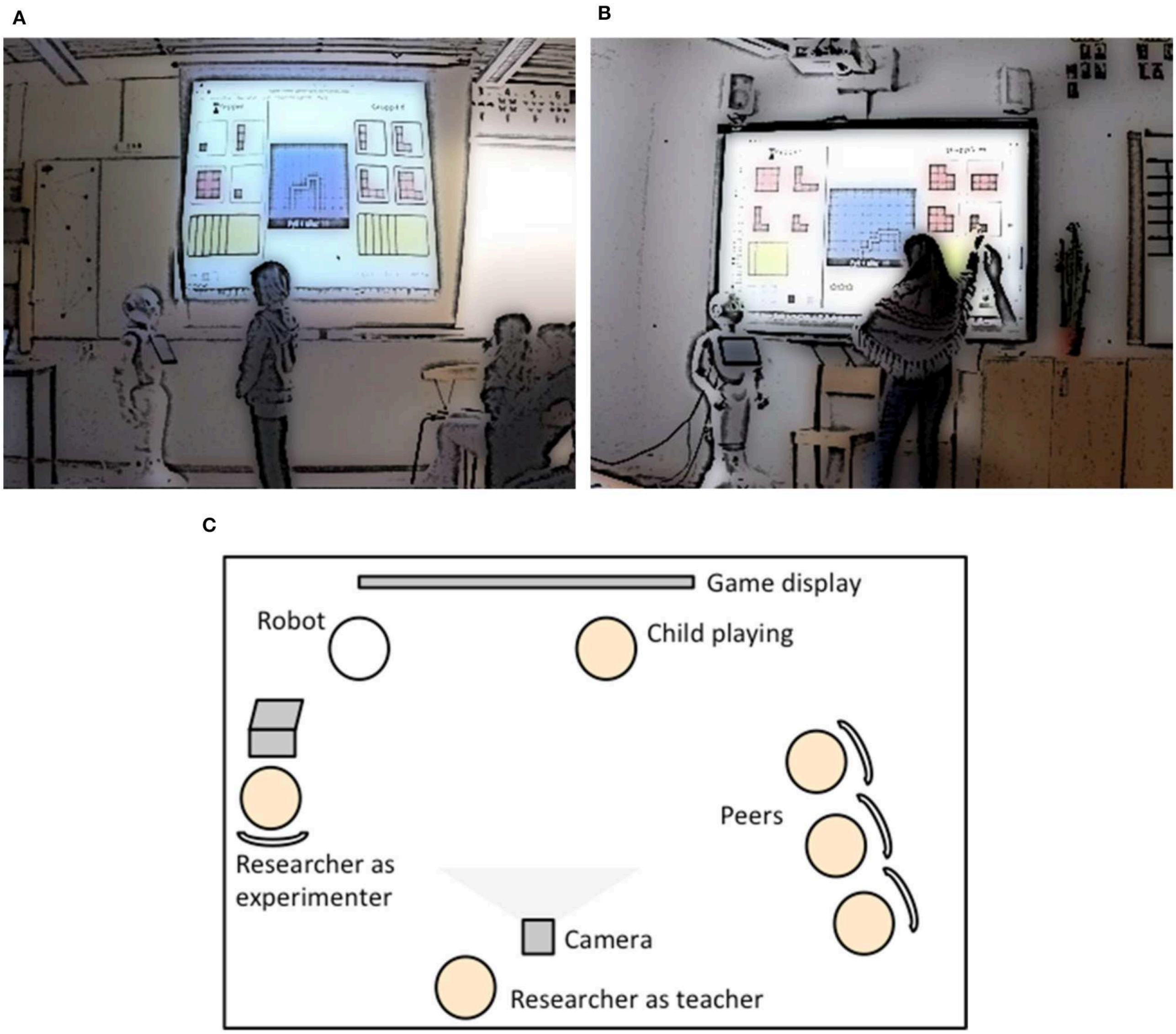


FIGURE 1 | (A) Still image of the field trial setup at School A. **(B)** Still image of the field trial setup at School B. **(C)** Top-view illustration of the field trial setups.

the game-playing child. The children in each group took turns playing the game with the robot; the children currently not playing were seated next to the scene, accompanied by one of the authors (researcher in teacher role), who is also a licensed teacher with 15 years of teaching experience. Her role was to facilitate and organize the children's collaboration during the sessions, observe the interaction, and manage a video recorder. Another author (the experimenter) was tasked with handling the technical aspects of the game, i.e., starting the game and making sure that everything was working, while also executing the children's choices of cards during their turns in School A where the display was not interactive. Finally, a video camera placed in the middle of each room captured the game display, the robot, and the child from behind. For illustrations of the interaction sessions and field trial setups, see **Figure 1**.

Participants

Two classes from each school participated in the study: two 2nd grade classes from School A and two 4th grade classes from School B, i.e., children of ages 8 and 10 years old. The classes were selected based on active interest from the children's respective teachers to enroll in our research project. Thus, the same children participated in a workshop on a previous occasion, which consisted of a robot-programming task and a post-workshop questionnaire (Pareto et al., 2019). The children had never played the Graphical Arithmetic Game before. In total 69 children across 19 groups participated in the trial: 28 children from the 2nd grade, and 41 children from the 4th grade. In the current study, we randomly sampled the sessions of nine groups for analysis ($n_{\text{children}} = 33$; 15 female, 18 male; 17 second-graders, and 16 fourth-graders).

Procedure

Prior to the study, the class teachers divided the children into groups of three or four. During the study, the teachers excused each group from their classroom to take part in the trial in the allocated room for 30 min each. The trial sessions then proceeded as follows: First, the researchers welcomed the children, described the aim of the study, and asked for confirmation of the children's previous assent to participate in the study. Second, the researchers briefly explained the game's aim and rules. They also explained that the children would take turns playing with the robot, but that the group members were encouraged to help the child playing the game, and to suggest answers to the robot's questions. Then, a video recorder was activated, and the game session was initiated. During the session, the children took turns on a voluntary basis to play the game together with the robot, where each child was allowed to play for about 6–7 min before being asked to take a seat, whereupon another child could volunteer. In the analyzed sessions, the children played on average 10 cards each, including the choices they made for the robot. After 30 min, the researchers thanked the groups for their participation, held a debriefing session about their experience, and followed them back to class.

Data Collection and Analysis

During the field trial, we collected video recordings of the interaction sessions. As mentioned previously, nine of the interaction videos were randomly selected for analysis, which amounted to a total of 3 h and 15 min of video data. For this study, we adopted a qualitative inductive approach, drawing on thematic analysis (Braun and Clarke, 2006) and Interaction Analysis (Jordan and Henderson, 1995).

Thematic and Interaction Analysis

The first phase in our analysis involved what Braun and Clarke (2006) refer to as *familiarizing yourself with your data*, i.e., the search for interesting areas of study in the material without explicit protocol. Specifically, the videos were viewed independently by two of the authors (henceforth referred to as coders), who made notes regarding their observations. These observations were discussed with the remaining authors through a joint data session where all authors viewed and discussed the content of selected videos. At this stage, consensus was reached that the data contained rich material regarding interaction trouble and repair strategies.

The next phase involved conducting interaction analyses of the sessions (Jordan and Henderson, 1995). First, the videos were divided between the two coders who each created interaction transcripts for half of the videos. The interaction transcripts contained high-level documentation of the sequential interaction processes, i.e., what each actor in the material was doing at specific times, descriptions of the interaction between the different actors, along with the coders' analyses of the interactions. One such transcript was produced for each game-playing child in their respective group ($M_{\text{duration}} = 5\,\text{min}\ 55$ sec; $min = 1\,\text{min}\ 38\,\text{s}$; $max = 10\,\text{min}\ 45\,\text{s}$), meaning that between three and four transcripts were produced for each interaction session (33 transcripts in total). Segments containing trouble in interaction were documented in detail, whereas segments containing fluent turn-taking and gameplay were just commented as such. To identify trouble, any situation, which seemingly disrupted the interaction flow, was considered.

The third phase in the analysis consisted of coding the data, in which the two coders independently coded half of the videos each. We followed the three-stage process for coding qualitative data suggested by Campbell et al. (2013): (1) developing a coding scheme with as high intercoder reliability as possible based on a sample of transcripts (typically 10%), (2) negotiating coding disagreements among coders until reaching acceptable levels of intercoder agreement (as the recommended approach in exploratory research), and (3) deploying the coding scheme to the full set of transcripts. During the first stage, a preliminary coding scheme was developed by the more experienced coder on a sample of two sessions, which was tested by the other coder under supervision. This procedure generated a list of codes, which each contained a qualitative description of an observed action along with a label. The codes were organized into inductively formulated code families, each denoting a common topic (Campbell et al., 2013). Four transcripts out of 33 (12%) were coded by both coders independently and checked for intercoder reliability. Following Campbell et al. (2013), intercoder reliability was calculated as the number of common instances of codes (i.e., agreement in coding) divided by the total instances of codes (i.e., agreement + disagreement). The average level of intercoder reliability was 73%. For the second stage, the coding disagreements (33 out of 125 codes) were analyzed and discussed by the coders. The most frequent differences were whether subtle non-verbal actions occurred or not (18 disagreements), and whether the child or the adult initiated a help action (7 disagreements). The coding scheme was refined to address these differences. For the third stage, the remaining sessions were divided between the two coders and coded independently. Given our inductive approach, the list of codes evolved and was continuously discussed, compared, and unified during the process, producing a joint coding scheme. The final coding scheme consists of seven (primary) code families and 36 (secondary) codes (see **Supplementary Materials Table A**).

The final phase in our analysis involved developing themes to describe the nature of trouble in CRI, and the following repair strategies. This phase was carried out by one of the authors who developed themes based on the coding scheme and interaction transcripts. The themes were discussed and reformulated through several iterations with the remaining authors.

RESULTS

In this study, we set out to explore trouble and repair in CRI. This analytical interest stemmed from our observations of children's group interactions with a robot tutee in a classroom setting, wherein trouble (and repair) seemed prevalent. By means of Interaction Analysis and thematic analysis, we explored situations of trouble and repair, which constituted 26.4% of the

TABLE 1 | Overview of themes and subthemes derived from the thematic analysis.

Research question	Main themes	Subthemes
RQ1: Sources of trouble	Active social norm violations	Makes irrelevant comments
		Interrupts Signals dismissal through non-verbal behavior
	Passive social norm violations	Fails to act at its designated turn in the game
		Fails to respond verbally
RQ2: Children's repair strategies	Adapt to the robot	Exaggerate articulateness
		Modifying tutoring approach
		Seeking to understand interaction form
	Establish distance to the robot	Making the robot invisible
		Give up
	Shift focus to human actors	Seek affirmation
		Request help
RQ3: Strategies of other actors	Offer help	Provide guidance to the child Intervene (or interfere) in the interaction

TABLE 2 | Number of occurrences and children who encountered each source of trouble.

Main themes and subthemes	No. of occurrences	Percentage of children who experienced each theme
Active social norm violation	64	88%
Makes irrelevant comments	28	55%
Interrupts	21	42%
Signals dismissal through non-verbal behavior	15	18%
Passive social norm violations	92	73%
Fails to respond verbally	83	70%
Fails to act at its designated turn in the game	9	18%

For main themes, child percentages are based on the whole sample (n = 33); for subthemes, child percentages are instead based on the number of children within each main theme.

sampled video corpus (the remaining segments of the video corpus depicted what we considered to be fluent interactions).

In this section, we present our findings in individual subsections for each research question. Within each subsection, main themes are represented through italicized, bold, font (i.e., ***main theme***), subthemes are indicated as such through bold font (i.e., **subtheme**), and translated excerpts derived from the Interaction Analysis are shown for illustration and discussion purposes. Individual children are denoted through their participant IDs (**C** for game-playing child accompanied by a number). **Table 1** provides an overview of all themes and subthemes.

The Sources of Trouble

We found that situations and/or behaviors that constituted trouble in this particular CRI situation (i.e., the sources of trouble) were related to the robot's social norm violations, which were either ***active social norm violations*** (41%) or ***passive social norm violations*** (59%). Although these violations could be traced back to technical issues or limitations with the robot, this analysis is concerned with exploring these situations from an interaction perspective. Hence, sources of trouble are considered from the perspective of how it might be interpreted in social interaction.

Trouble stemming from the robot's ***active social norm violations*** manifested in different ways. Yet, the commonality was that these behaviors were unexpected and undesirable. First, the robot sometimes **made irrelevant comments**, which constituted 44% of all active social norm violations. For example, when C3 was in the process of explaining to the robot that they needed to try again, the robot responded with the following contextually irrelevant comment:*"Yes I know that 7 + 3 = 10."* Second, the robot sometimes **interrupted** (33%) the child speaking. For instance, as C1 was in the process of providing a response to one of the robot's questions, the robot unexpectedly announced, *"Now we continue to play!"* which could be perceived as general disinterest or disregard for what the child had to say. Third, trouble also surfaced when the robot **signaled dismissal through non-verbal behavior** (23%), e.g., when it turned its back to the child in the middle of an interaction. In one situation, the child called out to the robot in order to encourage it to turn around and face him/her; instead of doing so, the robot merely responded: *"Yes, I hear you"* which, in interactions between humans, would likely be interpreted as disengaged and dismissive behavior.

Regarding the robot's ***passive social norm violations***, there were several situations when the robot simply failed to act as expected. For instance, the robot could lose its connection to the game and consequently **fail to act at its designated turn in the game**, yet this only accounted for 10% of its passive behavior. More common was the robot's **failure to respond verbally** (90%) to the child when such behavior seemed mandated. This could occur during the child's attempts to greet the robot, but also during dialogues connected to the gameplay. For instance, for C4, the robot inquired as to how they would receive points in the game, for which he provided a verbal explanation; the robot, however, only acknowledged his explanation non-verbally (by nodding), which caused trouble since he became uncertain as to whether the robot had actually understood.

In **Table 2**, data regarding the number of times each source of trouble was observed, along with the number of children across the whole dataset who encountered it, is presented. In total 32 of the 33 children in this study encountered some form of trouble during their sessions.

Children's Repair Strategies

Our analysis shows that children use different repair strategies in different situations. Specifically, the children either persisted in their attempts at interacting with the robot by modifying

TABLE 3 | Excerpt for C2.

C2		**Male, 2nd grade, school A, second player of his group**		
Time	**Actor**	**Verbal**	**Non-verbal**	**Note**
18.22	Robot	Why are there 10 squares in the enclosed area?		
18.25	Child	I don't know.		
	Robot		Nods	Trouble: fails to respond verbally
18.34	Child	Listen. I don't know	Moves closer to the robot and speaks close to its face	Repair: exaggerate articulateness

TABLE 4 | Excerpt for C4.

C4		**Male, 2nd grade, school A, first player of his group**		
Time	**Actor**	**Verbal**	**Non-verbal**	**Note**
10.22	Robot	Have you come up with the answer to this question: how do we get points?		The first part of the question denotes that the robot has asked this question before, but not perceived a response
10.28	Child	We will fill these boxes.	Points to the enclosed area of the game board	Repair: modifying tutoring approach
	Robot		Nods	Trouble: fails to respond verbally
10.33	Child	Then we get stars. There are points.	Points to the score meter	Repair: modifying tutoring approach

their behavior and ***adapting to the robot*** in different ways, ***distancing themselves from the robot***, or ***shifting focus to the human actors*** present (i.e., the researcher in a teacher role, the experimenter, or the child's peer group members). These main themes will be presented in turn below. Notwithstanding, the same child typically employed a variety of strategies within the same session, such that these categories are not by any means mutually exclusive on a child-by-child basis.

Regarding children's methods for ***adapting to the robot***, the children could **exaggerate articulateness**. In such cases, children could change the ways in which they communicated verbally with the robot by either shortening their responses to simple keywords, or strengthening the volume and clarity of their verbal communication. In the excerpt shown in **Table 3**, C2 is playing the game, whereby the robot poses a question.

As can be observed in this example, the robot poses an inquisitive question regarding the game board to gauge the significance of the value ten. C2 responds verbally that he does not know why there are ten empty boxes in the enclosed area on the game board. That the robot nods but fails to respond verbally indicates to C2 that the robot has not properly heard or understood his answer, causing temporary trouble. In response, C2 attempts to repair this trouble by trying to get the robot's attention (when he moves closer and says: *"Listen"*), and by trying to make his response audible.

In contrast to these exaggeration strategies, we also found that the children tried to adapt to the robot in more social ways, e.g., by **modifying their tutoring approach**. Specifically, this could entail the children elaborating upon a mathematical concept, or explaining in a different way than they had done initially. In some cases, these modified explanations were complemented by visual demonstration on the game display through a variety of gestures, such as pointing to elements in the game. The children could also ask the robot to repeat or explain itself (e.g., *"I didn't hear what you said in the beginning"* [C26]), or simply instruct the robot on which cards to play. Taken together, such adaptations could indicate that the children perceived the robot as a social other capable of perceiving and interpreting complex human reasoning. In the excerpt shown in **Table 4**, the robot asks C4 a question while he is playing the game.

Earlier in the interaction, the robot had already asked how they receive points, but had not perceived or understood the response. Against this experience, C4 thus tries to modify his response by complementing his verbal explanation with gestures directed at the game board. In response, the robot simply nods and fails to respond verbally, which is interpreted by C4 as a signal that the robot does not quite understand. Hence, C4 once again tries to modify his tutoring approach by explaining the game mechanics in a different way (with reference to the scoring of points by acquiring stars).

Another way in which children tried to adapt to the robot concerned their **seeking to understand the interaction form**, where they also seemed open toward interacting on the robot's terms. While this was also the case when the children exaggerated their articulateness, this subtheme differed in relation to the children's seeming curiosity. For instance, they could increase their proximity to the robot and perform exaggerated gestures in an attempt to make the robot perceive and recognize their interaction endeavor. Yet, unlike the situations where the children would only utter keywords, presumably for the sake of the robot's speech recognition difficulties, these communication attempts seemed more related to the children's desire to establish communication with the robot (e.g., C33 who asked the robot, *"What are you doing?"* when it failed to respond). In some cases, the children would wait patiently for the robot to act while they stood in front of it. In other cases, the children would mirror the robot's non-verbal behavior (e.g., C25, who switched from verbal communication to mirroring the robot's frequent head nodding).

In contrast to adaptive behaviors, the children also responded to trouble in interaction by ***establishing distance to the robot***. This strategy mainly occurred after a long sequence of trouble; hence, it was usually preceded by some form of overt expression of emotional distress such as discomfort or irritation. Some of the children established distance by **making the robot invisible**, i.e., a form of domination technique. Specifically, this could manifest itself through the children talking over the robot, i.e.,

TABLE 5 | Excerpt for C7.

C7		Female, 2nd grade, school A, fourth player of her group		
Time	**Actor**	**Verbal**	**Non-verbal**	**Note**
25.44	Robot	Okay, 5.		
25.50	Child	Yes	Leaning forward	Repair: exaggerate articulateness
	Child		Standing still	Repair: seeking to understand interaction form
25.55	Robot	Now we have 4 points.		
26,01	Child	I don't want to play anymore.		Give up

speaking simultaneously as the robot, but not directed at the robot as such. Some children simply ignored the robot's questions completely, whereas some children took a less overt approach and acknowledged the robot's questions, but provided an indifferent response (e.g., *"Mm"*). They could also interrupt the robot by quickly answering *"Yes"* or *"No"* at the start of the robot's utterance. Some children also chose to **give up** on the interaction, either by walking away and taking a seat, or by stating that they did not want to continue the interaction, as demonstrated in the following excerpt (see **Table 5**).

Finally, children's strategies consisted of ***shifting focus to the human actors*** who were present (either the researchers or their peer group members). This typically occurred when the children had exhausted other repair strategies more directed toward the robot. From these other actors, the children often **sought affirmation** regarding their responses to the robot's dialogue (e.g., checking with researchers or peers that their particular response would be appropriate), but also related to gameplay choices (e.g., asking peers or researchers to confirm that their card selection would afford points). In more difficult situations, however, the children would **request help** in open-ended and explicit ways, indicating both verbally and non-verbally that they did not understand how to proceed in the dialogue with the robot or the game.

Strategies of Other Actors

In terms of the other actors present in the interaction sessions (i.e., the child's group members, the researcher as experimenter, and the researcher in a teacher role), their repair strategies consisted of ***offering help*** in various ways. In most cases, they tried to **provide guidance to the child**, which meant that they addressed the game-playing child directly, and conveyed various forms of scaffolding for interacting with the robot successfully, but also regarding strategic moves in the game. They also **intervened (or interfered) in the interaction** by responding to the robot directly. For instance, the peers could call out the correct answer, or try to get the robot's attention, but this was quite rare. On a few occasions when the robot signaled dismissal through non-verbal behavior by turning its back to the children, the experimenter intervened as shown in **Figure 2**.

In terms of providing guidance to the child, the experimenter, who possessed technical knowledge about the robot and the game, could suggest specific verbal formulations that the robot would understand. The peer group members could also become involved in this process of trouble and repair, leading to a complex interaction situation. Often times, the peers drew on their previous experiences having heard similar questions from the robot during their turns. The excerpt in **Table 6** provides an illustration of when the experimenter and the peer group members provided guidance to C15. Right before the excerpt, C15 indicates that he does not know how to answer the robot's question and consequently stays silent while grabbing the robot's hand.

As can be observed in this excerpt, the experimenter tries to guide the child when she notes that the child does not know how to answer the question. She does so by involving C15's peer group members so that they can provide input on what to say to the robot based on their experience from the first player round. What follows is a series of trouble and repair strategies consisting of the robot failing to respond verbally, and C15 trying to make himself understood through exaggerated articulateness. These attempts go unsuccessful, and the interaction ends with the robot proclaiming that they should give up on that question and proceed instead.

DISCUSSION

This qualitative study has explored interaction trouble and repair in the field of CRI. Our analytical interest came about through our initial observations of children's classroom interactions with a robot tutee in the context of a collaborative game in mathematics, which took place in groups of children who took turns actively playing the game and teaching the robot.

Our research questions concerned what situations and/or behaviors constitute trouble in such a CRI setting, what repair strategies children employ to address interaction trouble, and how onlookers (researchers and peer group members) respond when witnessing the trouble. The results indicated that the primary source of interaction trouble related to the robot's social norm violations. Notwithstanding, there were a few additional situations, not due to the robot, which caused trouble as well. For instance, in one group in particular, the peer group members not actively playing the game with the robot disrupted the interaction by shouting out various directives at the game-playing child. Whereas, one of the children was able to ignore this behavior during her turn, another child became very distracted and began jumping around the room and throwing himself on the floor in an attempt to entertain his peers. However, for the sake of limiting our scope, we omitted such (rare) cases from further analysis.

Regarding the robot's social norm violations (RQ1), one could argue that these were, in a way, always a result of technical issues rather than an intentional design choice. Yet, the (social) interactional setting did not provide any actual

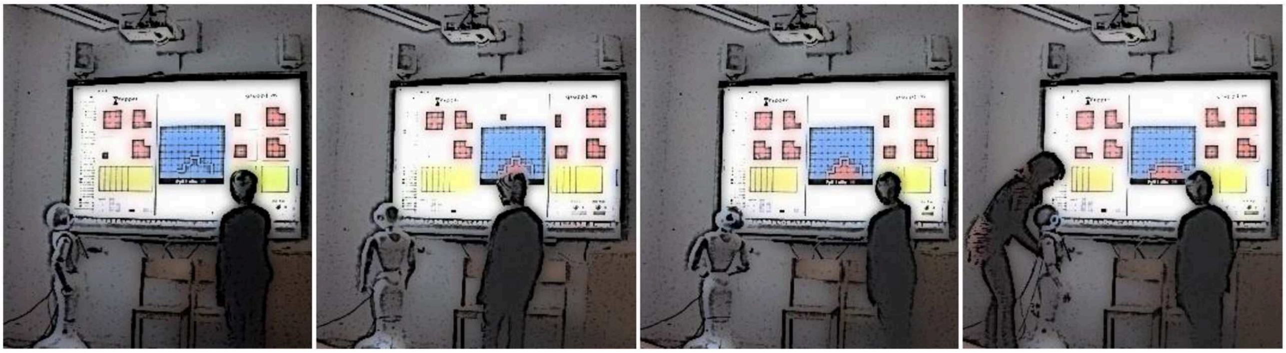


FIGURE 2 | Still images from an interaction session illustrating the experimenter intervening when the robot has turned its back to the child.

TABLE 6 | Excerpt for C15.

C15		**Male, 2nd grade, school A, second player of his group**		
Time	**Actor**	**Verbal**	**Non-verbal**	**Note**
14.48	Experimenter	It is your job to answer Pepper [the robot].		Repair: provides guidance to the child
14.51	Child		Releases the robot's hand and continues to play	
14.55	Experimenter	Ask your peers what they said.		
14.59	Peers	That you should get 10.		Repair: provide guidance to the child
15.00	Child	To get 10.	Talks close to the robot	Repair: exaggerate articulateness
	Robot		Nods	Trouble: fails to respond verbally
15.03	Experimenter	One more time.		Repair: provides guidance to the child
15.08	Child	To get 10.	Talks even closer to the robot	Repair: exaggerate articulateness
	Robot		Gazes at the game	Trouble: fails to respond verbally
15.11	Experimenter	Stand in front of it so Pepper [the robot] sees you.		Repair: provides guidance to the child
15.16	Child	(laughingly) To get 10.	Standing on his toes right in front of the robot	Repair: exaggerate articulateness
15.23	Robot	Let's leave this [question] now.	Gazes at the game	Trouble: makes irrelevant comment
15.25	Peers		Laughing	

opportunities for children to differentiate between social vs. technical issues, making it futile to discuss these differences from an interaction perspective. Our findings resonate with an earlier literature review on failure in HRI (Honig and Oron-Gilad, 2018). Indeed, all sources of trouble identified in the current study have been observed in HRI before, albeit with the obvious contextual variations. Compare, e.g., the similarities between our subthemes and the descriptions of errors and symptoms identified by Honig and Oron-Gilad (2018): **Interrupts** vs. "timing speech improperly," **makes irrelevant comments** vs. "producing inappropriate speech or erroneous instruction," **fails to act at its designated turn in the game** and **fails to respond verbally** vs. "producing no action or speech (irresponsiveness)," and finally, **signals dismissal through non-verbal behavior** vs. "producing unexpected or erratic behavior." It thus seems that these issues are not limited to a specific set of robot products, but actually a common challenge faced by several research projects in HRI; examples of this are, however, much rarer in CRI.

Turning to children's repair strategies (RQ2), these were many and varied in this study, including adapting to the robot's shortcomings in perception by exaggerating articulateness, adapting to its lack of knowledge in mathematics by modifying their tutoring approach, or by adapting to what they believed to be the robot's interaction modalities. We found that children used these strategies not only in response to the trouble currently taking place, but also as proactive measures throughout the interaction sessions. This suggests that children reiterated their understanding of the robot's capability as the interaction progressed. Children also shifted their focus to the human actors in the room, and sought their guidance with the interaction and task. They could also establish distance to the robot in various ways. Moreover, children used different strategies in close succession in a trial-and-error fashion. For instance, they could begin by modifying their tutoring approach, and then decide to request help from peers or researchers, and finally end up giving up on the interaction altogether.

As our analysis of the strategies of the peer group members and the researchers reveal (RQ3), the children did not necessarily need to request help as this was in many cases offered voluntarily. Typically, such guidance consisted of scaffolding the child currently interacting with the robot on what to say, and how to say it, in order for it to be perceptible to the robot. In other cases, peers and researchers intervened and spoke directly to the robot (or when the experimenter needed to physically turn the robot around to face the child); this type of intervention was, however, quite rare in our video data.

Taken together, the presence of additional actors in the room made various forms of support possible during the interaction. In contrast to one of the author's earlier studies of breakdowns in CRI (Serholt, 2018), children in the current study were perhaps able to avoid breakdowns largely due to the presence of other actors (researchers in particular), which enabled a form of collaborative repair work to take place. Indeed, children often turned to their peers and the researchers to repair troublesome situations. According to Serholt (2018), collaborations among peers during CRI can allow for a higher level of social support. However, it can also have certain drawbacks for the learning situation, such as children ignoring or mocking the robot that is supposed to facilitate their learning processes. Similar tendencies were found in the current study, specifically in relation to our subtheme **making the robot invisible**. While we did not observe mocking behaviors toward the robot *per se*, it is likely that the presence of adults (the researchers) actually discouraged children from such overt expressions of discontent. This should be considered from the wider perspective of implementing social robots in classrooms, where allowing children to interact with a robot on their own or in groups, vs. only in the presence of their teachers, requires understanding of the tradeoffs in order to reach a conscious and sensible solution. At present, research implies that children should not be left alone with educational social robots at all (Serholt et al., 2017; Newton and Newton, 2019).

Social robots are typically autonomous, embodied robots that may vary in form and behavior, but that are developed to follow certain social behaviors that is expected in its role. A previous study showed that people cooperated more with a robot whose social behavior was matched appropriately with a task (Goetz et al., 2003). This suggests that the willingness of children to collaborate with social robots in the classroom may depend on the extent to which its behavior fits the task and the overall situation. Another aspect concerns what kind of mental processes a robot in the classroom may facilitate when collaborating with children. An important related field to understand aspects of human cognition in HRI are social and cognitive neuroscience studies of human–robot and human–human interaction (Cross et al., 2019). For example, Rauchbauer et al. (2019) used functional magnetic resonance imaging (fMRI) to investigate neurological differences when people carried out a conversation with a robot compared to a person as an interaction partner in a task. The brain imaging findings revealed that human interaction led to engagement of brain regions associated with higher-order social cognitive processes, including the temporo-parietal junction. Performing the same task with a robotic interaction partner instead activated dorsal frontal and parietal brain regions. This indicates that human interactions engage more social motivation and mentalizing processes, while interactions with robots recruit additional executive and perceptual resources. This reveals some of the limitations of interactions with robots, and points toward the importance of peers and teachers to stimulate higher-order social cognitive processes among children.

From a design perspective, there are many ways in which these results can be considered and used. As suggested in previous work, robot interaction design may benefit from including socially based recovery strategies following a breakdown or trouble in interaction in order to promote long-term acceptance. Although we have studied a robot tutee only, we believe that our findings can be valuable for the development of social robots for children in general. Indeed, the social aspects of interaction with robots is not specific to the tutee role, even if, say, children's perceptions of the robot as a novice may have made them more forgiving toward its misunderstandings. According to Uchida et al. (2019), HRI researchers should not only focus on improving a robot's dialogue capability, but also consider ways to encourage cooperative intentions from users so that the user and robot will adopt an equal share of responsibility for breakdowns in dialogue. This is, indeed, interesting, and perhaps quite relevant for robot tutees, since much responsibility for joint understanding should probably fall on the tutor (child) rather than the tutee. We can already see this taking place in some of the children's adaptive strategies, specifically when they **modified their tutoring approach** in different ways. Of course, for this to be potentially beneficial, it would require the robot to be perceptive to these strategies.

Currently, off-the-shelf social robots are rather expensive, and extremely limited in functionality. Using a social robot in a classroom also requires technical expertise; not to mention the maintenance and updates required. For instance, when Davison et al. (2020) recently deployed a social robot in a classroom for 4 months in an unsupervised study, the researchers conducted all maintenance at particular times after school hours, meaning that there were times during the school day(s) when the robot could not be used as planned. It is paramount to apply and evaluate novel CRI systems, but considering the current technical limitations, these are far from ready to be implemented in schools to support teaching on a large scale. Conversational systems lack understanding of meaning and context and typically act on scripts in a pre-designed type of conversation (Serholt, 2018). Although there are noteworthy examples of somewhat long-term studies of autonomous social robots being conducted in classrooms (Serholt and Barendregt, 2016; Alves-Oliveira et al., 2019; Davison et al., 2020), many studies still require some degree of teleoperation for the interaction to work smoothly, particularly when it comes to verbal communication (Kory-Westlund and Breazeal, 2019; Vogt et al., 2019). This suggests that interaction trouble will likely continue to be a prominent feature of conversational interactions with autonomous social robots. It is our conviction that children cannot be expected to possess the skills necessary to repair all troublesome situations that follow, especially since educational robots can only be used in such delimited contexts (educational robots are seldom designed to function within more than a few specific educational activities), and duration during an ordinary school day. From this perspective, it is the responsibility of the designers of robots to make sure that the interaction works somewhat fluently. From a wider perspective of social robots in education, it is also necessary to consider the ethical aspects of implementing them in schools (Sharkey, 2016; Serholt et al., 2017).

Limitations and Future Directions

There are several limitations to this study that should be considered. First, the study was limited to two schools in Sweden, the sample size was rather small, while we only considered a particular CRI setup. This makes our findings difficult to generalize to other contexts. Nevertheless, social interaction with robots, and children's expectations of such interactions discernible through their repair strategies, constitutes a first step in understanding these issues more generally. This study did not focus on the children's experiences of the interaction, their views of social norms, or their preferences in teaching methods and learning experiences. We welcome future research that can demonstrate additional themes to explain trouble and repair, as well as other entryways to this topic relevant for CRI.

Second, the children in our study had some previous experience of programming the robot to execute simple dialogues. This could have influenced their perceptions of the robot as a machine with a limited social repertoire. Future research could potentially do a comparative study of how children handle trouble and repair depending on their levels of previous experience with robots.

Third, the interaction sessions analyzed in this study were quite brief and short-term. This means that the interactions were likely affected by a certain novelty effect, and that children's repair strategies could be developed even further after some time. The next step would be to investigate if the robot is perceived to add stress to the learning situation and how it is perceived during long-term use. Generally speaking, future research should continue working toward making HRI and CRI studies more long-term.

Fourth, another influential factor not yet touched upon in this paper relates to the mathematics game. Although our study was mainly concerned with exploring the social interaction between children and robots, the interactive display held a mediating role throughout the interactions. It constituted a boundary object for the children and robot to interact around, which conveyed awareness of the social situation they shared (e.g., the robot knew what cards the children played and commented on their actions). Hence, the task was not purely verbal; it was also graphically represented on the game board, around which the children and robot had a joint task. Future research should explore the influence of such boundary objects in CRI.

Fifth, a methodological limitation to this study is its lack of validation of the intercoder reliability level after the coding scheme was developed; instead, we relied on continuous intercoder discussions and agreements to address reliability. Although the aim of this study has been to provide a theoretical account of trouble and repair in CRI, some quantitative results are also presented in relation to the theme *sources of trouble*; thus, these findings should be interpreted with care.

Finally, although we strived toward making this study naturalistic and ecologically valid, it was not feasible to include the children's actual teachers in the study due to the need for technical expertise in operating the system. It is possible that regular teachers without any experience in robotics would employ other strategies for supporting the children than the researchers did, which should be considered in relation to our results. The study of teacher repair in CRI could be an interesting avenue for future research, which we intend to explore once our robot design has reached a more developed stage, also incorporating and evaluating a set of repair strategies in the face of trouble. Furthermore, it would be interesting to explore the connections between certain forms of trouble and certain forms of repair strategies. Due to the explorative nature and relative small-scale of this study, however, this was not possible to do here.

CONCLUSION

In this paper, we have explored trouble and repair strategies in children's interactions with a robot tutee in an educational setting. The aim of this study has been to shed light on the interaction issues in CRI under the premise that such issues can never be completely avoided or designed away. Trouble and repair in social interaction, while highly contextual, is also universal. Children make use of the strategies that they already know from human communication, but our study further demonstrates that having robots as social interaction partners introduces additional layers to the interaction. This makes this research, and similar future studies in this area, an important contribution not only to the design and evaluation of educational robots, but also for furthering our understanding of what it means for children to interact with and develop relationships with social robots.

DATA AVAILABILITY STATEMENT

The datasets generated for this study will not be made publicly available because video recordings contain personal identifiable information. Requests to access the datasets should be directed to Lena Pareto, lena.pareto@hv.se.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of CODEX—Rules and Guidelines for Research established by the Swedish Research Council. The protocol did not require ethical approval according to Swedish law as no sensitive personal information about the subjects was collected. All child subjects' parents/legal guardians gave written informed consent, and all child subjects gave written assent, as well as verbal assent at the start of each session, in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

SE and LP carried out the study and collected the empirical data, created the interaction transcripts, and coded the video data under the guidance of SS and SL. SS conducted the thematic analysis under the guidance of LP and SE, drafted the manuscript, and held the overall responsibility for bringing individual sections together. All authors contributed to the authoring of the final manuscript and contributed to the research design.

FUNDING

This work was partially supported by the Marcus and Amalia Wallenberg Foundation and was funded in part by the project START (Student Tutor and Robot Tutee). The authors are solely responsible for the content of this publication. It does not represent the opinion of the Wallenberg Foundation, and the Wallenberg Foundation is not responsible for any use that might be made of data appearing therein.

ACKNOWLEDGMENTS

We thank the children and their schools for participating in this study. We also wish to thank Wolmet Barendregt for her time in our research project and for her contributions to the design process.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frobt.2020.00046/full#supplementary-material

REFERENCES

Alves-Oliveira, P., Sequeira, P., Melo, F. S., Castellano, G., and Paiva, A. (2019). Empathic robot for group learning: a field study. *ACM Trans. Hum. Robot Interact.* 8, 1–34. doi: 10.1145/3300188

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: a review. *Sci. Robot.* 3:eaat5954. doi: 10.1126/scirobotics.aat5954

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qual. Res. Psychol.* 3, 77–101. doi: 10.1191/1478088706qp063oa

Campbell, J. L., Quincy, C., Osserman, J., and Pedersen, O. K. (2013). Coding in-depth semistructured interviews:problems of unitization and intercoder reliability and agreement. *Sociol. Methods Res.* 42, 294–320. doi: 10.1177/0049124113500475

Cross, E. S., Hortensius, R., and Wykowska, A. (2019). From social brains to social robots: applying neurocognitive insights to human-robot interaction. *Philos. Trans. R Soc. Lond. B Biol. Sci*. 374:20180024. doi: 10.1098/rstb.2018.0024

Davison, D. P., Wijnen, F. M., Charisi, V., van der Meij, J., Evers, V., and Reidsma, D. (2020). "Working with a social robot in school: a long-term real-world unsupervised deployment," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge: Association for Computing Machinery), 63–72. doi: 10.1145/3319502.3374803

Edwards, B. I., and Cheok, A. D. (2018). Why not robot teachers: artificial intelligence for addressing teacher shortage. *Appl. Artif. Intellig.* 32, 345–360. doi: 10.1080/08839514.2018.1464286

El-Hamamsy, L., Johal, W., Asselborn, T. L. C., Nasir, J., and Dillenbourg, P. (2019). "Learning by collaborative teaching: an engaging multi-party cowriter activity," in *Proceedings of the IEEE International Conference on Robot and Human Interactive Communication* (New Delhi: IEEE). doi: 10.1109/RO-MAN46459.2019.8956358

Giuliani, M., Mirnig, N., Stollnberger, G., Stadler, S., Buchner, R., and Tscheligi, M. (2015). Systematic analysis of video data from different human–robot interaction studies: a categorization of social signals during error situations. *Front. Psychol.* 6:931. doi: 10.3389/fpsyg.2015.00931

Goetz, J., Kiesler, S., and Powers, A. (2003). "Matching robot appearance and behavior to tasks to improve human-robot cooperation," in *Proceedings of the The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003 (Proceedings ROMAN 2003)* (Millbrae, CA: ROMAN), 55–60.

Honig, S., and Oron-Gilad, T. (2018). Understanding and resolving failures in human-robot interaction: literature review and model development. *Front. Psych.* 9:861. doi: 10.3389/fpsyg.2018.00861

Jones, A., and Castellano, G. (2018). Adaptive robotic tutors that support self-regulated learning: a longer-term investigation with primary school children. *Int. J. Soc. Robot.* 10, 357–370. doi: 10.1007/s12369-017-0458-z

Jordan, B., and Henderson, A. (1995). Interaction analysis: foundations and practice. *J. Learn. Sci.* 4, 39–103. doi: 10.1207/s15327809jls0401_2

Kory-Westlund, J. M., and Breazeal, C. (2019). A long-term study of young children's rapport, social emulation, and language learning with a peer-like robot playmate in preschool. *Front. Robot*. 6:81. doi: 10.3389/frobt.2019.00081

Lemaignan, S., Fink, J., Mondada, F., and Dillenbourg, P. (2015). "You're doing it wrong! studying unexpected behaviors in child-robot interaction," in *Proceedings* (Paris: Springer International Publishing), 390–400. doi: 10.1007/978-3-319-25554-5_39

Movellan, J. R., Tanaka, F., Fortenberry, B., and Aisaka, K. (2005). "The RUBI/QRIO project: origins, principles, and first steps," in *Proceedings of the 4th International Conference on Development and Learning* (Osaka: IEEE), 80–86. doi: 10.1109/DEVLRN.2005.1490948

Newton, D. P., and Newton, L. D. (2019). Humanoid robots as teachers and a proposed code of practice. *Front. Ed.* 4:125. doi: 10.3389/feduc.2019.00125

Obaid, M., Aylett, R., Barendregt, W., Basedow, C., Corrigan, L. J., Hall, L., et al. (2018). Endowing a robotic tutor with empathic qualities: design and pilot evaluation. *Int. J. Hum. Robot.* 15:18500251. doi: 10.1142/S0219843618500251

Pareto, L. (2014). A teachable agent game engaging primary school children to learn arithmetic concepts and reasoning. *Int. J. Artif. Intellig. Ed.* 24, 251–283. doi: 10.1007/s40593-014-0018-8

Pareto, L. (2017). "Robot as tutee," in *Robotics in Education,* eds M. Merdan, W. Lepuschitz, G. Koppensteiner, and R. Balogh (Cham: Springer International Publishing), 271–277. doi: 10.1007/978-3-319-42975-5_24

Pareto, L., Ekström, S., Barendregt, W., Serholt, S., and Kiesewetter, S. (2019). "Augmenting game-based learning with a robot tutee," in *Proceedings of the 13th European Conference on Games Based Learning (ECGBL)* (Odense), 560–568.

Rauchbauer, B., Nazarian, B., Bourhis, M., Ochs, M., Prévot, L., and Chaminade, T. (2019). Brain activity during reciprocal social interaction investigated using conversational robots as control condition. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 374:20180033. doi: 10.1098/rstb.2018.0033

Ros, R., Nalin, M., Wood, R., Baxter, P., Looije, R., Demiris, Y., et al. (2011). "Child-robot interaction in the wild: advice to the aspiring experimenter," in *Proceedings of the 13th International Conference on Multimodal Interfaces* (Alicante: ACM), 335–342. doi: 10.1145/2070481.2070545

Selwyn, N. (2019). *Should Robots Replace Teachers?* Cambridge: Polity Press.

Serholt, S. (2018). Breakdowns in children's interactions with a robotic tutor: a longitudinal study. *Comp. Hum. Behav.* 81, 250–264. doi: 10.1016/j.chb.2017.12.030

Serholt, S. (2019). "Interactions with an empathic robot tutor in education: students' perceptions three years later," in *Artificial Intelligence and Inclusive Education: Speculative Futures and Emerging Practices,* eds J. Knox, Y. Wang, and M. Gallagher (Singapore: Springer), 77–99. doi: 10.1007/978-981-13-8161-4_5

Serholt, S., and Barendregt, W. (2016). "Robots tutoring children: longitudinal evaluation of social engagement in child-robot interaction," in *Proceedings of the 9th Nordic Conference on Human-Computer Interaction (NordiCHI'16 ACM)* (Gothenburg: ACM). doi: 10.1145/2971485.2971536

Serholt, S., Barendregt, W., Vasalou, A., Alves-Oliveira, P., Jones, A., Petisca, S., et al. (2017). The case of classroom robots: teachers' deliberations on the ethical tensions. *AI Soc.* 32, 613–631. doi: 10.1007/s00146-016-0667-2

Sharkey, A. (2016). Should we welcome robot teachers? *Ethics Inform. Technol.* 18, 283–297. doi: 10.1007/s10676-016-9387-z

Uchida, T., Minato, T., Koyama, T., and Ishiguro, H. (2019). Who is responsible for a dialogue breakdown? an error recovery strategy that promotes cooperative intentions from humans by mutual attribution of responsibility in human-robot dialogues. *Front. Robot.* 6:29. doi: 10.3389/frobt.2019.00029

Vogt, P., van den Berghe, R., de Haas, M., Hoffmann, L., Kanero, J., Mamus, E., et al. (2019). "Second language tutoring using social robots. a large-scale study," in *Proceedings of the IEEE/ACM Int. Conference on Human-Robot Interaction (HRI 2019)* (Daegu: IEEE). doi: 10.1109/HRI.2019.8673077

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 26 June 2020
doi: 10.3389/frobt.2020.00078

Educational Robotics to Foster and Assess Social Relations in Students' Groups

Michela Ponticorvo[1]*, Franco Rubinacci[1], Davide Marocco[1], Federica Truglio[1] and Orazio Miglino[1,2]

[1] Natural and Artificial Cognition Lab, Department of Humanistic Studies, University of Naples "Federico II", Naples, Italy,
[2] Institute of Cognitive Science and Technology, National Research Council, Rome, Italy

Robotics has gained, in recent years, a significant role in educational processes that take place in formal, non-formal, and informal contexts, mainly in the subjects related to STEM (science, technology, engineering, and mathematics). Indeed, educational robotics (ER) can be fruitfully applied also to soft skills, as it allows promoting social links between students, if it is proposed as a group activity. Working in a group to solve a problem or to accomplish a task in the robotics field allows fostering new relations and overcoming the constraints of the established links associated to the school context. Together with this aspect, ER offers an environment where it is possible to assess group dynamics by means of sociometric tools. In this paper, we will describe an example of how ER can be used to foster and assess social relations in students' group. In particular, we report a study that compares: (1) a laboratory with robots, (2) a laboratory with Scratch for coding, and (3) a control group. This study involved Italian students attending middle school. As the focus of this experiment was to study relations in students' group, we used the sociometric tools proposed by Moreno. Results show that involving students in a robotics lab can effectively foster relations between students and, jointly with sociometric tools, can be employed to portrait group dynamics in a synthetic and manageable way.

Keywords: educational robotics, sociometric tools, social networks, assessment, students' groups, coding

OPEN ACCESS

Edited by:
Giovanni Pilato,
Institute for High Performance Computing and Networking (ICAR), Italy

Reviewed by:
Robertas Damasevicius,
Silesian University of Technology, Poland
Soheil Keshmiri,
Advanced Telecommunications Research Institute International (ATR), Japan

****Correspondence:***
Michela Ponticorvo
michela.ponticorvo@unina.it

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *29 February 2020*
Accepted: *07 May 2020*
Published: *26 June 2020*

Citation:
Ponticorvo M, Rubinacci F, Marocco D, Truglio F and Miglino O (2020) Educational Robotics to Foster and Assess Social Relations in Students' Groups. Front. Robot. AI 7:78. doi: 10.3389/frobt.2020.00078

INTRODUCTION

During the last decades, different activities have found their own space along with curricular ones in schools. Between these, educational robotics (ER) can be an effective teaching and learning tool (Miglino et al., 1999) as it allows for transferring knowledge such as mathematics, computer science, and physics (Lindh and Holgersson, 2007; Williams et al., 2007; Nugent et al., 2009) and allows one to train skills, including thinking skills and problem solving approaches (Hussain et al., 2006; Sullivan, 2008; Mikropoulos and Bellou, 2013; Atmatzidou and Demetriadis, 2016; Gabriele et al., 2017).

An interesting review dating back to 2012 by Benitti (Benitti, 2012) reports that the use of robots in school has positive outcomes for teaching concepts that are connected to STEAM areas (STEM plus arts), as it can have an impact on education in the fields of science, technology, and mathematics along the educational process starting with preschool up to higher education including university (Javidi and Sheybani, 2010; Alimisis, 2013; Chung et al., 2014; Eguchi, 2014a).

Along the years, ER has been widely introduced in school activities and has consolidated its presence, especially in classrooms of high schools.

ER implies an integrated approach to complement different areas and fields, enhancing interest, and curiosity in scientific issues (Arís and Orcos, 2019).

For today's society, mastering technology is fundamental and ER can be used to introduce technology and promote other skills. In fact, in parallel with STEAM-related issues, ER allows to promote skills like initiative, autonomy, teamwork, and creativity (Sica et al., 2019a), the so-called 21st century skills (Eguchi, 2014b), complex, and evolutionary systems management (Miglino et al., 2004; Whittier and Robinson, 2007; Rubinacci et al., 2017a), together with social skills and communication (Owens et al., 2008).

A relevant study by Kandlhofer and Steinbauer (2016) shows that ER leads to a better achievement in social skills and self-esteem in students that results in increased motivation (Bazylev et al., 2014), which is a pivotal element in enhancing learning.

On the educational science side, ER is based on the constructionist approach, where the students are at the center of the learning challenge because they are active agents who can determine their learning processes (Piaget, 1974; Papert, 1980; Papert and Harel, 1991).

This means that, during ER activities, learners build their own pathways to understand the world around them; they discover, they use information to creatively get more knowledge, and they participate actively in the educational challenges, guided by teachers (Sica et al., 2019b).

Moving from the individual to the group level, it is interesting to underline that most of ER activities must be run in groups, thus promoting collaborative work and collaborative learning (Denis and Hubert, 2001). Collaborative learning in ER has been examined by a certain number of studies, showing how it can contribute to foster social ties in groups of students at different ages. The very recent study by Gonnot et al. (2019) analyzes the use of social robots in a context of collaborative learning, investigating how adding a social dimension to robot can improve learning. Some other studies were devoted to understand if social robots could affect the collaboration between children at play (Strohkorb et al., 2016) and to propose a framework for robots as mediator tools (Mitnik et al., 2008).

Robots can be the core element of an educational framework for collaborative learning, if they are conceived as components of Internet-of-Things (Plauska and Damaševičius, 2014), and, thanks to their features that promote collaborative learning, they can be used adopting a constructivist approach, as said before, which is highly motivating for children and adolescents.

The study by Atmatzidou and Demetriadis (2012) deepens the reflection on the pedagogical approaches for ER in the school context, which is a high-impacting issue. They explore different collaboration scripts used as a guide in students' group work during the ER activity.

ROBOTICS AND GROUP DYNAMICS

Summarizing what literature taught, ER can be useful to promote the following: knowledge related to STEAM and skills such as computational thinking, problem solving, complex systems management, and collaborative learning, "inside the students," which means that the focus is on the personal side.

In the present study, we propose to change the focus to what happens "between the students" who are involved in ER activities, which means that we concentrate on the social side.

We believe that ER can be used to foster positive and collaborative relations between students and, at the same time, provide a context to assess the changing networks in the classroom (Rubinacci et al., 2017b; Truglio et al., 2018a). In particular, these recent studies proposed by the authors of the present paper indicate how ER can be exploited to favor positive ties and connections between students.

Now we make a step forward to verify this claim and to show that the use of sociometric tools in the context of ER can picture the classroom environment in critical moments that affect students' career and classroom climate (Truglio et al., 2018b). A low social inclusion at school can have a dramatic effect on relevant phenomena including school dropout (Frostad et al., 2015; Ricard and Pelletier, 2016), and the sociometric framework offers sensitive tools to observe micro and macro dynamics elicited by ER activities.

This happens because ER allows one to establish a bridge between students, who become interdependent as they are required to reach a shared goal (Burbaite et al., 2013; Kamga et al., 2016), to coordinate themselves, to learn to divide tasks in subtasks, and to complete them, taking into account other group members (in terms of opinions, ideas, skills, and abilities). As a consequence, also those students who are not well-included in the class have the opportunity to be involved in group activity and to improve relationships with other students.

In this paper, we would like to show how ER is indeed an adequate and useful framework to assess social relations and support positive connections among students in the peer group. In particular, our research hypothesis is that ER can be more effective in promoting positive ties and connections between students if compared with other activities. At the same time, our goal is to verify if the use of sociometric tools offers a valid framework to evaluate these ties. To address these issues, we have worked on a 2-month project (from September to November 2017). The trial took place in Naples and its surroundings, an area in Southern Italy, which is highly affected by school dropout resulting in threats at the social level (O'Higgins et al., 2007). In the section Robotics to Foster and Assess Social Relations in Students' Groups, we will describe this project in more detail including the results we obtained, and in the section Discussion and Conclusions, we will discuss these results and their implications.

ROBOTICS TO FOSTER AND ASSESS SOCIAL RELATIONS IN STUDENTS' GROUPS

In this section, we describe the study we have run in a secondary school in Italy. The proposed research project aims to assess whether the ER laboratory, through group activities, is an effective method to assess, and promote social relationships within a peer group in the class. To test our research hypothesis, we considered three groups and two activities: the ER laboratory and the coding laboratory with Scratch. The third group performed individual activities that were not intended to stimulate interactions between students. It was thus possible to picture the group dynamics at the beginning of the school year and the effect of the different activities on them.

Materials and Methods

Participants

The study involved 70 participants attending the first-year of middle school ("Scuole medie" in the Italian school system), aged between 10 and 11 years. Thirty-eight participants were females and 32 males; their mean age was 10.48 years.

We decided to focus on first-year students as there are weak ties between them, especially at the beginning of the school year.

For school needs, each group was randomly assigned to a condition of the experimental design. From discussion with school referents, we were assured that classes were composed so as to be homogenous in terms of grades from primary school, gender balance, and social skills.

In more details:

1. Group 1, formed by 23 students, carried out the ER laboratory.
2. Group 2, formed by 24 students, performed the coding laboratory with Scratch.
3. Group 3, composed of 23 pupils, was not involved in any group activity.

Group 3 was the control group and allowed to obtain the baseline to compare group activities about the effects on social relations, as the time lapse between the start and the end of group activities may anyway have an effect on links and relationships between peers.

The Tools for Group Activities: Lego Mindstorms NXT and Scratch

The robotics technology we used in the present project was Lego Mindstorms NXT (Klassner and Anderson, 2003). This robotics kit includes both a hardware side and a software side (NXT-G).

In the coding lab, we used Scratch (Maloney et al., 2010): it is a programming language that is freely available and is commonly used to approach children, kids, and teen students to coding as it offers the opportunity to create multimedia and interactive games simply and intuitively with images, music, and sounds. Together with coding, related to computational thinking, this software helps students to develop their skills related to creativity, systematic reasoning, and problem solving.

Sociometric Test

To assess social relations, the sociometric test of Moreno was administered to the students of the three groups, before and after the laboratory activities. The sociometric test allows one to effectively investigate interpersonal relationships inside the peer group and to highlight the status of the group components in terms of inclusion. Indeed, sociometry is a methodology that was proposed by Jacob L. Moreno in order to study the structure and interactions of people within a group (Moreno, 1941, 1951), and it has been employed in many different contexts including family therapy and educational contexts. If we consider educational contexts, the sociometric methodology can be useful to examine situations where there are conflicts among students, isolated subjects, lack of cooperation in working groups, etc. In the present project, it was the selected tool to picture and study the links between peers in the three groups involved in the activities.

The sociometric test proposed to the students concerned the criterion that is called affective-relational perspective. This perspective is related to the emotional aspect of a relationship and reflects students' affinities.

The criterion is operationalized in two sentences, which allow to highlight preferences and, conversely, rejections toward members of the group. These sentences ask to indicate the classmates who the responding participants would (or would not) want as roommates during a school trip.

Then the first step is to report data in a double-entry table named sociomatrix. In this table, on the axes of abscissas and ordinates, there are the names of the group members: horizontally we report the expressed choices (or rejections) and vertically the received choices (or rejections). The choices are indicated with "1" and the rejections with "−1."

Let us consider an example of a very small group of children, composed by A, B, C, and D.

We will then have a square matrix with four elements on the axes of abscissas and ordinates. If A chooses B, we will put a $+1$ at the intersection between A (horizontally) and B (vertically). If C rejects D, we will put a -1 at the intersection between C (horizontally), and D (vertically).

These sociometric data can be represented in the graphical form called sociogram too. It is a network graph with nodes and lines. Nodes represent the students, the components of a group, whereas lines are the links, the relations (different kinds of lines distinguish choices and rejections). Furthermore, each line has one or two arrows showing the direction of the relationship and if it is unidirectional or bidirectional. In the sociogram, A, B, C, and D will be the nodes, pictured as circles, and they will be connected by lines with different graphical features, corresponding to a different kind of relation (see the **Figures 1–12** in the section Discussion and Conclusions).

From the sociogram and the sociomatrix, it is possible to delineate the following: the total choices and rejections that each member of the group has received; the degree of reciprocity of choices and rejections; and the difference between ignored, rejected, isolated, and popular subjects.

The popular subjects are those who have received a large number of choices, so they are those who have greater influence and greater power within the group. The rejected subjects, on the other hand, are those who have received a large number of rejections. Finally, the isolated subjects are those who have received very low number of choices. This last category includes:

1. subjects who are ignored by the group, but who prove to be open and available to others by expressing their choices and
2. subjects who are ignored and tend to self-isolate by expressing neither choices nor rejections.

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
a				-1			-1	1	-1										1			1	
b			1	-1	1	1	-1				-1	-1	1				1	-1		1			
c		1		-1			-1		1		1		1		1		1	-1					
d			1				1		1	-1	1		1		1			1			1		
e		1	-1	-1		1			-1		-1		-1			1	1			1			
f		1			1		-1		-1					1		1	1		1	1		1	1
g		1	1	1					-1		1	1		-1							1		
h	1						-1		-1	1		-1		1									1
i			1	1	-1	-1	-1				1		1		1		1	-1					
j				-1	1		-1	1	-1		-1			1		1		-1					
k			1	1			-1		1				1					-1		1			
l		-1				1	-1		-1		-1			1	-1				1	-1		1	1
m		1	1	1			-1		1						1			-1		1			
n				-1	1	1	-1		-1	1						1	1	-1				1	1
o			1				-1		1		1		1					-1			1		
p					1	1	-1		-1					1					1			1	
q		1		-1	1	1	-1				-1		1	1		1		-1		1			
r				1			1			-1											-1		
s				-1	1	1	-1				1			1		1	1	-1					-1
t		1			1				-1				1				1	-1			1		1
u		-1	1		-1			-1	1			1											1
v					1	1				-1				1		1							
w						1	-1		-1	1		1		1				-1	1		1	1	

FIGURE 1 | Sociomatrix built on the *Educational Robotics group* for the affective criterion at the beginning of the scholastic year.

Along with the sociogram and the sociomatrix, it is possible to use statistical techniques on the indexes that are derived from the sociometric tools. The sociometric tool provides a rich amount of data on group interaction and dynamics.

Procedure

As hinted at previously, the three conditions to verify the effects of different activities on interpersonal relations in the peer group are the lab with robotics activities and the lab with Scratch about coding, together with the control group. The groups have been randomly assigned to one of the three experimental conditions (ER lab, coding lab with Scratch, and no group activities).

The sociometric test was proposed to the participants (the students belonging to the three experimental groups) in two moments, on September 25 (i.e., before the beginning of lab activities: pretest) and on November 29 (i.e., at the end of lab activities: post-test).

The activities covered 6 weeks, with a meeting for a week and each lasting 1 or 2 h (for a total of 10 h). To carry out the activities, students were divided in subgroups that were composed by different students at each meeting. In the next subsection, laboratory activities are described in more detail.

The Laboratory Activities

During the laboratory activities, which were scheduled as 6 weekly meetings lasting 1 or 2 h, participants followed different pathways.

In summary, in the robotics lab, the following activities were carried out: realization of posters dealing with technology, robot

	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w
a		1	1	-1	1	1	-1	1	-1	1	-1	1	1	1	1	1	1	-1	1	1	1	1	1
b	-1		1	1	1	1	-1	-1	-1	-1		-1	1	1	1	-1	1	-1	1	1	1		1
c	1	1								1			1	1	1		1			1	1		
d			1				1		1	-1	1	-1	1		1			1			1		
e				-1		1	-1		-1					1		1	1	-1	1	1		1	1
f	1	1	1	-1	1		-1	1	-1	1		1	1	1	1	1	1	-1	1	1	1	1	1
g			1	1					1		1		-1		-1			1			1		
h	1	1	1	-1	1	1	-1		-1	1	-1	1	1			1	1	-1	1	1		1	1
i	-1		1		-1	-1		-1		-1		-1	1	-1	1	-1		-1	-1			-1	-1
j	1	1	1		1	1	-1	1	-1		1	1	1	1	1	1	1	-1	1	1	1	1	1
k		1	1				-1		-1				1		1		-1	-1		1			
l	1			-1	1	1	-1	1	-1	1	-1			1		1		-1	1				1
m	1	1	1	1	1	1	-1	1	1	1	1	1		1	1	1	1	1	1	1	1	1	1
n	1	1	1	-1	1	1	-1	1	-1	1	1	1	1		1	1		-1	1	1		1	1
o		1	1				-1		-1				1				1	-1		1			
p	1			-1	1	1	-1		-1					1					1			1	
q	1	1	1	-1	1	1	-1		-1		-1			1		1		-1		1	1		
r				1			1				1		1								1		
s	1	1	1	-1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1		1	1	1	1
t		1		-1			-1		-1				1				1				1		
u					1		1		-1					-1	1		-1					1	
v	1	1			1	1	-1		-1	1		1	1	1		1	1	-1		1			1
w	1			-1	1	1	-1	1	-1	1		1		1		1			1		1	1	

FIGURE 2 | Sociomatrix built on the *Educational Robotics group* for the affective criterion at the end of the robotics laboratory.

building and programming, and building of road itineraries representing the environment where the robot moved.

For the coding lab with Scratch, the students were involved in the following: realization of posters regarding the topic of technology, creation of a *sprite* (an element of Scratch programming environment, which can be conceived as an agent; see Ponticorvo et al., 2017), creation of the *stage* (the place where sprites interact), coding of *sprite* behavior in a spatial labyrinth, and building of multimedia road itineraries representing the environment where the *sprite* moved. These activities were conceived in order to make comparable the tasks with the students attending the robotics lab and the coding lab.

What is different is that in the robotics lab, participants used tangible materials, so as to build the robot and to realize road itineraries, whereas those in the coding lab have carried out their activities exclusively with software, then in a digital environment.

Previous work conducted by our research group indicates that this element can be relevant in promoting different cognitive and social processes (Di Fuccio et al., 2015; Ferrara et al., 2016).

In more detail, the ER lab's schedule was the following: during the first meeting, by a frontal and interactive lesson, the researcher talked about technology and introduced the definition of a robot as an artifact with a sensory-motor system. At the end of this first meeting, participants are divided in five subgroups to build autonomously and collaboratively a poster about technology and robots. During the second meeting, students were divided again in five subgroups, different from the previous ones. Every subgroup built a robot using the tools

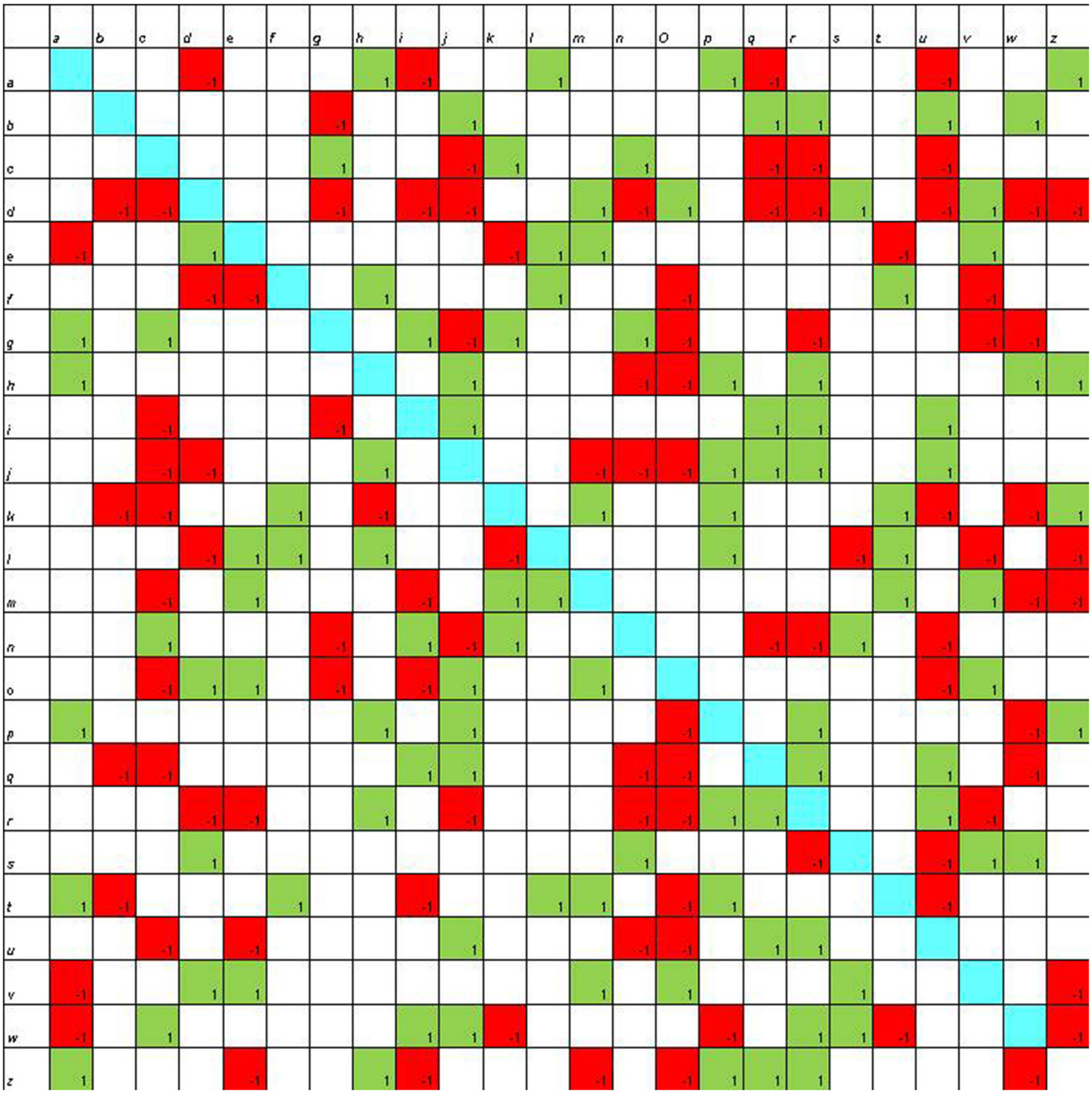


FIGURE 3 | Sociomatrix built on the *coding with Scratch group* for the affective criterion at the beginning of the scholastic year.

described above. Students had to collaborate and work in groups to reach a common goal.

In the third meeting, the software to program the robot was introduced and students used it to implement the robot control system. Participants were again divided in subgroups and worked together to build their strategy for the robots. In the fourth and fifth meetings, they built the street pathways for the robot taking inspiration from their own city and elaborating them in a creative way. The sixth meeting was devoted to writing the code and transferring it in the robot to follow the street pathway, always working in subgroups.

The coding laboratory with *Scratch* was structured in a very similar way: in the first meeting by a frontal and interactive lesson, the researcher talked about technology, and introduced the software *Scratch* for programming. At the end of the meeting, students were divided into five subgroups to build autonomously and collaboratively a poster about technology and *Scratch*. In the second meeting, run in the computer classroom, *Scratch* was introduced in its basic functionalities. Later participants were divided in subgroups and had realized together some elements in *Scratch*.

In the third meeting, how to program the elements in *Scratch* had been shown and then they were divided in subgroups to decide their strategy to program to follow a spatial labyrinth.

During the fourth and fifth meetings, students, divided in subgroups, implemented multimedia, taking inspiration from their city. In the last meeting, a street pathway had been implemented and new subgroups had been formed to write the code and the sequences to follow the multimedia street pathway.

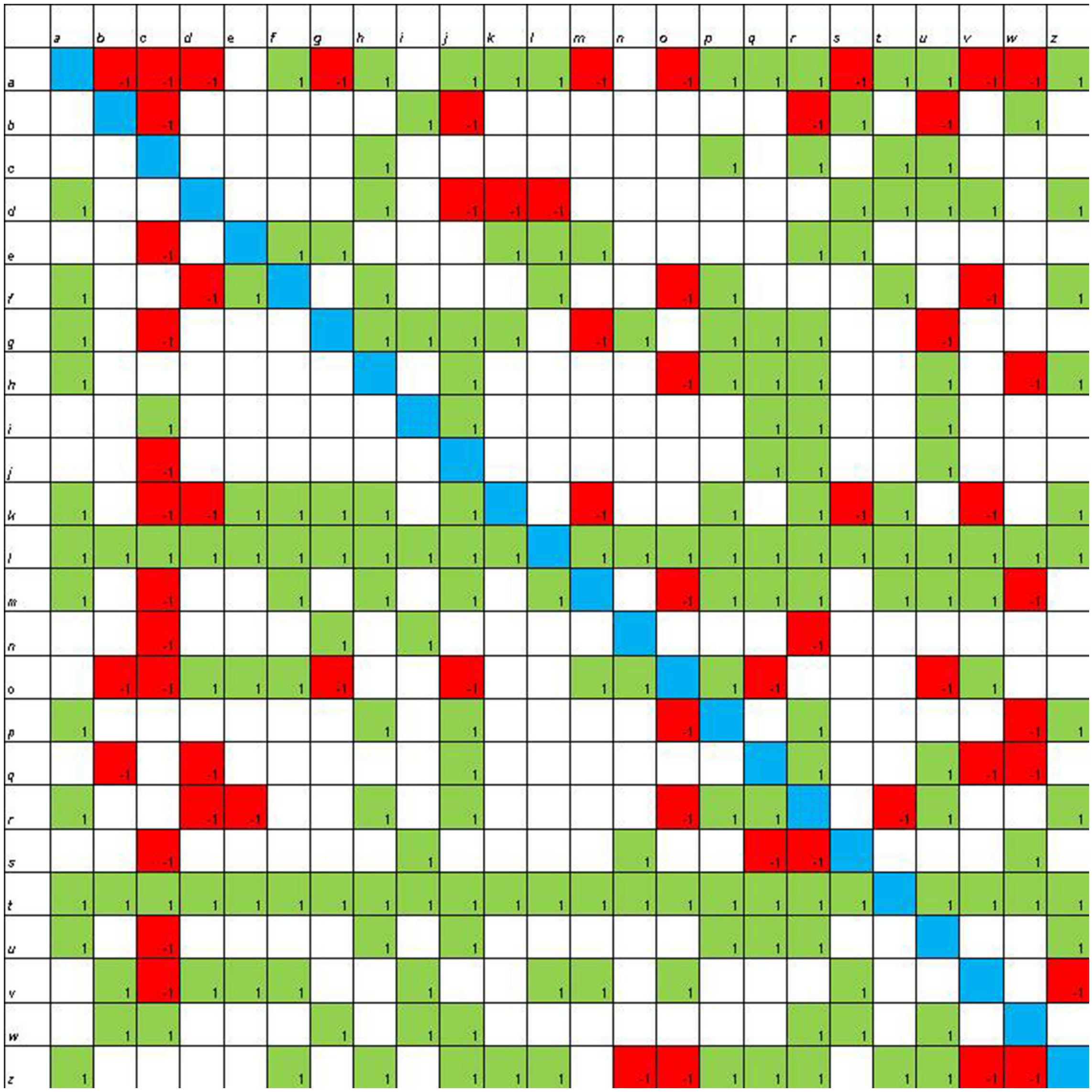


FIGURE 4 | Sociomatrix built on the *coding with Scratch group* for the affective criterion at the end of laboratory activities.

Results

In this section, we report sociograms and sociomatrices for each condition at the beginning and at the end of the intervention, we compare the indexes for the three conditions, and then we confront the number of selections and rejections at the beginning and at the end of the project using *t*-test.

Sociogram and Sociomatrix Analysis

Sociograms and sociomatrices were built on the three groups both for pretest and post-test. Group 1 carried out the ER laboratory, Group 2 performed the coding laboratory with Scratch, and Group 3 was the control group. Here we report sociograms and sociomatrices about the affective criterion.

In the sociomatrices, to delineate the choices of the members of the group, the number 1 was used inside a green box, and to indicate the rejection, the −1 was used in a red box. Furthermore, the total choices and the total rejections (both expressed and received) were recorded for every student in the peer group.

At the beginning of the school year (**Figure 1**), it is possible to observe that there are six students who are able to attract a good number of choices (8 and 9, the highest values) and two students who receive more than 10 rejections.

After the laboratory activities (**Figure 2**), it is evident that peers make much more selections and more students receive a high number of choices (16 students receive 10 or more choices). Also the rejections increase and four students receive more than 10 rejections.

In the group that was involved in the coding activities, at the beginning of the school year, there are three students who attract 8–9 choices and one student who receives 10 rejections (**Figure 3**).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
A			-1					1				1	-1				1	1	-1			1	
B			-1		1			1	-1		-1								-1				
C		-1		-1		1		1			1		-1	1	-1				1				-1
D			-1						-1	-1	-1		1	1	1		1			1			
E		1				1			-1	-1			1	1		-1	-1					-1	1
F		1	1		1				-1				-1	1	-1				-1	-1		1	1
G			-1	1	1				-1	-1			-1		1					1			
H	1	1							-1		-1			1							-1		
I				-1	-1																	1	
J	1			-1					1			1		-1			1	1	1				
K		-1	1										-1	1							-1		
L	1								-1	-1						1	1	1			-1	1	
M	1		-1	1			1			-1	-1			1	1	-1	-1						
N			-1	1	1	1		1	-1	-1			1		1		-1						
O				1					-1	-1			1	1		1	-1						
P				1		-1			-1	1	-1	1			1		1		-1				
Q	1								-1	1		1				1		1			-1	-1	
R	1	-1				-1						1				-1	1					1	-1
S		-1	1					1			1		-1								-1		
T	1	-1					1	1	-1			1	1	1							-1	-1	
U	1		-1				1	-1		-1	-1	-1	-1					1					
V			-1	-1	1	1		1	1	-1		1			-1	-1							
W			-1		1	1							1						-1				

FIGURE 5 | Sociomatrix built on the control group for the affective criterion at the beginning of the scholastic year.

Also in this case, the number of choices increases (**Figure 4**): seven students receive 10 or more selections and only one student receives more than 10 rejections.

In the control group, at the first assessment (**Figure 5**), there are two participants who receive 8–9 selections and one participant who collects more than 10 rejections.

At the end of the project (**Figure 6**), the number of choices increases and three participants receive more than 10 rejections.

The sociomatrix represents the basis for other analysis and allows one to have a relevant number of information in a synthetic way.

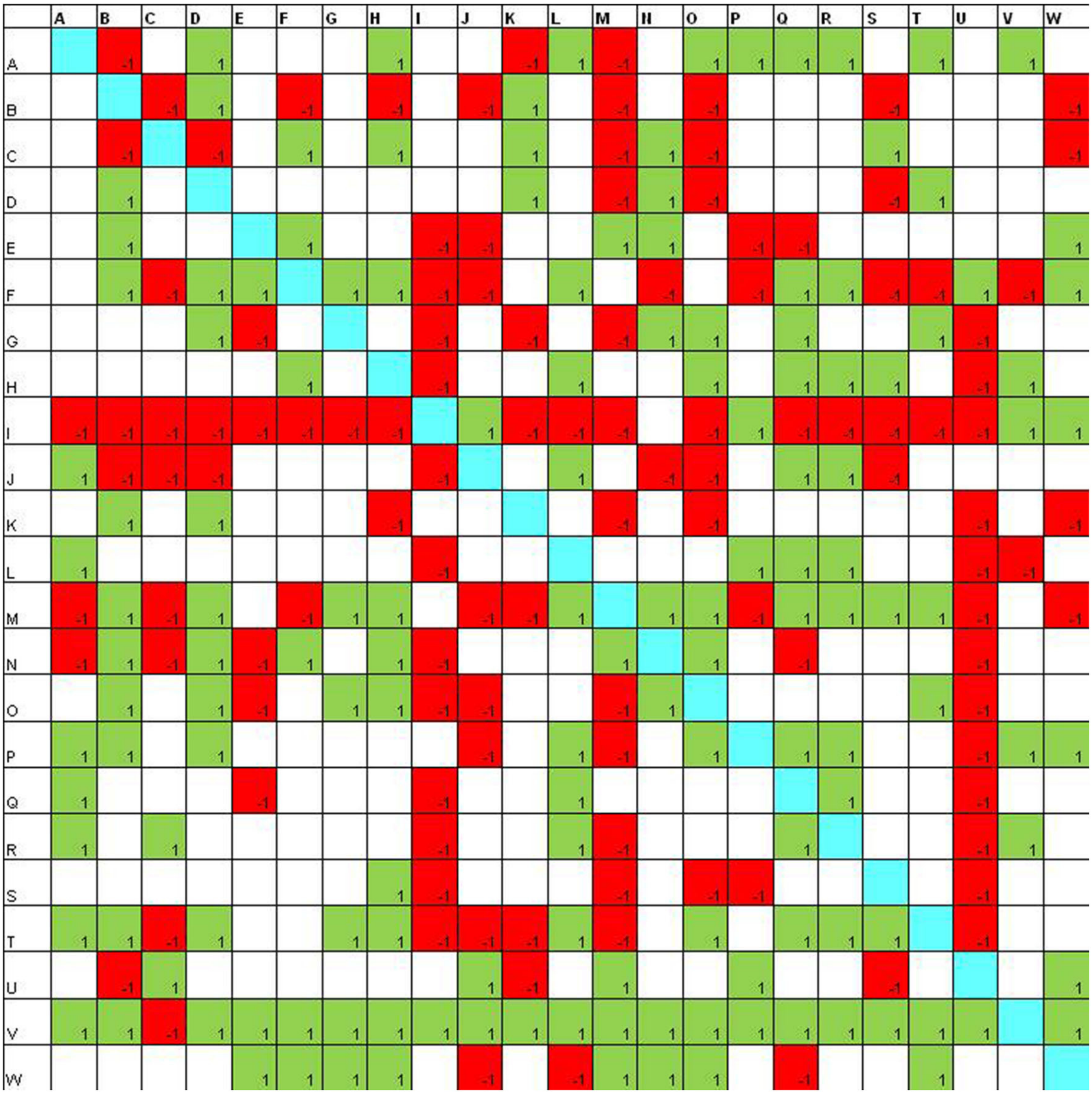
Starting from the sociomatrices, we built the sociograms and calculated various indexes, as described by Garcia-Magarino et al. (2019), with the software Gephi, an open-source software package for analysis and visualization of social networks (Bastian et al., 2009).

Here we report the sociograms for the three experimental groups at the pretest and post-test, considering the total one, i.e., the one that considers both selections and rejections and then, separately, the selections and the rejections.

The qualitative comparison of the sociograms at these two moments shows some interesting dynamics. In the robotics group (**Figures 7**, **8**), the network becomes more connected: in particular, the selection one shows much more links. Considering the node in the rejections groups, at post-test, there is only one node that receives a high number of rejections.

For the coding group (**Figures 9**, **10**), we observe that there are much more choices, especially selections as rejections link decreases.

If we consider rejects, in the control condition, rejects increase significantly, whereas in the robotics condition, the number of rejects remains essentially the same.



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
A		-1		1				1			-1	1	-1		1	1	1	1		1		1	
B			-1	1		-1		-1		-1	1		-1		-1				-1				-1
C		-1		-1		1		1			1		-1	1	-1				1				-1
D		1									1		-1	1	-1				-1	1			
E		1				1			-1	-1			1	1		-1	-1						1
F		1	-1	1	1		1	1	-1	-1		1		-1		-1	1	1	-1	-1	1	-1	1
G				1	-1				-1		-1		-1	1	1		1			1	-1		
H						1			-1			1			1		1	1	1		-1	1	
I	-1	-1	-1	-1	-1	-1	-1	-1		1	-1	-1	-1		-1	1	-1	-1	-1	-1	-1	1	1
J	1	-1	-1	-1					-1			1		-1	-1		1	1	-1				
K		1		1				-1					-1		-1						-1		-1
L	1								-1							1	1	1			-1	-1	
M	-1	1	-1	1		-1	1	1		-1	-1	1		1	1	-1	1	1	1	1	-1		-1
N	-1	1	-1	1	-1	1		1	-1				1		1		-1				-1		
O		1		1	-1		1	1	-1	-1			-1	1						1	-1		
P	1	1		1						-1		1	-1		1		1	1			-1	1	1
Q	1				-1				-1			1						1			-1		
R	1		1						-1			1	-1				1				-1	1	
S								1	-1				-1		-1	-1					-1		
T	1	1	-1	1			1	1	-1	-1	-1	1	-1		1		1	1	1		-1		
U		-1	1							1	-1		1			1			-1				1
V	1	1	-1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1		1
W					1	1	1	1		-1		-1	1	1	1		-1			1			

FIGURE 6 | Sociomatrix built on the control group for the affective criterion at the end of laboratory activities.

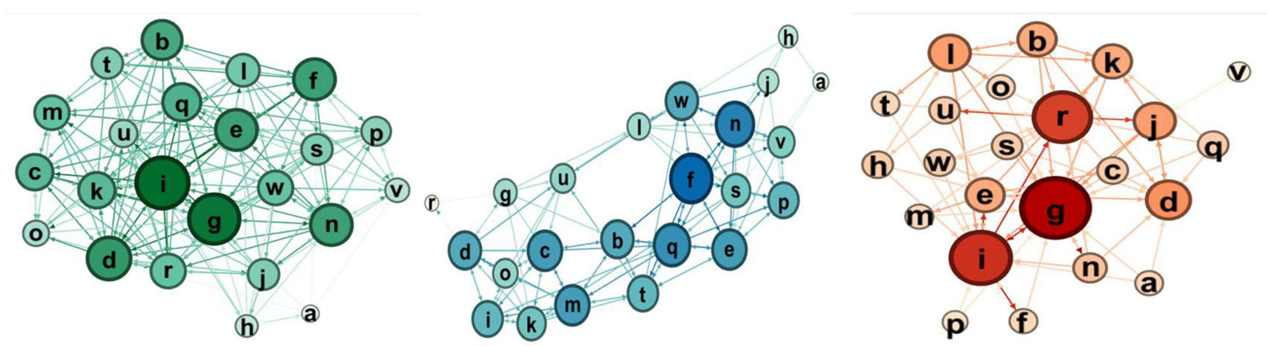


FIGURE 7 | Sociograms representation of *Educational Robotics group* at the pretest. In green the total one, in blue the selection one, and in red the rejection one.

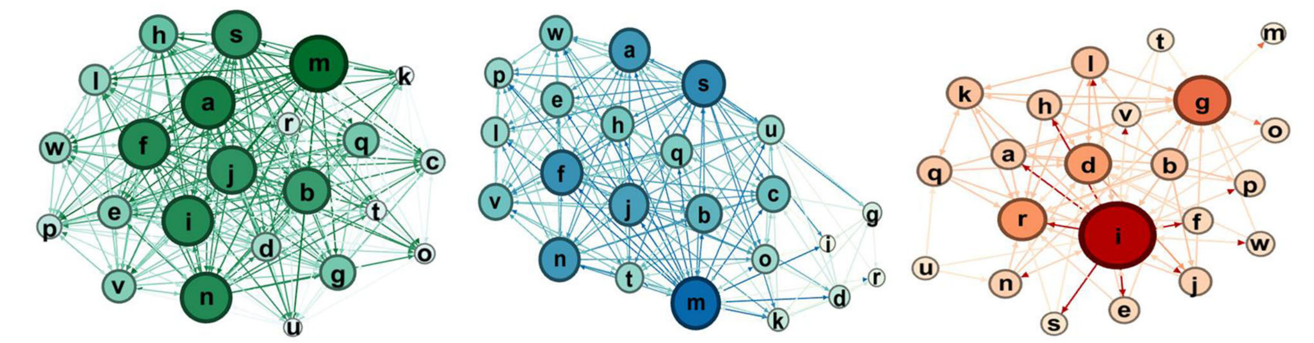


FIGURE 8 | Sociograms representation of *Educational Robotics group* at the post-test. In green the total one, in blue the selection one, and in red the rejection one.

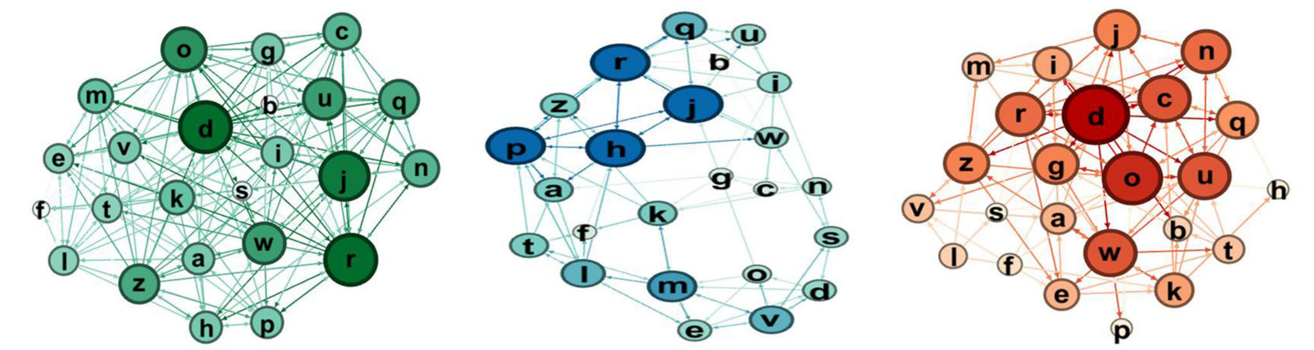


FIGURE 9 | Sociograms representation of the coding group at the pretest. In green the total one, in blue the selection one, and in red the rejection one.

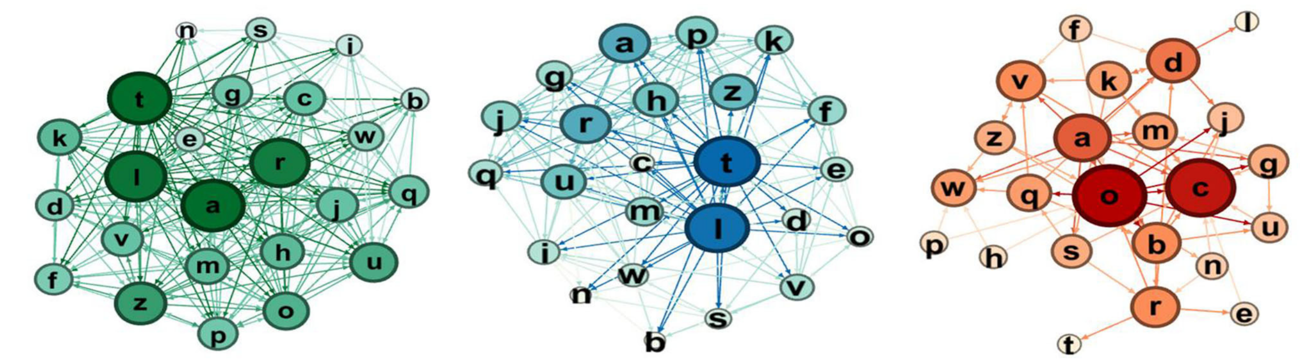


FIGURE 10 | Sociograms representation of the coding group at the post-test. In green the total one, in blue the selection one, and in red the rejection one.

For the control group (**Figures 11**, **12**), the comparison between the pretest and the post-test indicates that the group has more links, as we expected for the time lapse, but it is interesting to notice that the rejection links increase.

In **Table 1**, the analysis run with the Gephi software is reported at the pretest and the post-test. The average corresponds to the ratio between connections (edges) and the number of participants (nodes). Here, the Social Intensity, Cohesion, Dissociation, and Coherence indexes are also reported (Garcia-Magarino et al., 2019).

Social Intensity Index measures the percentage of relations (reciprocal or not) on the number of theoretically possible combinations. It indicates how the students are connected, either positively, or negatively. Usually, a high value of the index means that students know each other well.

Cohesion Index is the ratio between reciprocal relations and possible relations. Cohesion is useful to understand if students rely on the others in the group. It is the level of reciprocal acceptation between students and can highlight popular students.

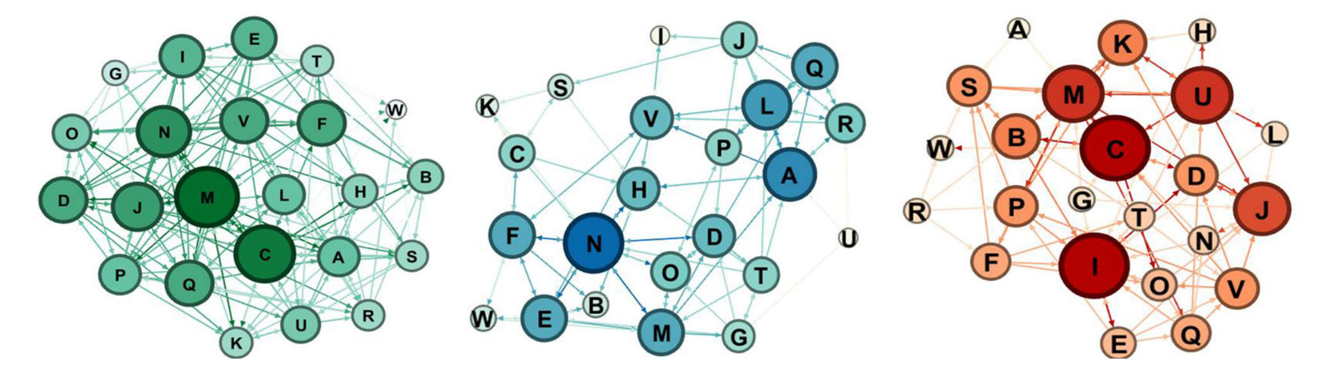


FIGURE 11 | Sociograms representation of the control group at the pretest. In green the total one, in blue the selection one, and in red the rejection one.

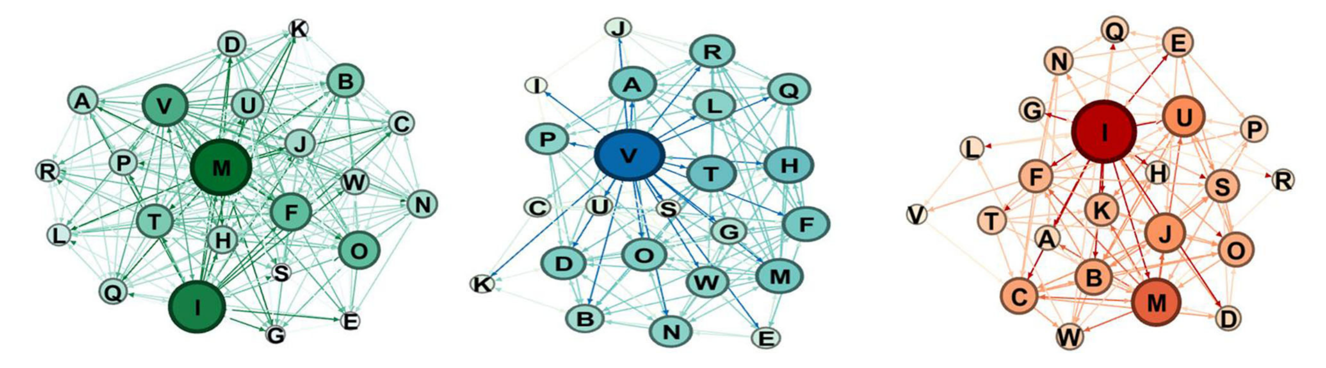


FIGURE 12 | Sociograms representation of the control group at the post-test. In green the total one, in blue the selection one, and in red the rejection one.

Dissociation Index represents the opposite of previous metrics because it is centered on the ratio between reciprocal rejects and the number of possible combinations. This index shows the average ratio of reciprocal rejects and if there are unpopular students.

Coherence refers to the ratio between reciprocal selections and selections received by other students. In other words, it represents the reciprocity in students' selection. It is useful to highlight if students tend to have reciprocal relations.

These indexes vary between 0 and 1.

Table 1 summarizes the indexes for the experimental group at the pretest and the post-test.

Coherently with what we have observed by the sociograms, the indexes get better between the pretest and the post-test. There is a notable increase in Social Intensity and Selection indexes and a low increase in Rejection for the robotics lab.

These analyses indicate that the robotics lab can be effective in promoting dynamics that can lead to a modification of the status of each participant at a personal level and of the group as a dynamic entity.

In the coding laboratory, there is a little increase in Social Intensity and Selection indexes and a little decrease in Rejection: this indicates that the network has changed slightly. In the control group, all indexes increase a little, as expected because of the interaction related to school.

In the three experimental conditions, the indexes show an increase between pretest and post-test selections.

To better understand the effects produced by the robotics lab in comparison with the coding activity, we run the statistical analyses whose results are reported in the next section.

Statistical Analysis on Choices and Rejections

In this section, we report the analysis on the number of choices and rejections in the robotics lab, the coding lab, and the control group: in particular, we analyzed the difference between selections and rejections at the beginning and at the end of the project. Is there a difference considering the beginning of activities and the end? Results on this research question are reported in **Table 2**.

Is robotics more effective than the other conditions to foster relations in the peer group? To answer this question, we have compared the three conditions in two moments: the pretest and the post-test running a one-way ANOVA with the software SPSS®.

At the pretest, the ANOVA revealed no significant differences between the three conditions: $F_{(2,67)} = 1.803$; $p = 0.173$ for selections and $F_{(2,67)} = 0.574$; $p = 0.566$ for rejections.

On the contrary, at the post-test, the difference is significant if we consider the selections: $F_{(2,67)} = 7.569$; $p = 0.001$ (for selections). The *post-hoc* comparisons (Bonferroni method)

TABLE 1 | Indexes for the experimental groups at pretest and at post-test obtained with the Gephi software.

	Nodes		Edges		Average		Index		Coherence	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Robotics Group										
Total (Social Intensity index)	23	23	196	323	8.52	14	0.387	0.638	0.56	0.62
Selection (Cohesion index)	23	23	126	239	5.48	10.39	0.25	0.472	0.67	0.65
Rejection (Dissociation index)	23	22	70	84	3	3.8	0.138	0.182	0.2	0.33
Coding Group										
Total (Social Intensity index)	24	24	206	255	8.59	10.6	0.373	0.462	0.53	0.42
Selection (Cohesion index)	24	24	107	193	4.46	8	0.194	0.35	0.64	0.48
Rejection (Dissociation index)	24	24	99	162	4.125	2.7	0.179	0.123	0.30	0.16
Control Group										
Total (Social Intensity index)	23	23	184	5 262	8	11.4	0.364	0.518	0.55	0.53
Selection (Cohesion index)	23	23	98	146	4.3	6.3	0.194	0.289	0.65	0.41
Rejection (Dissociation index)	23	23	86	116	3.7	5	0.17	0.229	0.33	0.34

Average is the ratio between connections (edges) and the participants (nodes). The column Index reports the Social Intensity, Cohesion, Dissociation indexes; the column Coherence indicates the coherence of the social network.

TABLE 2 | Comparison between pretest and post-test about received selections and rejects in the three experimental conditions ($p < 0.05$ are marked with an asterisk).

	Selections		Rejects	
	t-test	*p*-value	*t*-test	*p*-value
Robotics	7.71507274	≃0*	1.834425	0.079561
Coding	6.87784727	≃0*	3.279852	0.003286*
Control	3.73671073	0.001079202	2.998793	0.006408*

indicate that a statistically detectable difference emerges between the robotics condition and the control group: average difference $= 4.043$; $p = 0.001$.

DISCUSSION AND CONCLUSIONS

ER is nowadays a frequent appointment in curricular pathways; the experiment we have described and the related data indicate a notable change in the interpersonal relations within the group that attended the robotics lab in the direction of their improvement. This change emerges in the comparison with the control group and the coding lab. This result can be motivated by the shift of the learning perspectives, which becomes more active, and consequently by the different way students interact with each other. Indeed, according to the constructivist approach, this kind of activities offers the students the possibility to establish relations with their peers in a different way in order to understand their psychological affinities. To solve the robotics tasks, the participants must act in an interdependent way, whereas the majority of curricular activities are individual. Allowing to move from individual to group activity forces to build an interdependent relation: the students who are not well-included in the peer group have a new chance to be an active part in solving the tasks thus improving relationships with other students. The present study has indeed some limitations; for example, it was run on already established groups (classes), so it was not possible to vary the group composition. Moreover, the groups were followed along a relatively short period of time, and it would be interesting to verify if the positive changes were stable over time.

From these results, it is possible to deduce that labs and related activity can be an effective methodology to promote and support new and satisfying relations between students. The data reported in the Results section indicate that there is an increase in selections in the ER condition, which is higher than the other conditions, thus showing that the ER activities can have some specific features that are functional to improve the relations between peers, which is, in turn, a protective factor to prevent dropout.

Some issues remain still open: in the group that was involved in the robotics lab, a small number of participants remain rejected. Does it depend on the individual participant or from the group organization? And if it depends on the individual, which are the psychological variables that are relevant?

Future research will be devoted to address this question, along with the comparison with the educational robotic lab with different activities that foresee interactions with tangibles (such as laboratories of craft, art, workshop on music, etc.). These new experiments will investigate if social relations can be enhanced specifically by running a lab or if this improvement is comparable to the effects of other group activity involving manipulation.

This first project was followed by a wider experience under the Codinc project (Coding for inclusion) in the period January–May 2019. In this European-funded project, ER, together with sociometric tools, has been the context to assess peer relations and has become the core of Codinc methodology, as it offers the opportunity to portray interpersonal relationship in a situation that is different from the common interactions of the peers at school.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

MP, FR, and OM conceived the original idea and planned the experiments. FT carried out the experiment. MP, FR, FT, and DM run the statistical analyses. MP took the lead in writing the manuscript with the support of FR. FT and OM supervised the findings of this work. All authors discussed the results and contributed to the final manuscript.

FUNDING

This research has been supported by the EU project CODINC (592121-EPP-1-2017-1-BE-EPPKA3-IPI-SCO-IN) from Erasmus+ and STOP (Programme H2020-EU.1.3.3., MSCA-RISE-2018 – No. 823978).

ACKNOWLEDGMENTS

The authors thank the manager, the teachers, the students, and their parents of the school where the experiment was run for their collaboration.

REFERENCES

Alimisis, D. (2013). Educational robotics: open questions and new challenges. *Themes Sci. Technol. Educ.* 6, 63–71.

Arís, N., and Orcos, L. (2019). Educational robotics in the stage of secondary education: empirical study on motivation and STEM skills. *Educ. Sci.* 9:73. doi: 10.3390/educsci9020073

Atmatzidou, S., and Demetriadis, S. (2016). Advancing students' computational thinking skills through educational robotics: a study on age and gender relevant differences. *Robot. Auton. Syst.* 75, 661–670. doi: 10.1016/j.robot.2015.10.008

Atmatzidou, S., and Demetriadis, S. N. (2012). "Evaluating the role of collaboration scripts as group guiding tools in activities of educational robotics: conclusions from three case studies," in *2012 IEEE 12th International Conference on Advanced Learning Technologies* (Rome: IEEE), 298–302. doi: 10.1109/ICALT.2012.111

Bastian, M., Heymann, S., and Jacomy, M. (2009). "Gephi: an open source software for exploring and manipulating networks," in *International AAAI Conference on Weblogs and Social Media* (San Jose, CA).

Bazylev, D., Margun, A., Zimenko, K., Kremlev, A., and Rukujzha, E. (2014). Participation in robotics competition as motivation for learning. *Proc. Soc. Behav. Sci.* 152, 835–840. doi: 10.1016/j.sbspro.2014.09.330

Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: a systematic review. *Comput. Educ.* 58, 978–988. doi: 10.1016/j.compedu.2011.10.006

Burbaite, R., Stuikys, V., and Damasevicius, R. (2013). "Educational robots as collaborative learning objects for teaching computer science," in *2013 International Conference on System Science and Engineering (ICSSE)* (Budapest: IEEE), 211–216. doi: 10.1109/ICSSE.2013.6614661

Chung, C. C., Cartwright, C., and Cole, M. (2014). Assessing the impact of an autonomous robotics competition for STEM education. *J. STEM Educ. Innov. Res.* 15:24.

Denis, B., and Hubert, S. (2001). Collaborative learning in an educational robotics environment. *Comput. Hum. Behav.* 17, 465–480. doi: 10.1016/S0747-5632(01)00018-8

Di Fuccio, R., Ponticorvo, M., Di Ferdinando, A., and Miglino, O. (2015). "Towards hyper activity books for children. connecting activity books and montessori-like educational materials," in *Design for Teaching and Learning in a Networked World*, eds G. Conole, T. Klobučar, C. Rensing, J. Konert, and E. Lavoué (Cham: Springer), 401–406. doi: 10.1007/978-3-319-24258-3_31

Eguchi, A. (2014a). "Robotics as a learning tool for educational transformation," in *Proceeding of 4th International Workshop Teaching Robotics, Teaching with Robotics & 5th International Conference Robotics in Education* (Padova).

Eguchi, A. (2014b). Educational robotics for promoting 21st century skills. *J. Autom. Mob. Robot. Intell. Syst.* 8, 5–11. doi: 10.14313/JAMRIS_1-2014/1

Ferrara, F., Ponticorvo, M., Di Ferdinando, A., and Miglino, O. (2016). "Tangible interfaces for cognitive assessment and training in children: logicart," in *Smart Education and e-Learning 2016*, eds V. L. Uskov, R. J. Howlett, and L. C. Jain (Cham: Springer), 329–338. doi: 10.1007/978-3-319-39690-3_29

Frostad, P., Pijl, S. J., and Mjaavatn, P. E. (2015). Losing all interest in school: social participation as a predictor of the intention to leave upper secondary school early. *Scan. J. Educ. Res.* 59, 110–122. doi: 10.1080/00313831.2014.904420

Gabriele, L., Marocco, D., Bertacchini, F., Pantano, P., and Bilotta, E. (2017). An educational robotics lab to investigate cognitive strategies and to foster learning in an arts and humanities course degree. *Int. J. Online Eng.* 13:7. doi: 10.3991/ijoe.v13i04.6962

Garcia-Magarino, I., Plaza, I., Igual, R., Lombas, A. S., and Jamali, H. (2019). An agent-based simulator applied to teaching-learning process to predict sociometric indices in higher education. *IEEE Trans. Learn. Technol.* 1, 1–14. doi: 10.1109/TLT.2019.2910067

Gonnot, A., Michel, C., Marty, J. C., and Cordier, A. (2019). "Social robots in collaborative learning: consequences of the design on students' perception," in *11th International Conference on Computer Supported Education* (Heraklion). doi: 10.5220/0007731205490556

Hussain, S., Lindh, J., and Shukur, G. (2006). The effect of LEGO training on pupils' school performance in mathematics, problem solving ability and attitude: Swedish data. *J. Educ. Technol. Soc.* 9, 182–194.

Javidi, G., and Sheybani, E. (2010). Making youth excited about STEM education. *J. Comput. Sci. Coll.* 26, 140–147.

Kamga, R., Romero, M., Komis, V., and Mirsili, A. (2016). "Design requirements for educational robotics activities for sustaining collaborative problem solving," in *International Conference EduRobotics 2016,* (Cham, Springer), 225–228. doi: 10.1007/978-3-319-55553-9_18

Kandlhofer, M., and Steinbauer, G. (2016). Evaluating the impact of educational robotics on pupils' technical-and social-skills and science related attitudes. *Robot. Auton. Syst.* 75, 679–685. doi: 10.1016/j.robot.2015.09.007

Klassner, F., and Anderson, S. D. (2003). Lego mindstorms: not just for K-12 anymore. *IEEE Robot. Autom. Mag.* 10, 12–18. doi: 10.1109/MRA.2003.1213611

Lindh, J., and Holgersson, T. (2007). Does lego training stimulate pupils' ability to solve logical problems? *Comput. Educ.* 49, 1097–1111. doi: 10.1016/j.compedu.2005.12.008

Maloney, J., Resnick, M., Rusk, N., Silverman, B., and Eastmond, E. (2010). The scratch programming language and environment. *ACM Trans. Comput. Educ. (TOCE)* 10:16. doi: 10.1145/1868358.1868363

Miglino, O., Lund, H. H., and Cardaci, M. (1999). Robotics as an educational tool. *J. Interact. Learn. Res.* 10, 25–47.

Miglino, O., Rubinacci, F., Pagliarini, L., and Lund, H. H. (2004). Using artificial life to teach evolutionary biology. *Cogn. Process.* 5, 123–129. doi: 10.1007/s10339-004-0009-z

Mikropoulos, T. A., and Bellou, I. (2013). Educational robotics as mindtools. *Themes Sci. Technol. Educ.* 6, 5–14.

Mitnik, R., Nussbaum, M., and Soto, A. (2008). An autonomous educational mobile robot mediator. *Auton. Robot.* 25, 367–382. doi: 10.1007/s10514-008-9101-z

Moreno, J. L. (1941). Foundations of sociometry: an introduction. *Sociometry* 4, 15–35. doi: 10.2307/2785363

Moreno, J. L. (1951). *Sociometry, Experimental Method and the Science of Society* (Manchester: The North West Psychodrama Association).

Nugent, G., Barker, B., Grandgenett, N., and Adamchuk, V. (2009). "The use of digital manipulatives in k-12: robotics, GPS/GIS and programming," in *Frontiers in Education Conference FIE '09. 39th IEEE* (San Antonio, TX), 1–6, 18–21. doi: 10.1109/FIE.2009.5350828

O'Higgins, N., D'Amato, M., Caroleo, F. E., and Barone, A. (2007). Gone for Good? Determinants of school dropout in Southern Italy. *Giornale degli Economisti e Annali di Econ.* 66, 207–246.

Owens, G., Granader, Y., Humphrey, A., and Baron-Cohen, S. (2008). LEGO((R)) Therapy and the social use of language programme: an evaluation of two social skills interventions for children with high functioning autism and asperger syndrome. *J. Autism Dev. Disord.* 38, 1944–1957. doi: 10.1007/s10803-008-0590-6

Papert, S. (1980). *Mindstorms: Computers, Children and Powerful Ideas*. New York, NY: Basic Books.

Papert, S., and Harel, I. (1991). *Constructionism*. New York, NY: Ablex Publishing Corporation.

Piaget, J. (1974). *To Understand is to Invent*. New York, NY: Basic Books.

Plauska, I., and Damaševičius, R. (2014). "Educational robots for internet-of-things supported collaborative learning," in *International Conference on Information and Software Technologies*, (Cham, Springer), 346–358. doi: 10.1007/978-3-319-11958-8_28

Ponticorvo, M., Di Fuccio, R., Di Ferdinando, A., and Miglino, O. (2017). An agent-based modelling approach to build up educational digital games for kindergarten and primary schools. *Expert Syst.* 34:e12196. doi: 10.1111/exsy.12196

Ricard, N. C., and Pelletier, L. G. (2016). Dropping out of high school: the role of parent and teacher self-determination support, reciprocal friendships and academic motivation. *Contem. Educ. Psychol.* 44, 32–40. doi: 10.1016/j.cedpsych.2015.12.003

Rubinacci, F., Ponticorvo, M., Gigliotta, O., and Miglino, O. (2017a). "Breeding robots to learn how to rule complex systems," in *Robotics in Education,* (Cham, Springer), 137–142. doi: 10.1007/978-3-319-42975-5_13

Rubinacci, F., Ponticorvo, M., Passariello, R., and Miglino, O. (2017b). Robotics for soft skills training. *Res. Educ. Media* 10, 20–25. doi: 10.1515/rem-2017-0010

Sica, L. S., Ponticorvo, M., and Miglino, O. (2019a). "Enhancing digital creativity in education: the docent project approach," in *International Conference in Methodologies and Intelligent Systems for Techhnology Enhanced Learning* (Cham, Springer), 103–110. doi: 10.1007/978-3-030-23884-1_14

Sica, L. S., Ponticorvo, M., Miglino, O., and Rubinacci, F. (2019b). "Between creative process and creative product: DoCENT Mooc for the integration of technology, creativity and learning," in *Psychobit 2019* (Naples).

Strohkorb, S., Fukuto, E., Warren, N., Taylor, C., Berry, B., and Scassellati, B. (2016). "Improving human-human collaboration between children with a social robot," in *Robot and Human Interactive Communication (ROMAN), 2016 25th IEEE International Symposium* (New York, NY: IEEE), 551–556. doi: 10.1109/ROMAN.2016.7745172

Sullivan, F. R. (2008). Robotics and science literacy: thinking skills, science process skills and systems understanding. *J. Res. Sci. Teach.* 45, 373–394. doi: 10.1002/tea.20238

Truglio, F., Marocco, D., Miglino, O., Ponticorvo, M., and Rubinacci, F. (2018a). "Educational robotics to support social relations at school," in *International Conference on Robotics and Education RiE 2017* (Cham, Springer), 168–174. doi: 10.1007/978-3-319-97085-1_17

Truglio, F., Ponticorvo, M., and Rubinacci, F. (2018b). Educational robotics improves social relations at school. *ERCIM NEWS*, *(114)*, 38–39.

Whittier, L. E., and Robinson, M. (2007). Teaching evolution to non-English proficient students by using lego robotics. *Am. Secondary Educ.* 35, 19–28.

Williams, D., Ma, Y., Prejean, L., Lai, G., and Ford, M. (2007). Acquisition of physics content knowledge and scientific inquiry skills in a robotics summer camp. *J. Res. Technol. Educ.* 40, 201–216. doi: 10.1080/15391523.2007.10782505

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 04 August 2020
doi: 10.3389/frobt.2020.00101

The Effects of Feedback on Children's Engagement and Learning Outcomes in Robot-Assisted Second Language Learning

Mirjam de Haas[1]*, *Paul Vogt*[1] *and Emiel Krahmer*[2]

[1] Department of Cognitive Science and Artificial Intelligence, Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, Netherlands, [2] Department of Communication and Cognition, Tilburg Center for Cognition and Communication, Tilburg School of Humanities and Digital Sciences, Tilburg University, Tilburg, Netherlands

To investigate how a robot's use of feedback can influence children's engagement and support second language learning, we conducted an experiment in which 72 children of 5 years old learned 18 English animal names from a humanoid robot tutor in three different sessions. During each session, children played 24 rounds in an "I spy with my little eye" game with the robot, and in each session the robot provided them with a different type of feedback. These feedback types were based on a questionnaire study that we conducted with student teachers and the outcome of this questionnaire was translated to three within-design conditions: (teacher) preferred feedback, (teacher) dispreferred feedback and no feedback. During the preferred feedback session, among others, the robot varied his feedback and gave children the opportunity to try again (e.g., "Well done! You clicked on the horse.", "Too bad, you pressed the bird. Try again. Please click on the horse."); during the dispreferred feedback the robot did not vary the feedback ("Well done!", "Too bad.") and children did not receive an extra attempt to try again; and during no feedback the robot did not comment on the children's performances at all. We measured the children's engagement with the task and with the robot as well as their learning gain, as a function of condition. Results show that children tended to be more engaged with the robot and task when the robot used preferred feedback than in the two other conditions. However, preferred or dispreferred feedback did not have an influence on learning gain. Children learned on average the same number of words in all conditions. These findings are especially interesting for long-term interactions where engagement of children often drops. Moreover, feedback can become more important for learning when children need to rely more on feedback, for example, when words or language constructions are more complex than in our experiment. The experiment's method, measurements and main hypotheses were preregistered.

Keywords: child-robot interaction, second-language learning, robot tutor, feedback, engagement, preschool children

Edited by:
Dirk Ifenthaler,
Universität Mannheim, Germany

Reviewed by:
Hideyuki Nakanishi,
Osaka University, Japan
Emilia I. Barakova,
Eindhoven University of Technology, Netherlands
Hatice Kose,
Istanbul Technical University, Turkey

***Correspondence:**
Mirjam de Haas
mirjam.dehaas@tilburguniversity.edu

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *25 October 2019*
Accepted: *25 June 2020*
Published: *04 August 2020*

Citation:
de Haas M, Vogt P and Krahmer E (2020) The Effects of Feedback on Children's Engagement and Learning Outcomes in Robot-Assisted Second Language Learning. Front. Robot. AI 7:101. doi: 10.3389/frobt.2020.00101

1. INTRODUCTION

A recent trend in education is to have social robots take on the role of educational tutors to support, for example, second language learning (Westlund and Breazeal, 2015; Belpaeme et al., 2018; Vogt et al., 2019). In educational settings, learning a (second) language typically involves social interactions between the child and the teacher. During these interactions, children constantly receive feedback about their performance. It has been shown that human feedback can have a clear impact on children's learning process and outcomes (Wojitas, 1998; Hattie and Timperley, 2007). Feedback is therefore an important part of the social interactions that facilitate language learning, which begs the question what the impact of various feedback types is when feedback is provided by a robot rather than a human.

Throughout many years researchers have investigated how (human) feedback can have an influence on second language learning. Focusing on children learning a second language, research has shown that receiving feedback benefits children's language development more than receiving no feedback (Mackey and Silver, 2005). Moreover, different types of feedback can help children in several ways. You can, for example, use positive feedback to reward and motivate children when they are correct, or use negative feedback to correct children's language and thereby improve children's learning gain (Hattie and Timperley, 2007).

While there have been many studies about robots for educating children, only few of these have investigated the effects that different types of feedback can have on children's engagement and learning performance (Gordon et al., 2016; De Haas et al., 2017; Ahmad et al., 2019). Usually, studies design feedback strategies for robot tutors based on results from educational studies involving only humans without investigating the effect that these strategies have on children's engagement and/or performance (e.g., Mazzoni and Benvenuti, 2015; Westlund and Breazeal, 2015; Gordon et al., 2016; Kennedy et al., 2016). However, it is not evident that the effect of human strategies will be the same when a robot uses them, because a robot has substantial cognitive and physical limitations compared to a human. For example, robots cannot produce the same facial expressions as humans or humans' subtle cues, thus are limited in providing facial cues that humans use to provide non-verbal feedback.

One recent study manipulated non-verbal and verbal feedback based on the child's emotional state (Ahmad et al., 2019). Results showed that children's engagement over time remained relatively high and children's word knowledge increased over time with positive or neutral feedback. While their results suggest that robot feedback can have a positive effect on children's engagement and learning gain, they did not compare different variations of positive and negative feedback or compared it against no feedback.

The results of Ahmad et al. (2019) are consistent with findings from human studies and demonstrate that feedback does not only enhance children's language performance, but also engages children. Positive feedback engages because it validates children's answers and thus boosts their confidence (Henderlong and Lepper, 2002; Zentall and Morris, 2010). Similarly, negative (corrective) feedback corrects and teaches the child the correct word which could result in a motivated child. However, both positive and negative feedback can also decrease engagement. On the one hand, too many repetitions of positive feedback can become meaningless for a child and can result in less intrinsic motivation (Henderlong and Lepper, 2002; Boyer et al., 2008). On the other hand, negative feedback can decrease the child's confidence and thereby decrease the engagement between the teacher and child (Wojitas, 1998).

Consequently, if used correctly, feedback can result in increased learning gains. Children become more intrinsically motivated by positive feedback, which increases the children's interest and their task engagement and therefore their skills. These increased skills will motivate the children further and engage the children to a greater extent (Blumenfeld et al., 2006).

This paper describes a study that investigated how preschool children respond to different types of feedback provided by a robot tutor. In the experiment, children interacted with a humanoid robot tutor in three different second-language sessions, and in each session the children received a different type of feedback. These types of feedback were designed based on a survey among student teachers, resulting in a strategy preferred by these student teachers, a strategy dispreferred by them and a strategy using no feedback at all. We analyzed the effect of these different types of feedback on the children's task engagement and learning gain over time.

2. BACKGROUND

2.1. Feedback

Numerous studies have shown that feedback facilitates second language learning (Lyster and Ranta, 1997; Henderlong and Lepper, 2002; Long, 2006; Hattie and Timperley, 2007). It can help to improve pronunciation, word-choice and grammar, and makes it easier for children to understand what is correct or incorrect in the foreign language. Feedback is not only used to correct children, but for example also by teachers to contribute positively to children's own feeling of competence and success and therefore encourage children to continue with a task (Blumenfeld et al., 2006; Hattie and Timperley, 2007). The type of feedback provided, however, matters (Shute, 2008). You can, for example, provide explicit negative feedback by indicating that something is wrong with children's answers, but without specifying what was wrong (e.g., "That's wrong."). It is also possible to provide corrective feedback by correcting children's answers or hinting toward it (e.g., "You said runned, but it should have been ran" or "it should not have been runned, but...?"). Prompting children with an extra attempt ("Try again.") is an implicit way of saying something was wrong. Hattie and Timperley (2007) propose a combination of these three types as good way of providing feedback. The combination provides children with explicit notions where the mistake was made, what went wrong and makes them to try again. Nevertheless, sometimes separate feedback is also sufficient. For example, using explicit negative feedback (i.e., stating explicitly that something is wrong) seems to be most beneficial for children who are struggling with a task, such as novel learners (Kluger and DeNisi, 1996; Shute, 2008).

Teachers, however, mostly provide negative feedback implicitly by using recasts (i.e., a type of feedback in which the teacher repeats the incorrect phrases in a correct form), but they still try to make sure that children reach their goal (Lyster and Ranta, 1997; Long, 2006). Although these recasts have been found to be used more often than the other feedback types, they seem to be less effective in helping the learner to reach their learning goal. Lyster and Ranta (1997) investigated the role of negative feedback and found that when teachers explicitly mentioned the fact that an error was made in their negative feedback, it led to a higher learning gain than when they did not, which suggests that explicit negative (or corrective) feedback can be more effective than implicit feedback by using recasts.

Feedback is not always negative or corrective, it can also be positive. In general, teachers mostly use positive feedback explicitly (praise) and not implicitly (Hattie and Timperley, 2007). The advantage of praise is that it approves children's answers and makes the task encouraging and motivating (Henderlong and Lepper, 2002). When children receive positive feedback, they become happy, and are therefore more committed and intrinsically motivated to complete a task. However, there are also downsides to providing positive feedback. When children receive too much positive feedback, they rely on the feedback and will not learn when they do not receive the feedback anymore (Henderlong and Lepper, 2002). In addition, when the use of praise is non-specific or ambiguous, such as saying "good job" or "beautiful" makes children not understand what part of their answer elicited the feedback and they will not know how to respond (Hamilton and Gordon, 1978). Thus, positive feedback should refer to the learning task and at the same time remain motivating enough in order to be effective.

2.1.1. Feedback, Engagement, and Learning

Engagement seems to have a positive effect on language learning (Christenson et al., 2012). A considerable amount of studies have shown that robots are engaging interaction partners for both adults and children (see for an overview Kanero et al., 2018). Engagement normally starts high due to the novelty effect but then seems to decrease over time (Kanda et al., 2007; Westlund and Breazeal, 2015; Rintjema et al., 2018). When talking about engagement, it can be helpful to distinguish between two kinds of engagement: robot-engagement, referring to how engaged a child is with the robot, and task-engagement, which focuses on how engaged a child is with the learning task. Clearly, these are not necessarily the same: a child can be very engaged with their social partner, the robot, but not with the task, or visa versa. Moreover, the effect of these different engagement types on learning gain can differ. For example, one study by Kennedy et al. (2015) used a highly engaging robot partner and, as a result, children were so distracted by the robot that they focused less on the task and therefore learned less. In their study, children who were highly engaged with the robot, learned less instead of more while it is possible that children who are highly engaged with the task, will still learn more. Consequently, it is useful to measure both types of engagement: task-engagement and robot-engagement.

Research in HRI has looked at many ways of keeping general engagement high, but did not investigate the role that different types of feedback could play here. For example, Ahmad et al. (2019) looked at the role of adaptive feedback on the children's emotion on engagement, but they did not investigate the effect of different types of feedback.

Feedback, however, can have an influence on children's motivation and their self-evaluation (Zentall and Morris, 2010), which—in turn—can influence engagement. Blumenfeld et al. (2006) suggested a feedback loop: in order to increase children's engagement, children first have to be motivated, which will then increase their interest in the task, which in turn will engage children followed by the children's learning gain. When children improve their language skills, this can lead to even higher motivation and further result in a higher engagement.

The influence of feedback on motivation depends on the type of feedback. For instance, praise that is specifically linked with the children's effort (e.g., "You are a good drawer" after drawing a picture) motivates children more than other types of praise, even when only 75% of the praise is linked with effort (Zentall and Morris, 2010). Moreover, Corpus and Lepper (2007) showed that for preschool children all praise enhanced motivation when they compared it with neutral feedback ("OK"). They compared motivation of preschool children with older children, and found that only for older children (fourth and fifth graders) the type of praise had an influence on their motivation, while preschool children benefited from all feedback equally. Another study found similar results: Morris and Zentall (2014) measured ambiguous praise ("Well done!", "Yeah," "Awesome") and found higher persistence, higher self-evaluations and fewer fixations on later mistakes. Apparently, children interpret ambiguous praise in the most beneficial manner for themselves. However, they also found that the use of gestures ("Thumbs up" and "High five") resulted in the highest self-evaluations.

The reason why feedback has an influence on motivation and therefore engagement can be explained by the Self-Determination Theory (Deci and Ryan, 1985). This theory poses that learners continue a task longer when their motivation is based on intrinsic aspects, such as pleasure and satisfaction, compared to when motivation is based on external rewards (Deci and Ryan, 1985). This intrinsic motivation arises particularly when a task contains autonomy and competence and is strengthened by a sense of relatedness between learner and teacher (Ryan and Deci, 2000). For example, autonomy increases when a learner can choose themselves what kind of activity to do, or when he or she receives informative rewards and non-controlling instructions. A higher degree of autonomy leads to increased intrinsic motivation and, in turn, higher levels of engagement. Moreover, competence increases with praise (Blanck et al., 1984), because it enhances the children's feeling of being capable to successfully complete a challenging task. Competence, especially in combination with autonomy, plays a considerable role in retaining intrinsic motivation. There are also disadvantages of praise, for example, when children first receive praise but are not able to successfully complete the task, their motivation can decrease (Zentall and Morris, 2012). Moreover, too much positive feedback can decrease the children's own curiosity (Henderlong and Lepper, 2002).

Negative feedback has been found to decrease intrinsic motivation, specifically the feeling of competence (Deci et al., 1991). It can potentially decrease children's self-efficacy or their active participation and engagement in the learning task, because children become unmotivated when receiving negative feedback (Wojitas, 1998). On the other hand, negative feedback can also have a positive influence on motivation, as it can help children to understand what they are trying to learn and to correct their mistakes (Hattie and Timperley, 2007). Kluger and DeNisi (1996) suggest that, similar as with praise, the effect of feedback is not only dependent on a link between behavior and feedback, but also on how the feedback was provided and how the learner interprets the feedback.

The combination of praise and negative feedback can be challenging enough for children, but at the same time motivates children enough to want to continue with the task. For example, if children additionally receive negative feedback to correct their mistakes and hear praise when they correctly answer a question, this can enhance the effect of both feedback types. Summarizing, feedback has the potential to both engage and disengage children (Dempsey and Sales, 1993), depending on the type of feedback given. Feedback (especially praise) can increase the intrinsic motivation of children, which increases their engagement. Engaged children are more motivated, learn faster, will be more likely to complete the task and to repeat the task, which leads to a better result (Dörnyei, 1998). However, it is not clear yet whether the rules that apply to human teacher-child interactions also apply to robot-child interactions.

2.1.2. Feedback in Child-Robot Interaction

Studies with educational robots for children that have explicitly looked at the role of feedback are sparse. While many studies have incorporated the use of feedback, specifically praise (Mazzoni and Benvenuti, 2015; Westlund and Breazeal, 2015; Gordon et al., 2016; Kennedy et al., 2016), they did not test the effect of feedback on the children's engagement or learning gain nor the effects that different forms of feedback may have. These studies investigated the role of praise either by incorporating it as part of a robot's strategy (Westlund and Breazeal, 2015; Kennedy et al., 2016), by looking at specific responses of children on occurrences of praise (Serholt and Barendregt, 2016) or on the effect of timing of the praise (Park et al., 2017). It seems that children notice the praise and react to it, however, these studies did not investigate its direct effect on engagement and learning gain. For example, Kennedy et al. (2016) compared a high verbal availability robot and a low verbal availability robot. The high verbal availability robot used—among other social behaviors—more expressive praise than the low verbal availability robot. Children of approximately 8 years old practiced different French grammar rules with one of the robots. The authors found no significant difference in learning gain for the robot that used more expressive positive feedback, but the children reported to have noticed the praise and payed attention to it.

In another study, Serholt and Barendregt (2016) investigated children's responses to the robot's praise. In their long-term study, the robot gave praise on the children's performance of the previous session. Positive feedback did not go unnoticed, 70% of the children acknowledged the robot during feedback through verbal or gestural responses such as smiling. Similarly, Park et al. (2017) explored whether the timing of a robot's praises would influence the engagement of children. Children had to tell a robot a story and the robot reacted on their emotional level as a form of feedback. For example, when children had a high energy level, the robot played a large excited motion. Park and colleagues compared two conditions, one with a robot that reacted every 5 s on the child without changing its energy level, and one with a robot that reacted during breaks between child speech and changed the energy level of its responses appropriately. The children seemed to be more engaged with latter robot that changed its feedback to their energy level. Likewise, Westlund and Breazeal (2015) used a non-humanoid robot to teach children a second language and found that children learned with a social robot more than with a non-social robot. Both robots used positive phrases when children were correct, e.g., "Good job!" or "You're working hard!" and only provided hints with an incorrect answer, e.g., "I think it was that one." However, the social robot added expressive phrases based on the child's emotional state (e.g when children were excited, the robot first reacted with "woo hoo" before the feedback).

While many robots use praise, which is an explicit form of positive feedback, explicit negative feedback is not often used in child-robot studies. Typically, studies incorporated implicit feedback by using hints (e.g., "I think it was the other one," Gordon et al., 2016) or by introducing doubts ("Are you sure?" Mazzoni and Benvenuti, 2015).

Three studies that specifically investigated the effect that feedback has on learning and/or engagement are those by De Haas et al. (2017), Resing et al. (2019), and Ahmad et al. (2019). De Haas et al. (2017) conducted a between-subject study with 4-year-old pre-school children that compared the effect that three different feedback strategies (peer-like, adult-like, and no feedback) had on learning gain and engagement. The feedback strategies did not affect the learning gain or the engagement measured through eye-contact. Instead, children showed a substantial amount of individual differences in how they engaged with the robot across the three feedback conditions. Some children focused completely on the robot, while other children focused more on the researcher by asking for more guidance. Even though children did not seem to benefit from the different types of feedback, this study consisted of only one session which—due to the novelty effect—may have disturbed the effect that different forms of feedback may have.

Resing et al. (2019) reported a study where 6 till 9-year-old children had to solve a puzzle together with an owl-like robot that either helped them by giving feedback or did not provide any help. The help-providing robot used both verbal and non-verbal feedback. It shook its head and had blinking eyes when their answer was incorrect as a way of providing non-verbal (explicit) negative feedback, or nodded and said "Well done!", with (different) blinking eyes as a form of explicit positive feedback. Children trained by the robot with feedback became better in solving new puzzles than children trained with the other robot. However, again, children showed large individual differences in the number of corrections they needed.

Ahmad et al. (2019) addressed individual differences between children and compared in a between-subjects design a robot that adapted its feedback with one that did not. They studied how children between 10 and 12 years old responded to the robot's feedback during 2 weeks. The robot adapted its feedback behavior to the children's emotional state. For example, when children were rated as happy the robot used that in its feedback ("You are looking happy, and I'm happy that you are in front of me. Let's learn another word"). During the game, the robot kept referring to the game outcome, only in the post-test the robot provided feedback on learning performance ("I am happy that you got it wrong in session one, but this time your answer is correct' or 'It's sad that you didn't remember this word, the correct answer is..."). Ahmad and colleagues found that the children's engagement remained relatively high (or stable) when interacting with the adaptive robot, while their engagement lowered over time with the non-adaptive one. Moreover, children's learning gain was higher with the adaptive robot, compared to the non-adaptive one. While these results are promising, this study did not investigate the effect of different feedback strategies.

Generally, developers of robot tutors base the educational strategies of the robot on the already existing human studies and use those strategies in their child-robot interactions without studying whether these strategies are similarly effective. Most child-robot studies use praise as a motivator in their experiments and are hesitant to use explicit negative feedback. It is not clear what type of negative feedback works best for robots, although in educational studies it seems that mentioning the children's mistake seems to be more effective for language learning. In this paper, we address this gap in knowledge by investigating the effect of different forms of feedback on both task-, robot-engagement and learning gain.

2.2. Teachers' Feedback

In preparation of the present study, we carried out a survey among student teachers concerning their views on how a robot should provide feedback. The aim of this survey was two-fold: (1) To gain insights how student teachers' would think the robot should provide feedback to children giving correct and incorrect answers in a tutoring setting, and with varying levels of the children's engagement at the time feedback is given. (2) To create a data set with different feedback phrases that student teachers would use. We interviewed student teachers instead of practicing teachers, because students are more likely to work with technologies in the future, such as social robots, than teachers who already worked for many years. Moreover, receiving many responses was more feasible with student teachers than with teachers.

In our survey, we showed 27 student teachers 40 video fragments of both engaged and disengaged children interacting with a robot in a second language tutoring experiment reported in De Wit et al. (2018). All fragments showed a robot teaching 5- to 6-year-old Dutch children animal words in English as a second language. In each fragment, the robot expressed an English word and asked the child to select—on a tablet—the animal he or she thought that the word referred to. The fragment ended right after the child answered to this request. After watching each fragment, the student teachers were asked to provide a feedback suggestion. The survey was carried out in a between-subject design with two conditions: in one condition (closed questions), student teachers could choose between six feedback strategies (three positive and three negative), and in the other condition (open-ended questions) they could freely write the feedback themselves. This closed questions survey would provide insights of what strategy student teachers would choose, and the open questionnaire would create a data set of different feedback phrases.

We did not find a difference between student teachers' suggestions for engaged or disengaged children. However, we found that the suggested forms of feedback differed substantially between the closed and open-ended questionnaires: In the closed questions survey, the majority of the student teachers chose to use an explicit positive phrasing together with an explanation in the form of a translation ["Goed zo! Een 'hippo' is een nijlpaard" (Dutch)—"Well done! A 'hippo' is a hippo" (English)], and they chose a correction of the child's answer through repetition and translation of the target words ["Een hippo is een nijlpaard, je moet de nijlpaard aanraken" (Dutch)—"A 'hippo' is a hippo, you have to touch the hippo" (English)] as a means of providing implicit negative feedback.

In the case of the open-ended survey, the student teachers chose for both positive and negative feedback to only provide an explicit phrasing without repeating the target words for both positive feedback ["Goedzo" (Dutch)—"Well done" (English)] and negative feedback ["Helaas dat was niet goed" (Dutch)—"Unfortunately, that was not correct" (English)]. Moreover, we found that in the open-ended questionnaire student teachers varied their phrasing of the feedback considerably. These results indicate that student teachers do not have a straightforward strategy for choosing how to provide feedback.

After the surveys were analyzed, we discussed the findings with a subset of the student teachers. They suggested two main reasons why these results differed. Firstly, correction and explanation (e.g., through repetition of target words) is essential for negative feedback. This was the main reason why they chose to repeat the target words in the closed-ended questionnaire. Secondly, they indicated that variation in the form by which feedback is provided is also crucial. The robot should not repeat the same phrase throughout the whole lesson. Student teachers participating in the open-ended questionnaire focused more on creating varying feedback phrases and less on the repetition of the target word.

Based on these findings, we concluded that the "preferred" feedback strategy would combine the results from the closed questions survey with the open-ended survey: take an explicit feedback phrase (e.g., "Well done" or "That's wrong"), add a repetition of the target word, and provide children an extra attempt when their answers are incorrect. Since variation is key, the feedback phrases should vary, based on the data set created by the open-ended survey.

2.3. This Study

The present study investigates whether 5- and 6-year-old children are more engaged with the task and with the robot, and learn more words when participating in a second language (L2) training with a robot that provides feedback as recommended by the student teachers (preferred feedback), compared to a robot that provides feedback contrary to what was recommended by the student teachers (dispreferred feedback), and compared to a robot that provides no feedback at all (no feedback). As our survey with student teachers revealed, providing adequate feedback is a complex matter that consists of multiple strategies, which are hard to separate, thus making it difficult to investigate such individual factors experimentally. We therefore decided to combine multiple factors in our preferred and dispreferred feedback strategies, and explore to what extent these strategies, as performed by a robot, influence children's engagement and learning gain in an L2 tutoring scenario.

Every child receives three sessions with different robots, each providing a different form of feedback, thus allowing us to investigate how children react to the different forms of feedback using a within-subjects design. We based the training sessions on previous studies in which children played an "I spy with my little eye" game with a NAO robot to learn different L2 words (De Wit et al., 2018; Schodde et al., 2019).

Based on previous findings in literature regarding the role of feedback in second language learning, and previous studies that address feedback in child-robot interactions (Ahmad et al., 2019), we hypothesize that children will be more task- and robot-engaged when receiving (either preferred or dispreferred) feedback than when they do not receive feedback (H1a). Especially positive feedback is expected to increase the children's intrinsic motivation for the task and thus their engagement. We also hypothesize that children will remember more words when receiving feedback than when receiving no feedback (H1b). Feedback can help to understand whether an answer is correct or not and may indicate what the correct form should be, thus providing insight into the learning process and helps to improve the learning performance.

Moreover, we hypothesize that children will be more task- and robot-engaged with (H2a) and will remember more words from (H2b) a robot that provides feedback as preferred by a student teacher compared to a robot that provides dispreferred feedback. When feedback is varied (as in the preferred feedback strategy), children are expected to pay more attention to it, boosting their confidence and with that their task-engagement. The varied feedback of the robot can additionally increase the children's interest in the robot and with that their robot-engagement. In contrast, when a robot repeatedly uses the same phrase as feedback (dispreferred feedback), children might get tired of this repetition and as a result will pay less attention to the robot. Additionally, children can practice with the preferred feedback once more in the case of a mistake and thus improve their knowledge, which they cannot with the dispreferred feedback strategy and which might lead to an increase in their task-engagement. Moreover, the preferred feedback also provides children with an explicit notion where the mistake has been made, what went wrong and how they can fix it by trying again (the three rules of good feedback according to Hattie and Timperley, 2007).

3. METHODS

The research questions, hypotheses and analyses in this study have been preregistered at AsPredicted[1] and the source code has been made publicly available[2].

3.1. Design

The study was a within-subjects design, where all participants were assigned to all feedback strategies/conditions (each session a different strategy). The strategies for providing feedback were based on the survey asking student teachers how they would make the robot provide feedback in situations comparable to the ones in this experiment, translating to a preferred strategy and dispreferred strategy. The order of the feedback strategies and word sets were counterbalanced using a 3 × 3 latin-square to reduce an order effect. The three strategies/conditions were

1. Preferred feedback
2. Dispreferred feedback
3. No feedback

Each child received three sessions with the robot, and could learn 18 words in total and 6 in each session. In all conditions, all sessions were the same, except for the words learned, the feedback strategy that the robot used and the shirt the robot was wearing (to give the impression that children were playing with three different robots, see **Figure 1**).

3.2. Participants

In total, 72 native Dutch-speaking children aged 5 and 6 years participated in the current study. The participants were recruited from three elementary schools located in the Netherlands. Bilingual children were excluded from the study. A pre-test showed that 12 children were familiar with more than half of the target words and these children were excluded from the study in accordance with the exclusion criteria of our preregistration. Furthermore, four children dropped out of the study for various reasons like unwillingness to continue (3) or sickness (1). This resulted in 56 children (28 boys, *Mage* = 5 years and 6 months, *SDage* = 5 months) participating in the final experiment. All parents gave informed consent for the participation of their child.

3.3. Materials

The Softbank Robotics NAO robot and a Microsoft Surface tablet computer were used. The lessons involved one-on-one interactions between robot and child. We did not rely on automatic speech recognition because speech recognition has been shown to not work well with this age group (Kennedy et al., 2017). Instead the experimenter used a Wizard of Oz technique when the child had to say something to the robot in the beginning of the experiment. The robot was placed in a crouching position in an angle of 90 degrees next to the sitting child to give the robot

[1] https://aspredicted.org/qg6dx.pdf
[2] https://github.com/l2tor/feedback-study/

FIGURE 1 | (A–C) show the different shirts for each sessions. All children saw the robot wearing the red shirt during the first session and all children saw the robot wearing the yellow shirt during the last session.

the same perspective of the child, while still being able to face the child. The tablet was placed on top of a small box in front of the robot and child. A video camera placed on a tripod facing the child to record the child's responses and facial expressions. A second camera was placed from the side to get a more complete overview of the interactions. Each session was distinguished by a different color shirt and robot name (see **Figure 1**). We used the different shirts and names to make it known to children that they would play with three different robots, with different robot behaviors (namely the robot feedback strategies). The shirts were not linked to feedback conditions or different word sets, but rather to the lesson number. In other words, all children started with the robot wearing the red shirt called Luka during the first session and ended with the robot wearing the yellow shirt called Charlie.

3.3.1. Target Words

In total 18 target words were selected and during each lesson, children learned six target words. Target words were selected such that children can be expected to have acquired those in their first language but arguably not in their second language. Moreover, we selected words that would not be too similar in their L1 and in their L2 [e.g., not "Olifant" (Dutch) and "Elephant" (English)]. All 18 words were divided in three word sets based on their frequency in the children's first language. We used the dataset of Schrooten and Vermeer (1994) and placed each word in a frequency bin. Words in the same bin were randomly assigned to the different word sets. For example, the word "dog" was from the same frequency bin as the words "bird" and "horse" and were thus added to different word sets. See **Table 1** to see all target words with their frequency. We used cartoon-like images of the target animals during the experiment (see **Figure 2** for examples).

3.3.2. Pre-test

Before the children started the three sessions, we tested their L2 knowledge of the 18 target words with a comprehension test which was a picture-selection task. In this test, children were presented with a pre-recorded target word spoken by a bilingual speaker of Dutch and English and asked to choose which one out of four pictures matched this word ["Waar zie je een dog?" (Dutch) "Where do you see a: dog?" (English)]. The presentation of the target words in the pre-test was randomized for each child. We presented each target word one time during the pre-test.

3.3.3. Post-test

The children's long-term knowledge was tested between 2 and 3 weeks after the last session with the comprehension test. The test was the same as the pre-test only this time, each target word was presented three times in a random order to reduce chance level performance due to guessing. The reason for not doing so in the pretest was to reduce the chance of children learning from this task (Smith and Yu, 2008). A word was registered as correct if it was selected correctly at least twice out of the three trials. Additionally, we tested three different pictures of the animals in order to generalize the children's knowledge. To be more specific, we used a cartoon-like picture, a drawn picture (the same as in the experiment) and a photograph of the target animal.

In addition to the measurements described in this paper we also carried out a perception questionnaire of the robot at the end of all sessions. We will not discuss those results because this questionnaire is beyond the scope of this paper.

3.4. Tutoring Sessions

The lessons were based on the children's game "I spy with my little eye" and on the interaction described in Schodde et al. (2019). The whole interaction was in the children's L1, except for the target words. Before the three tutoring sessions, children had a group introduction to the robot and took a pre-test.

The tutoring session had four parts which were all repeated during all three tutoring sessions:

1. Start phase. The robot explained that he was a friend of the group introductory robot, he asked for the child's name, age and some questions about their favorite animals and games. The robot finished with saying that "I spy with my little eye" is his favorite game and that he wants to play that with the children. He then explained the rules of the game.

TABLE 1 | Target words with their frequency scores in Dutch taken from Schrooten and Vermeer (1994).

Word set 1			Word set 2			Word set 3		
Dutch	**English**	**Freq**	**Dutch**	**English**	**Freq**	**Dutch**	**English**	**Freq**
Hond	Dog	98	Vogel	Bird	72	Paard	Horse	64
Kikker	Frog	27	Kip	Chicken	30	Konijn	Rabbit	48
Vlinder	Butterfly	22	Nijlpaard	Hippo	16	Varken	Pig	36
Papagaai	Parrot	9	Slang	Snake	14	Eekhoorn	Squirrel	13
Haai	Shark	9	Slak	Snail	14	Zeehond	Seal	10
Neushoorn	Rhino	9	Walvis	Whale	9	Hert	Deer	9

Words that have a higher score are more familiar to children in Dutch.

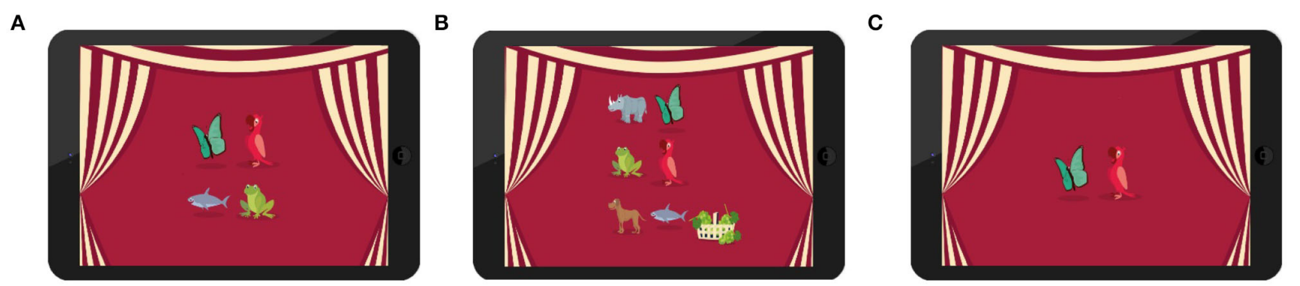


FIGURE 2 | (A) Training rounds. Each round the robot named one animal that children had to find **(B)** In-game test. Children had to drag a grape to the animal that the robot named **(C)** second attempt after wrong answer. Children were allowed to correct themselves in the preferred feedback condition. In this example, the child wrongly chose a butterfly instead of a parrot and could correct his/her mistake by selecting the correct one.

2. Concept binding of the target words. To teach children the target words, the tablet showed an animal on the screen, the robot said the L2 word with the L1 translation and asked the child to repeat the word [e.g., "Een vogel is een bird in het engels, zeg mij maar na bird" (Dutch). "A bird is a *bird* in English, repeat after me *bird*" (English)]. Only after the child had repeated the animal, they continued to the next animal. When a child did not repeat the robot, the experimenter asked the child to listen to the robot and repeat after the robot. If a child was very hesitant to repeat the word, the experimenter would say it together with the child.
3. Training rounds. After the concept binding the robot explained to the child that he would ask for an animal and that the child had to search for it on the tablet screen. They first practiced with an L1 word that was no target ("Ik zie, ik zie wat jij niet ziet en het is een eenhoorn, zoek maar naar de eenhoorn," "I spy with my little eye a unicorn, please search for the unicorn"). For each target word the tablet showed the target animal with three distractors (see **Figure 2A**). After the L1 practice round, the robot and child also practiced once in L2. After these two practice rounds they started the training of the target words. The robot constantly asked the child to search for a target word ("Ik zie, ik zie wat jij niet ziet en het is een <target word> zoek maar naar de <target word>, 'I spy with my little eye a <target word>, please search for the <target word >"). Depending on the condition the robot provided feedback or not and the child continued to the next animal. There were 24 rounds in total, each animal was trained four times, which made the L2 exposure to all animals ten times in total for all conditions (twice in the concept binding, eight times during the practice rounds).
4. In-game test. After each session there was an in-game test that tested the short-term memory of the target words. The tablet screen showed all animals of that tutoring sessions and a bucket of grapes (see **Figure 2B**). Each round, the robot named an animal and the child had to feed this animal with one of the grapes. The robot asked the animals in random order and after each round the order of presenting the animals on the screen was shuffled.

All conditions had the exact same design, meaning that the lesson structure was the same, the tablet output was the same and the behavior of the robot was the same, except for the feedback. In all conditions, the robot used the standard following-gaze feature of NAO.

3.5. Feedback Conditions

All feedback was provided in the children's L1 to keep the L2 exposure consistent between conditions. A comparison of the different types of feedback can be found in **Table 2**. The feedback conditions were based on the student teachers' preferred response for the robot (preferred feedback), the opposite (dispreferred feedback) and a control condition was added where the robot did not use any feedback. Preferred and dispreferred feedback different on multiple aspects:

TABLE 2 | An example of the robot's feedback in the different feedback conditions.

	Correct answer		Incorrect answer	
Condition	**Dutch**	**English**	**Dutch**	**English**
Preferred	Goed gedaan, het was een vogel.	Well done, it was a bird.	Helaas, je hebt een vogel aangeraakt. Laten we het nog eens proberen!	Unfortunately, you selected a bird. Let's try again!
Dispreferred	Goed gedaan.	Well done.	Helaas, dat is niet goed.	Unfortunately, that was not correct.
No feedback	-	-	-	-

TABLE 3 | The preferred feedback utterances.

Positive		Negative	
Dutch	**English**	**Dutch**	**English**
Goed gedaan!	Well done!	Helaas dat was niet goed.	Unfortunately, that was not correct.
Knap hoor.	Impressive.	Sorry deze is niet goed.	Sorry but this is not correct.
Ja goed gedaan!	Yes, well done!	Helaas, probeer het nog een keer.	Unfortunately, try again.
Ga zo door!	Keep going!	Jammer, we proberen het nog eens.	What a pity, let's try again.
Super!	Great!	Ah jammer, denk nog even goed na.	Ah pity, think again.
Heel knap gedaan.	Really impressive.	Super goed geluisterd, maar dat was niet goed, probeer het nog eens.	You listened very well, but this was not correct, try again.

The robot's feedback varied between six different options.

1. Variation. The robot used a variety of positive and negative feedback in the preferred feedback condition and no variation in the dispreferred feedback condition. We based the phrases on the student teachers' open-ended survey and can be found in **Table 3**. The robot randomly chose between six verbal phrases for positive feedback and negative feedback and the same phrase was never used twice in a row. We only added variation to the preferred strategy because the student teachers considered this crucial.
2. Extra attempt. The robot let children to try again after an incorrect answer in the preferred feedback condition and not in the other conditions. This was based on the student teachers' closed-ended answers where they relied heavily on the answer with the extra attempt. During the extra attempt, the tablet would only display the correct target word and the children's incorrect answer to help the children distinguish the two answers (see **Figure 2C**). After children correctly answered their second attempt, they received positive feedback.
3. Repetition. In the preferred condition, the robot would repeat the target word, either in addition to positive feedback or in addition to noting the mistake including the child's wrong answer. However, this was only done in 50% of all feedback to reduce the amount of repetition and because the student teachers did not always use a repetition (only in the closed-ended questionnaire and not in the open-ended questionnaire). The robot would only repeat the target word in the children's L1 (i.e., Dutch) to keep the amount of L2 exposure consistent over all children and to only focus on the effect of feedback.
4. Non-verbal feedback behavior. The robot used some non-verbal behavior when the child was correct in the preferred feedback condition, but not in the dispreferred feedback condition. This non-verbal behavior consisted of the robot nodding and displaying a rainbow colored pattern in the LED-eyes to indicate happiness.

After the feedback was provided (or after the child's answer in the no feedback condition), the game continued to the next target word.

3.6. Procedure

3.6.1. Robot Introduction and Pre-test

One week before the experiment, the children participated in a group introduction to familiarize themselves with the robot. During this introduction, based on Vogt et al. (2017), children learned how the robot moves and how to talk to it, and they played a game where they had to imitate the robot and they danced together. Unlike the robots during the experiment, this robot was not wearing a shirt. After this group introduction the children carried out a pre-test on their prior English knowledge in one-on-one sessions, as explained in section 3.3.2.

3.6.2. Experiment

At least 1 week after this group introduction and the pre-test, we started the first tutoring sessions with the children. Children participated in a quiet room away from their classrooms. After the child was collected from her or his classroom for the first session, he or she was told that he or she would play a game on a tablet with a friend of the introduction robot. This was repeated every new session so each child saw four "different" robots in total (one introduction robot and three robots in the tutoring sessions). When the child entered the room with the robot, the experimenter told the child to sit in front of the tablet next to the robot and started the experiment. After the child finished the

24 rounds of "I spy with my little eye" and the subsequent in-game post-test, the experimenter filled in a questionnaire with the child about the robot. When this questionnaire was completed the experimenter brought the child back to the classroom. This was repeated for three times with at least 1 day in between the different sessions.

The complete interaction was autonomous, except for the detection of children's speech when they repeated the target words as instructed. For detecting the child's speech, the experimenter would press a button on a control panel once the child had repeated the robot's utterance. The interaction was a one-on-one interaction, but the experimenter stayed in the same room to intervene when necessary. For example, when a child did not repeat after the robot, the experimenter would try to encourage the child to repeat after the robot. Moreover, when the child had a question, the experimenter would say that she did not know the answer and directed the child's attention back to the robot. In other cases, when a child had to go to the bathroom, the experimenter paused the experiment and walked with the child to the bathroom and back. The duration of each session was around 11 min (Preferred: M = 14 min, SD = 2 min, Dispreferred: M = 11 min, SD = 1.5 min, No feedback: M = 10 min, SD = 1 min).

3.6.3. Post-test

Two weeks after the last lesson, the children were collected from the classroom once more for the post-test.

3.7. Engagement Coding and Analyses

3.7.1. Engagement Coding

Engagement was determined by manual coding of half of the data. Before coding, the two raters followed a coding training and practiced with different videos. Each video was rated on a Likert scale from 1 to 9, with 1 being a low level of engagement and 9 being highly engaged. We measured *task-engagement* that includes the attention that the child payed to the robot as instructor, but also to the task displayed on the tablet screen. Children were fully engaged, when they were completely "absorbed" in the activity, were open for new information, were very motivated, enjoyed the task and wanted to play with the robot (Laevers, 2005). Additionally, we rated *robot-engagement* that measures the children's attention and interest at the robot as a social interaction partner. Children were fully engaged with the robot, when they were interacting with the robot as a social conversation partner.

The coding scheme was based on the ZIKO coding scheme (Laevers, 2005). The ZIKO scheme describes a measurement for, among others, children's engagement. It is designed for child-task engagement in open classroom settings. We adapted the scheme to include specific cues for this experiment by including cues such as, attention toward the experiment leader instead of the robot or tablet and child is randomly clicking on the tablet in order to continue.

Each engagement level had specific cues for the rater to look for. For example, children scored high on task-engagement when they were not only looking at the task and robot, but also actively searching for the different animals on the tablet and were fully committed to the task. In contrast, when children turned away from the robot and task, did not perform anything related to the task and were fiddling, this resulted in a low engagement. Children who played the game but did not pay all their attention to it received an average task-engagement rating. In the case of robot-engagement we added social engagement cues, such as looking at the robot, having spontaneous conversations with the robot, but it also included the children's posture toward the robot (a closed posture indicating a low robot engagement and an open posture indicating a high robot engagement).

For all specific cues and information, see the coding scheme in the **Supplementary Material** and on GitHub[3].

For the engagement coding, we pseudo-randomly selected half of the children, excluding children who took a break during the interaction (for example when they had to go to the bathroom), which happened in 11 cases. Twenty percent of the selected data was coded by two raters and their inter rater agreement was considered moderate to good using the intraclass correlation coefficient (ICC_{task} = 0.70, 95% CI[0.37, 0.76], ICC_{robot} = 0.80, 95% CI[0.62, 0.90]) (Koo and Li, 2016). For analyses, we only used the data of the first rater. We extracted two 2-min video fragments of the interaction: one at the beginning of the training rounds during the interaction and one at the end of the interaction.

The engagement rating of both fragments were combined to get a more reliable measure of the child's overall engagement during the lesson. This resulted in 210 engagement ratings in total.

3.7.2. Analyses

To investigate the effect of the different feedback strategies on children's engagement, we conducted a repeated measures ANOVA with the feedback strategy as the independent variable (three levels) and engagement as a dependent variable.

In addition, to investigate the effect of the feedback strategies on learning gain, we carried out a two-way repeated measures ANOVA with the children's scores as a dependent variable and two strategies: (1) feedback strategy (three levels) and (2) test moment (the pre-test and the delayed post-test).

Using planned contrasts, we compared the effect of preferred and dispreferred feedback with no feedback on engagement and learning gain for H1 and preferred feedback and dispreferred feedback for H2. Moreover, to investigate the effect of the feedback strategies on short-term learning gain, a one-way repeated measures ANOVA with feedback strategy as the independent variable and the results of the in-game test as the dependent variable was performed.

4. RESULTS

We have made the data set for this experiment publicly available[4]. In this section we report the children's engagement and their learning gain during the sessions. In addition, we report on

[3]https://github.nl/l2tor/codingscheme. Please note that the coding scheme rates children on a scale of 1–5 including half points, which we have converted to a scale of 1–9 for convenience.

[4]https://doi.org/10.34894/ZEIKLY

TABLE 4 | Average task- and robot-engagement rating over time (SD).

Feedback strategy	All lessons		Lesson 1		Lesson 2		Lesson 3	
	Task	Robot	Task	Robot	Task	Robot	Task	Robot
Preferred	6.17 (1.43)	6.14 (1.74)	6.77 (1.25)	6.85 (1.70)	6.15 (1.72)	6.30 (1.60)	5.54 (1.16)	5.25 (1.62)
Dispreferred	5.26 (1.48)	4.47 (1.80)	5.06 (0.98)	4.00 (1.25)	5.18 (1.59)	5.18 (2.05)	5.46 (1.69)	4.00 (1.66)
No feedback	5.27 (1.82)	4.74 (1.58)	6.00 (1.83)	5.21 (1.76)	5.41 (1.38)	4.41 (1.20)	4.10 (1.79)	4.45 (1.67)
Overall	5.57 (1.63)	5.12 (1.85)	6.07 (1.57)	5.45 (1.95)	5.53 (1.57)	5.26 (1.81)	5.10 (1.64)	4.56 (1.69)

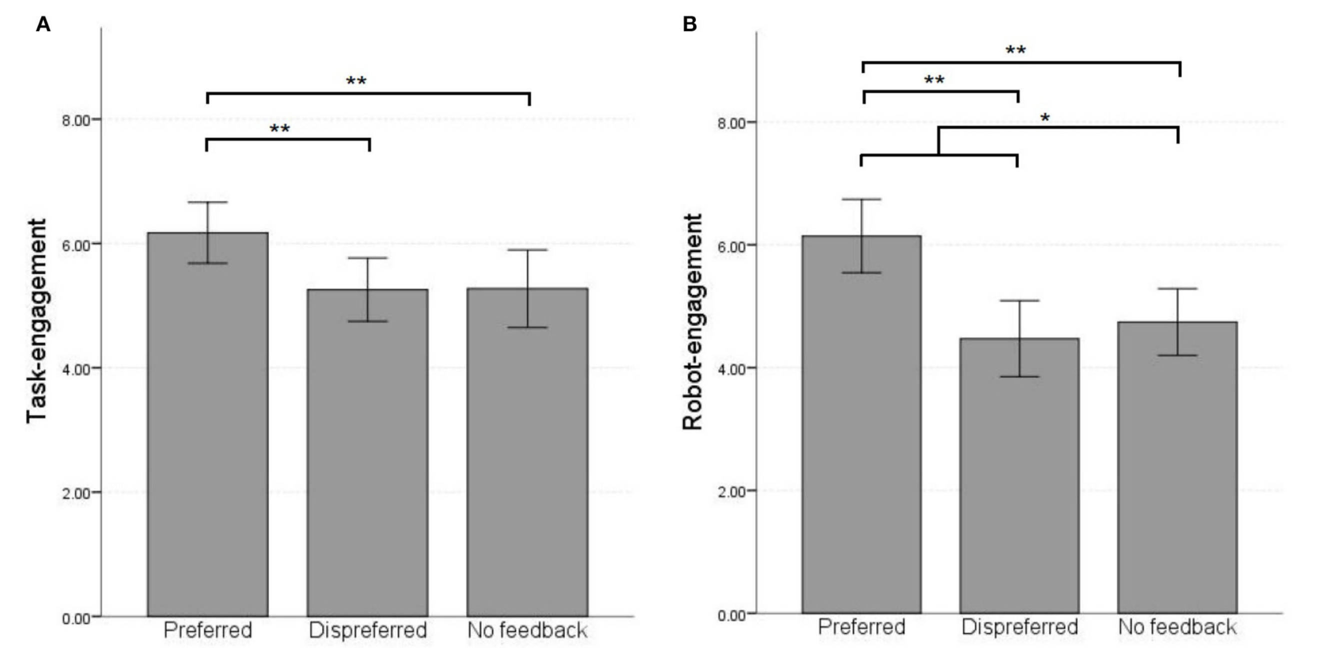


FIGURE 3 | Average engagement ratings per condition. Error bars show 95% confidence interval. *$p < 0.05$, **$p < 0.01$. **(A)** Task-engagement. **(B)** Robot-engagement.

the possible relation between learning gain and the children's engagement. Children received positive feedback during all 24 rounds in the preferred feedback condition and on average 14.30 times during the dispreferred feedback condition.

4.1. Engagement

Table 4 shows the overall results of both engagement types for the different lessons and different conditions. Overall, task-engagement ($M = 5.57$, $SD = 1.63$) was slightly higher than robot-engagement ($M = 5.12$, $SD = 1.85$). The two engagement types were moderately correlated [$r_{(105)} = 0.50, p < 0.01$], indicating that they both measure a different type of engagement.

4.1.1. Task-Engagement

Contrary to our expectations, planned contrast analyses for comparing both preferred feedback and dispreferred feedback combined ($M = 5.71, SD = 1.52$) with no feedback ($M = 5.27, SD = 1.82$) showed no significant difference in task-engagement [$F_{(1,\,34)} = 3.96, p = 0.06, \eta_p^2 = 0.10$]. However, as **Figure 3** shows, children are more engaged with preferred feedback ($M = 6.17, SD = 1.43$) than with dispreferred feedback [$M = 5.26, SD = 1.48$; $F_{(1,\,34)} = 13.49, p = 0.001, \eta_p^2 = 0.28$]. Further analysis using *post-hoc* comparisons with Bonferroni correction revealed that children were significantly more engaged in the preferred feedback condition than the no feedback condition [$t_{(34)} = 3.26, p = 0.003$, $M_{diff} = 0.9$]. There was no significant difference between dispreferred and no feedback [$t_{(34)} = -0.06, p = 0.96, M_{diff} = -0.01$].

Task-engagement dropped significantly over time (see **Figure 4**). A repeated measures ANOVA with a Huynh-Feldt correction was performed, because our data violated the assumption of sphericity. The analyses showed that task-engagement differed significantly between the lessons [$F_{(1.64,\,55.90)} = 7.16, p = 0.003, \eta_p^2 = 0.17$]. *Post-hoc* tests using the Bonferroni correction revealed that task-engagement dropped significantly between lesson 1 ($M = 6.07, SD = 1.56$) and 2 [$M = 5.53, SD = 1.57$; $t_{(34)} = 2.82, p =$

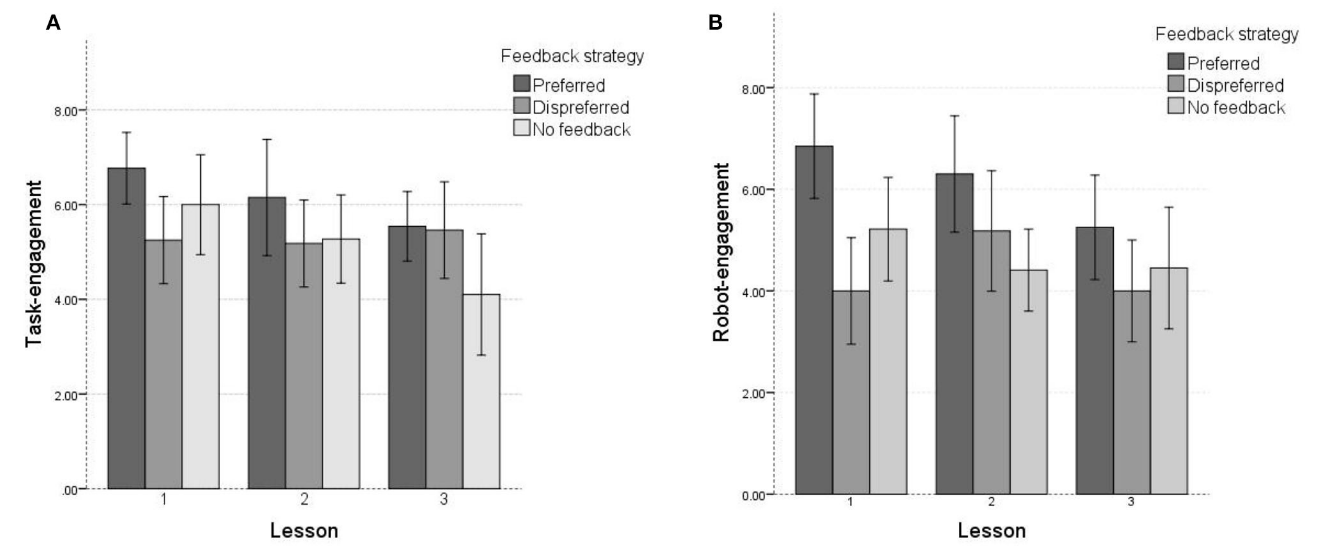


FIGURE 4 | Average task- and robot-engagement ratings over time and per condition. Error bars show 95% confidence interval. Note that a child who, for example, received preferred feedback in lesson 1 received different feedback in lesson 2 and in lesson 3. **(A)** Task-engagement. **(B)** Robot-engagement.

TABLE 5 | The task-engagement order effects visualized, a decreasing arrow shows decreasing task-engagement and visa versa.

Lesson 1		Lesson 2		Lesson 3
P	↘	**D**	↘	**N****
P	↘	**N**	↘	**D***
D	⟶	P	↘	N
D	⟶	N	⟶	P
N	⟶	P	⟶	D
N	⟶	D	↗	P

*P stands for preferred feedback, D for dispreferred feedback and N for no feedback. Task-engagement differed significantly for the first two orders with *indicating a p < 0.05 and **p < 0.01.*

$0.008, M_{diff} = 0.54$], and lesson 3 [$M = 5.10, SD = 1.64$; $t_{(34)} = 3.13, p = 0.004, M_{diff} = 0.97$] but not between lesson 2 and 3 [$t_{(34)} = 1.68, p = 0.102, M_{diff} = 0.43$].

We further tested whether there was an interaction effect between the feedback strategy and the session in which it was used. To this end, we used a mixed ANOVA with order as between factor and feedback strategy as within factor, because this accounts for the order in which participants received the different feedback strategies (for example, it might have had an influence on their task-engagement when they received no feedback first and the preferred feedback during the third session). There was a significant interaction effect between order and feedback strategy [$F_{(10, 58)} = 4.43, p < 0.001, \eta_p^2 = 0.433$] indicating that the effect of feedback on task-engagement varied as a function of when this feedback in the experiment it was administered taking into account that overall task-engagement decreased over time. As **Table 5** illustrates, children's task-engagement dropped over time, but not for all orders of the feedback strategies. The task-engagement dropped in most situations after children received preferred feedback, task-engagement never increased after dispreferred feedback and it either dropped or remained the same for no feedback. An exploratory repeated measures ANOVA on each order indicated that task-engagement differed significantly when preferred feedback ($M = 7, SD = 1.36$) was provided first, then dispreferred feedback ($M = 5.56, SD = 1.61$) and lastly no feedback [$M = 4.38, SD = 1.85; F_{(2, 14)} = 18.11, p < .001, \eta_p^2 = 0.72$] and furthermore, when preferred feedback ($M = 6.4, SD = 1.08$) was provided first, then no feedback ($M = 5.7, SD = 1.82$) and lastly dispreferred feedback [$M = 3.9, SD = 1.82; F_{(2, 8)} = 8.11, p = 0.012, \eta_p^2 = 0.67$]. All other orders did not differ significantly (all $p > 0.1$).

4.1.2. Robot-Engagement

Similarly as for task-engagement, we compared the average children's robot-engagement score during both the feedback conditions ($M = 5.31, SD = 1.95$) with the no feedback condition ($M = 4.74, SD = 1.58$) using planned contrast analyses. Unlike for task-engagement, we found a significant difference in robot-engagement between feedback and no feedback [$F_{(1, 34)} = 4.39, p = 0.044, \eta_p^2 = 0.11$], albeit with a relatively small effect size. Moreover, children scored higher

TABLE 6 | Average score per condition (SD).

Feedback strategy	Pre-test	Post-test	In-game
Preferred	1.88 (1.38)	2.71 (1.77)	2.80 (1.42)
Dispreferred	1.77 (1.28)	2.59 (1.65)	2.82 (1.62)
No feedback	2.00 (1.31)	2.55 (1.76)	2.75 (1.43)
Total	5.64 (2.00)	7.86 (4.10)	8.38 (3.20)

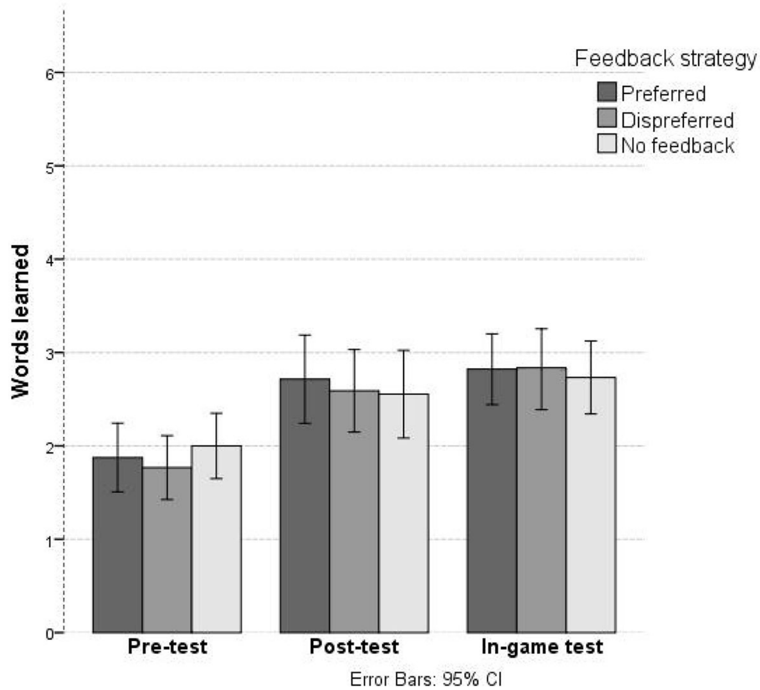


FIGURE 5 | Learning gain per condition. Error bars show 95% confidence interval.

for robot-engagement in the preferred feedback condition ($M = 6.14, SD = 1.74$) than in the dispreferred feedback condition [$M = 4.47, SD = 1.80$; $F_{(1,\ 34)} = 43.19, p < 0.01, \eta_p^2 = 0.56$]. Furthermore, *post-hoc* comparisons with Bonferroni correction revealed that children were significantly more engagement in the preferred feedback condition than in the no feedback condition [$t_{(34)} = 6.57, p < 0.01, M_{diff} = 1.40$]. There was no significant difference between robot-engagement in the dispreferred feedback condition and the no feedback condition [$t_{(34)} = 4.61, p = 1.0, M_{diff} = -0.27$].

As **Figure 4B** showed, robot-engagement also dropped over time. A repeated measures ANOVA showed a significant difference between the lessons [$F_{(2,\ 68)} = 4.56, p = 0.014, \eta_p^2 = 0.12$]. Again, note that the effect size is relatively small. Pairwise comparisons with a Bonferroni correction showed that robot-engagement dropped significantly between lesson 1 and 3 [$t_{(34)} = 2.67, p = 0.04, M_{diff} = 0.99$]. There was no significant difference between lesson 1 and lesson 2 [$t_{(34)} = 0.87, p = 1, M_{diff} = 0.29$] and lesson 2 and 3 [$t_{(34)} = 2.27, p = 0.09, M_{diff} = 0.7$].

Similarly as with task-engagement, we investigated whether there was an interaction effect between the feedback strategy and the lesson in which the feedback strategy was used. To test this, we used a mixed ANOVA with order as between factor and feedback strategy as within factor. For robot-engagement, there was no order effect [$F_{(10,\ 58)} = 1.58, p = 0.14$] which indicates that the children's robot-engagement was not influenced by different orders of feedback.

4.2. Learning Gain

Children made on average 9.75 mistakes during the 24 rounds (Preferred: $M = 9.95, SD = 5.56$; Dispreferred: $M = 9.30, SD = 5.22$; No feedback: $M = 9.75, SD = 5.41$). **Table 6** and **Figure 5** show the descriptive statistics for the target word knowledge scores for all conditions. Children performed above chance level in the pre-test [chance level = 4.5, $t_{(55)} = 4.27, p < 0.001, M_{diff} = 1.14$] and post-test [chance level = 2.61, $t_{(55)} = 9.58, p < 0.001, M_{diff} = 5.25$]. As expected, children performed better on the post-test than on the pre-test [$t_{(55)} = -3.88, p < 0.001, d = 0.52$], so children clearly learned some vocabulary.

The two-way repeated measures ANOVA with planned contrasts for both preferred feedback and dispreferred feedback (Pre-test: $M = 1.82, SD = 1.33$, Post-test: $M = 2.65, SD = 1.70$) showed no difference in learning gain compared to no feedback [Pre-test: $M = 2.00, SD = 1.31$, Post-test: $M = 2.55, SD = 1.76$); $F_{(1, 55)} = 0.47, p = 0.83$]. Furthermore, while children score numerically higher on word knowledge in the preferred feedback condition (Pre-test: $M = 1.88, SD = 1.38$, Post-test: $M = 2.71, SD = 1.77$) than in the dispreferred (Pre-test: $M = 1.77, SD = 1.28$, Post-test: $M = 2.59, SD = 1.65$), this difference was not significant [$F_{(1, 55)} = 0.45$, $p = 0.51$].

Table 6 also shows the results of the children's in-game tests. Children scored higher than chance in all conditions [chance level = 3, $t_{(55)} = 12.57, p < 0.001, M_{diff} = 5.38$]. Again, feedback strategy did not influence their learning gain, there were no significant differences [$F_{(2, 110)} = 0.122, p = 0.89$].

4.3. Relation Between Learning Gain and Engagement

To investigate whether there was a relation between both engagement types and learning gain, we performed a Pearson correlation analysis and in contrast with what we expected, we found no significant correlation between task-engagement and learning gain [Preferred: $r_{(35)} = 0.05, p = 0.78$, Dispreferred: $r_{(35)} = 0.09, p = 0.62$, No feedback: $r_{(35)} = 0.12, p = 0.50$]. Likewise, we did not find a significant correlation between robot-engagement and learning gain [Preferred: $r_{(35)} = 0.15, p = 0.40$, Dispreferred: $r_{(35)} = 0.09, p = 0.62$, No feedback: $r_{(35)} = 0.02, p = 0.90$].

5. DISCUSSION

The aim of this study was to understand the effects that different types of robot feedback have on children's engagement both with the task, the robot and their learning gain. We derived different types of feedback from a survey with student teachers and implemented them in three different robots, each robot teaching children words from a second language in a single session. One robot provided (teacher) preferred feedback, one provided (teacher) dispreferred feedback, and one provided no feedback at all. All children attended three sessions, each with a different feedback strategy. We studied how this choice of feedback influenced children's task- and robot-engagement and their learning gains.

5.1. Engagement

The analyses of both engagement types suggest that children seem to be generally engaged with the task and the robot during the three sessions. This accords with human studies indicating that feedback can make tasks encouraging and engaging (Henderlong and Lepper, 2002).

Contrary to our expectations, when the robot provided feedback (either preferred or dispreferred), this did not lead to increased task-engagement compared to when the robot provided no feedback (H1a). Children who received no feedback were, on average, rated as equally engaged as children who did receive feedback. However, the type of feedback did seem to have an influence on task-engagement of the children: children became more engaged with a robot that provided preferred feedback than with one that used dispreferred feedback or indeed no feedback (H2a). Moreover, the robot's feedback did result into a higher robot-engagement compared to no feedback (H1a). Children who received feedback (either preferred or dispreferred), were rated more engaged with the robot than children who did not receive any feedback. However, it is worth pointing out that the numeric effects for robot- and task-engagement were rather comparable, even though the former but not the latter was found to be statistically significant. Similar to task-engagement, children were most engaged with a robot that provided preferred feedback (H2a) in comparison to dispreferred and no feedback. Interestingly, the difference between robot-engagement for preferred feedback and dispreferred feedback was larger than the difference for task-engagement.

Preferred and dispreferred feedback differed on multiple aspects (variation, extra attempt, repetition of answer, non-verbal behavior) and when combined, these factors seem to have an influence on engagement. While it is hard to identify exactly to what extent each of these factors contribute to children's task- and robot-engagement, we believe that some aspects might have had a larger effect on both engagement types than others.

For example, variation in feedback, as is realized in the preferred feedback condition, could have had relatively strong effect on children's task- and robot-engagement. A robot that provides more variation in the way feedback is offered could spark children's interest and keep them interested and motivated in continuing the task over a longer period of time and at the same time also make them more interested in the robot. In contrast, a robot who continually uses the same feedback phrase or no feedback at all might have a negative impact on children's interest in the robot and their robot-engagement and moreover reduce their motivation to continue with a task and, thus, be less successful in keeping them task-engaged.

It is furthermore possible that the extra attempt after an incorrect answer in the children's L1 may have task-engaged the children more in the preferred feedback condition than in the other two conditions. The fact that children heard the correct L1 word, could try again and received praise afterwards, may have had a positive effect on their task-engagement. This is in line with how teachers tend to provide feedback, praising demotivated children to try to engage them again (Hattie and Timperley, 2007). Some children also mentioned the extra attempt as the robot helping them getting the correct answer, this might increase their sense of relatedness to the robot which could have increased their robot-engagement.

Lastly, the non-verbal communication of the robot in the preferred condition may have increased children's robot-engagement as well. The robot displayed rotating colored eyes and nodded each time when children were correct. This is in agreement with the results of Morris and Zentall (2014), who found that children showed more intrinsic motivation when the robot used non-verbal behaviors such as thumbs up, and the findings of Serholt and Barendregt (2016), who found that children paid most attention to the robot when it provided feedback accompanied by an arm gesture. Future studies that take variation of feedback in combination with different types of non-verbal behavior into account will be needed to develop a full picture of this finding (De Wit et al., 2020). Besides gesturing, also gaze is a known non-verbal factor that can influence engagement (Mwangi et al., 2018). However, in the current experiment gaze was not factor of interest, since the robot's gaze behavior was identical in all three conditions.

As mentioned, it is not possible with the current experiment to determine which factor had the largest effect on task-engagement or robot-engagement. For this more research is needed. In the current experiment, we explored to what extent by student teachers preferred feedback strategy would differ from a dispreferred feedback strategy or no feedback strategy. We found that preferred feedback has a beneficial effect on both engagement types. However, to identify the effect of different factors that define the preferred feedback strategy has on engagement and which factor contribute to which engagement type, future experiments could be set up in which each factor is varied between conditions.

Also consistent with other studies is that both task- and robot-engagement seemed to drop over time (Kanda et al., 2007; Coninx et al., 2015; De Wit et al., 2018), and this drop appeared to be similar for all three conditions, although the differences between the conditions stayed over time. Adding more variation to the robot's feedback, as well as varying other parts of its behavior, might help reduce a drop in engagement. Ahmad et al. (2019) suggested that children seemed to stay engaged with a robot that is adaptive, which lends some support to the importance of individualized variation.

Interestingly, we found an interaction effect between task-engagement and the order of feedback strategies but not between robot-engagement and order. In particular, we observed that children's task-engagement dropped after receiving preferred feedback and that their task-engagement was similar or lower before receiving preferred feedback. Receiving no feedback or dispreferred feedback might have demotivated children, and, conversely, receiving various feedback information on their performance, might have increased their motivation again and therefore their task-engagement. Visa versa, after children received preferred feedback and continued in the dispreferred or no feedback condition, their task-engagement decreased again. However, some caution to this explanation must be applied, as the findings might have been influenced by individual differences as well.

5.2. Learning Gain

As expected, children learned from all three sessions with the robot. They did not learn many words per session though, which is in line with previous research with this young age group (Westlund and Breazeal, 2015; Vogt et al., 2019). Our results also show that these learning effects were retained in the longer run, because we conducted a post-test 2 weeks after the last session, suggesting that the target words remained in children's memory (Axelsson et al., 2016).

Contrary to our expectations, children did not learn more in the feedback conditions than when receiving no feedback (H1b), nor did it matter for the learning gain whether feedback was of the preferred or dispreferred variety (H2b). This was not only the case for the post-test, but also applied to the in-game test that was taken immediately after each training round.

What these results suggest is that children could learn from the teaching sessions without the need for feedback, and that the contribution of feedback to learning might have been smaller than we anticipated. This can be explained by the fact that children could rely on cross-situational learning (Smith and Yu, 2008), because children saw four depictions of possible meanings each time they heard a target word, with the distractors changing while the target stayed the same across situations. Hence, children could infer the meaning of a target based on the co-variation in meanings offered with the different occurrences of the target word, which seems to largely drive the learning, and feedback does not appear to contribute to this learning process.

It is conceivable that the learning task itself might have been too easy for the children to really benefit from the feedback. Moreover, since the children could press any animal they wanted to go forward in the game, they did not have to pay attention to the feedback of the robot. For future research, it would be interesting to conduct a study in such a way that feedback becomes more central to the interaction or more content-related, and where the learning task is more complex (e.g., learning about difficult sentence structures or unfamiliar grammar). This might shed further light on the influence of feedback on learning in child-robot interaction.

It is interesting to note that we did not observe learning differences between preferred and dispreferred feedback, which might be due to the feedback being completely offered in the children's L1. As a result, children did not receive a explicit translation between L1 and L2 as part of their (corrective) negative feedback. This might explain why children did not learn the L2 translation of a concept better during negative preferred feedback. It seems plausible that the addition of L2 to the negative (corrective) feedback would have resulted in higher learning gains (Hall, 2002; Scott and de la Fuente, 2008). However, we did not add this L2-L1 translation to our negative feedback for methodological reasons to keep the different conditions comparable. In particular, we made sure that there was an identical number of L2 exposures in every condition, since the number of L2 exposures could also affect learning (Ellis, 2002).

5.3. Relation Between Engagement and Learning

Various studies have found that increased engagement leads to better learning performance (Christenson et al., 2012). However, in our data we did not observe a relation between task- or robot-engagement and learning. Children who were more engaged with the task or with the robot did not learn more words than children who were less engaged. This might be due to the relatively small learning gain of children in the different conditions. They learned on average close to two out of six words during each session and this might not have been enough to observe a correlation with both engagement types. Moreover, it is conceivable that individual differences between children might have played a role as well. Effects of engagement on learning seemed to differ substantially from one child to the next, which is consistent with earlier research with this age group interacting with a robot (De Haas et al., 2017). Finally, we conjecture that in future research with more varied and more prominent feedback (along the lines sketched above), we might indeed observe that more engagement leads to better learning results.

5.4. Strengths and Limitations

This study has at least four strengths: First, we systematically compared different feedback strategies, derived from actual strategies suggested by young student teachers. Second, we tested a large group of young children to measure the effects of feedback. Third, the study was a carefully constructed experiment, of which all hypotheses and analyses have been preregistered (Simmons et al., 2011). Fourth, we measured two types of engagement to account for the children's engagement with the task and with their engagement with the robot as social partner.

Our study has also at least four limitations. First, we only measured comprehension and not active production of words. However, as speech recognition of the robot is not reliable yet, a more interactive task would have to rely fully on the experimenter in a Wizard of Oz setting (Kennedy et al., 2016). Since we aimed for an autonomously operating system, our task was designed to teach only passive understanding of L2 by using a tablet to record children's responses.

Second, our task was very repetitive. The only variation we introduced was the feedback that the robot would provide in the preferred feedback condition. Children did not have control over when to play with the robot and they were not able to change the task. It is a challenge to design a task that is adaptive to children's preferences, while still being educationally responsible and technical feasible. Providing such autonomy to children could increase their intrinsic motivation, which would increase their engagement and their learning performance (Ryan and Deci, 2000; van Minkelen et al., 2020).

Third, the robot could not react to the children's perceived engagement level during the experiment. While a human teacher would constantly monitor children's engagement and adapt the task accordingly to make it more personalized, the robot in our experiment simply continued to the next word and kept the interaction the same throughout all sessions, disregarding the child's engagement. Being able to automatically recognize a child's engagement would allow the robot to personalize feedback and other behaviors based on this engagement (Gordon et al., 2016; Ahmad et al., 2019).

Finally, we investigated the main effect of feedback on engagement and learning gain and showed that the preferred feedback had an influence on engagement with the task and with the robot. However, preferred and dispreferred feedback varied on multiple factors (variation, extra attempt, repetition of answer, non-verbal behavior), and consequently we cannot attribute the effect on engagement to only one of these factors, only the combination. Future research should look at individual aspects of feedback if technically feasible to measure the effectiveness for engagement.

6. CONCLUSION

The study presented in this paper explored whether robot feedback affects children's task- and robot-engagement and learning gain in second language learning. We compared three robot behaviors: one based its feedback on student teachers' preferred feedback strategies, one that did the opposite and one that did not use any feedback. The preferred strategy varied its feedback, gave children an additional attempt when they answered incorrectly, repeated the target word and gave non-verbal feedback. In contrast, the dispreferred feedback strategy did not vary its feedback, did not provide children with an additional attempt, did not repeat the target word and did not give non-verbal feedback. We found that children in the preferred feedback condition were more engaged than children in the dispreferred feedback and no feedback conditions, both with the task as with the robot. However, the feedback strategy did not influence children's learning gain; they did not retain more word knowledge with one of the different conditions. Moreover, we did not observe a relation between learning and engagement.

Our results are especially interesting for long-term interactions where engagement of children often drops. Providing feedback in an even more varied and motivating manner might help children to remain engaged in long-term scenarios. We expect that in the long-term such varied and motivating feedback can also improve children's learning gains, especially when the learning tasks become more difficult and children cannot just learn from inferring associations through cross-situational learning.

DATA AVAILABILITY STATEMENT

All datasets generated for this study are included in the article/**Supplementary Material**.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Research Ethics and Data Management Committee of Tilburg School of Humanities and Digital Sciences. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin.

AUTHOR CONTRIBUTIONS

MH, PV, and EK contributed to the conception and design of the study. MH developed, conducted, analyzed the study, and wrote the first draft of the paper. PV and EK supervised and critically reviewed the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work has been supported by the EU H2020 L2TOR project (grant 688014).

ACKNOWLEDGMENTS

We would like to thank the children, their parents, and the schools for their participation. Furthermore, we would like to thank Marlies de Groot for lending her voice for the pre- and post-test, Jan de Wit for his help in the implementation, Mehrnaz Pour Morshed for her help collecting the data, and Reinjet Oostdijk for her help in annotating the videos.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/frobt.2020.00101/full#supplementary-material

REFERENCES

Ahmad, M. I., Mubin, O., Shahid, S., and Orlando, J. (2019). Robot's adaptive emotional feedback sustains children's social engagement and promotes their vocabulary learning: a long-term child-robot interaction study. *Adapt. Behav.* 27, 243–266. doi: 10.1177/1059712319844182

Axelsson, E. L., Williams, S. E., and Horst, J. S. (2016). The effect of sleep on children's word retention and generalization. *Front. Psychol.* 7:1192. doi: 10.3389/fpsyg.2016.01192

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018). Social robots for education: a review. *Sci. Robot.* 3:eaat5954. doi: 10.1126/scirobotics.aat5954

Blanck, P. D., Reis, H. T., and Jackson, L. (1984). The effects of verbal reinforcement of intrinsic motivation for sex-linked tasks. *Sex Roles* 10, 369–386. doi: 10.1007/BF00287554

Blumenfeld, P. C., Kempler, T. M., and Krajcik, J. S. (2006). *Motivation and Cognitive Engagement in Learning Environments*. New York, NY: Cambridge University Press, 475–488. doi: 10.1017/CBO9780511816833.029

Boyer, K. E., Phillips, R., Wallis, M., Vouk, M., and Lester, J. (2008). "Balancing cognitive and motivational scaffolding in tutorial dialogue," in *Intelligent Tutoring Systems*, eds B. P. Woolf, E. Aïmeur, R. Nkambou and S. Lajoie (Berlin; Heidelberg: Springer), 239–249. doi: 10.1007/978-3-540-69132-7_28

Christenson, S. L., Wylie, C., and Reschly, A. L. (2012). *Handbook of Research on Student Engagement*. New York, NY: Springer Science & Business Media. doi: 10.1007/978-1-4614-2018-7

Coninx, A., Baxter, P., Oleari, E., Bellini, S., Bierman, B., Blanson Henkemans, O., et al. (2015). Towards long-term social child-robot interaction: using multi-activity switching to engage young users. *J. Hum. Robot Interact.* 5:32. doi: 10.5898/JHRI.5.1.Coninx

Corpus, J. H., and Lepper, M. R. (2007). The effects of person versus performance praise on children's motivation: gender and age as moderating factors. *Educ. Psychol.* 27, 487–508. doi: 10.1080/01443410601159852

De Haas, M., Baxter, P., De Jong, C., Krahmer, E., and Vogt, P. (2017). "Exploring different types of feedback in preschooler and robot interaction," in *ACM/IEEE International Conference on Human-Robot Interaction* (Vienna), 127–128. doi: 10.1145/3029798.3038433

De Wit, J., Brandse, A., Krahmer, E., and Vogt, P. (2020). "Varied human-like gestures for social robots: investigating the effects on children's engagement and language learning," in *ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, UK). doi: 10.1145/3319502.3374815

De Wit, J., Schodde, T., Willemsen, B., Bergmann, K., De Haas, M., Kopp, S., et al. (2018). "The effect of a robot's gestures and adaptive tutoring on children's acquisition of second language vocabularies," in *ACM/IEEE International Conference on Human-Robot Interaction* (Vienna), 50–58. doi: 10.1145/3171221.3171277

Deci, E. L., and Ryan, R. M. (1985). The general causality orientations scale: self-determination in personality. *J. Res. Pers.* 19, 109–134. doi: 10.1016/0092-6566(85)90023-6

Deci, E. L., Vallerand, R. J., Pelletier, L. G., and Ryan, R. M. (1991). Motivation and education: the self-determination perspective. *Educ. Psychol.* 26, 325–346. doi: 10.1080/00461520.1991.9653137

Dempsey, J. V. and Sales, G. C. (1993). *Interactive Instruction and Feedback*. Englewood Cliffs, NJ: Educational Technology Publications. doi: 10.5555/563081

Dörnyei, Z. (1998). Motivation in second and foreign language learning. *Lang. Teach.* 31, 117–135. doi: 10.1017/S026144480001315X

Ellis, N. C. (2002). Frequency effects in language processing: a review with implications for theories of implicit and explicit language acquisition. *Stud. Second Lang. Acquis.* 24, 143–188. doi: 10.1017/S0272263102002024

Gordon, G., Spaulding, S., Westlund, J. K., Lee, J. J., Plummer, L., Martinez, M., et al. (2016). "Affective personalization of a social robot tutor for children's second language skills," in *30th AAAI Conference on Artificial Intelligence, AAAI 2016* (Phoenix, AZ), 3951–3957.

Hall, C. J. (2002). The automatic cognate form assumption: evidence for the parasitic model of vocabulary development. *Int. Rev. Appl. Linguist. Lang. Teach.* 40, 69–87. doi: 10.1515/iral.2002.008

Hamilton, V. J., and Gordon, D. A. (1978). Teacher-child interactions in preschool and task persistence. *Am. Educ. Res. J.* 15, 459–466. doi: 10.3102/00028312015003459

Hattie, J., and Timperley, H. (2007). The power of feedback. *Rev. Educ. Res.* 77, 81–112. doi: 10.3102/003465430298487

Henderlong, J., and Lepper, M. R. (2002). The effects of praise on children's intrinsic motivation: a review and synthesis. *Psychol. Bull.* 128, 774–795. doi: 10.1037/0033-2909.128.5.774

Kanda, T., Sato, R., Saiwaki, N., and Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term human. *IEEE Trans. Robot.* 23, 962–971. doi: 10.1109/TRO.2007.904904

Kanero, J., Geçkin, V., Oranç, C., Mamus, E., Küntay, A. C., and Göksun, T. (2018). Social robots for early language learning: current evidence and future directions. *Child Dev. Perspect.* 12, 146–151. doi: 10.1111/cdep.12277

Kennedy, J., Baxter, P., and Belpaeme, T. (2015). "The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning," in *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Portland, OR), 67–74. doi: 10.1145/2696454.2696457

Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2016). "Social robot tutoring for child second language learning," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (Christchurch: IEEE Press), 231–238. doi: 10.1109/HRI.2016.7451757

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., et al. (2017). "Child speech recognition in human-robot interaction: evaluations and recommendations," in *ACM/IEEE International Conference on Human-Robot Interaction* (Vienna), 82–90. doi: 10.1145/2909824.3020229

Kluger, A. N., and DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol. Bull.* 119, 254–284. doi: 10.1037/0033-2909.119.2.254

Koo, T. K., and Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J. Chiropract. Med.* 15, 155–163. doi: 10.1016/j.jcm.2016.02.012

Laevers, F. (2005). *Well-being and Involvement in Care Settings. A Process-oriented Self-evaluation Instrument (SIC's)*. Leuven: Kind En Gezin; Research Centre for Experiential Education.

Long, M. H. (2006). *Problems in SLA (Second Language Acquisition Research)*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lyster, R., and Ranta, L. (1997). Corrective feedback and learner uptake: negotiation of form in communicative classrooms. *Stud. Second Lang. Acquis*. 19, 37–66. doi: 10.1017/S0272263197001034

Mackey, A., and Silver, R. E. (2005). Interactional tasks and English L2 learning by immigrant children in Singapore. *System* 33, 239–260. doi: 10.1016/j.system.2005.01.005

Mazzoni, E., and Benvenuti, M. (2015). A robot-partner for preschool children learning english using socio-cognitive conflict. *Educ. Technol. Soc*. 18, 474–485. Available online at: https://www.jstor.org/stable/jeductechsoci.18.4.474

Morris, B. J., and Zentall, S. R. (2014). High fives motivate: the effects of gestural and ambiguous verbal praise on motivation. *Front. Psychol*. 5:928. doi: 10.3389/fpsyg.2014.00928

Mwangi, E., Barakova, E. I., Díaz, M., Mallofré, A. C., and Rauterberg, M. (2018). "Dyadic gaze patterns during child-robot collaborative gameplay in a tutoring interaction," in *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, (Nanjing), 856–861. doi: 10.1109/ROMAN.2018.8525799

Park, H. W., Gelsomini, M., Lee, J. J., and Breazeal, C. (2017). "Telling stories to robots: the effect of backchanneling on a child's storytelling," in *ACM/IEEE International Conference on Human-Robot Interaction* (Vienna), 100–108. doi: 10.1145/2909824.3020245

Resing, W. C., Bakker, M., Elliott, J. G., and Vogelaar, B. (2019). Dynamic testing: can a robot as tutor be of help in assessing children's potential for learning? *J. Comput. Assist. Learn*. 35, 540–554. doi: 10.1111/jcal.12358

Rintjema, E., van den Berghe, R., Kessels, A., de Wit, J., and Vogt, P. (2018). "A robot teaching young children a second language: the effect of multiple interactions on engagement and performance," in *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (Chicago, IL: ACM), 219–220. doi: 10.1145/3173386.3177059

Ryan, R. M., and Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *Am. Psychol*. 55:68. doi: 10.1037/0003-066X.55.1.68

Schodde, T., Hoffmann, L., Stange, S., and Kopp, S. (2019). Adapt, explain, engage—a study on how social robots can scaffold second-language learning of children. *J. Hum.-Robot Interact.* 9. doi: 10.1145/3366422

Schrooten, W., and Vermeer, A. (1994). *Woorden in Het Basisonderwijs. 1500 Woorden Aangeboden Aan Leerlingen*. [*Words in Primary Education. 1500 Words Presented to Students*]. Tilburg: Tilburg University Press.

Scott, V. M., and de la Fuente, M. J. (2008). What's the problem? L2 learners' use of the L1 during consciousness-raising, form-focused tasks. *Modern Lang. J*. 92, 100–113. doi: 10.1111/j.1540-4781.2008.00689.x

Serholt, S., and Barendregt, W. (2016). "Robots tutoring children: longitudinal evaluation of social engagement in child-robot interaction," in *Proceedings of the 9th Nordic Conference on Human-Computer Interaction, NordiCHI '16* (New York, NY: ACM), 64:1–64:10. doi: 10.1145/2971485.2971536

Shute, V. J. (2008). Focus on formative feedback. *Rev. Educ. Res*. 78, 153–189. doi: 10.3102/0034654307313795

Simmons, J. P., Nelson, L. D., and Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychol. Sci*. 22, 1359–1366. doi: 10.1177/0956797611417632

Smith, L., and Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* 106, 1558–1568. doi: 10.1016/j.cognition.2007.06.010

van Minkelen, P., Gruson, C., van Hees, P., Willems, M., De Wit, J., Aarts, R., et al. (2020). "Using self-determination theory in social robots to increase motivation in L2word learning," in *ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, UK). doi: 10.1145/3319502.3374828

Vogt, P., De Haas, M., De Jong, C., Baxter, P., and Krahmer, E. (2017). Child-robot interactions for second language tutoring to preschool children. *Front. Hum. Neurosci*. 11:73. doi: 10.3389/fnhum.2017.00073

Vogt, P., Van Den Berghe, R., De Haas, M., Hoffman, L., Kanero, J., Mamus, E., et al. (2019). "Second language tutoring using social robots: a large-scale study," in *ACM/IEEE International Conference on Human-Robot Interaction* (Daegu), 497–505. doi: 10.1109/HRI.2019.8673077

Westlund, J. K., and Breazeal, C. (2015). "The interplay of robot language level with children's language learning during storytelling," in *ACM/IEEE International Conference on Human-Robot Interaction* (New York, NY: ACM), 65–66. doi: 10.1145/2701973.2701989

Wojitas, O. (1998). Feedback? No, just give us the answers. *Times Higher Educ. Suppl*. 25:109162.

Zentall, S. R., and Morris, B. J. (2010). "Good job, you're so smart": the effects of inconsistency of praise type on young children's motivation. *J. Exp. Child Psychol*. 107, 155–163. doi: 10.1016/j.jecp.2010.04.015

Zentall, S. R., and Morris, B. J. (2012). A critical eye: praise directed toward traits increases children's eye fixations on errors and decreases motivation. *Psychonom. Bull. Rev*. 19, 1073–1077. doi: 10.3758/s13423-012-0294-y

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer EB declared a past co-authorship with one of the authors MH to the handling Editor.

frontiers in Robotics and AI

ORIGINAL RESEARCH
published: 07 October 2020
doi: 10.3389/frobt.2020.00099

A Comparison of Social Robot to Tablet and Teacher in a New Script Learning Context

*Zhanel Zhexenova[1], Aida Amirova[2], Manshuk Abdikarimova[1], Kuanysh Kudaibergenov[1], Nurakhmet Baimakhan[1], Bolat Tleubayev[1], Thibault Asselborn[3], Wafa Johal[3,4], Pierre Dillenbourg[3], Anna CohenMiller[2] and Anara Sandygulova[1]**

[1] Department of Robotics and Mechatronics, School of Engineering and Digital Sciences, Nazarbayev University, Nur-Sultan, Kazakhstan, [2] Graduate School of Education, Nazarbayev University, Nur-Sultan, Kazakhstan, [3] CHILI Lab, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, [4] School of Computer Science and Engineering, University of New South Wales, Sydney, NSW, Australia

This research occurred in a special context where Kazakhstan's recent decision to switch from Cyrillic to the Latin-based alphabet has resulted in challenges connected to teaching literacy, addressing a rare combination of research hypotheses and technical objectives about language learning. Teachers are not necessarily trained to teach the new alphabet, and this could result in a challenge for children with learning difficulties. Prior research studies in Human-Robot Interaction (HRI) have proposed the use of a robot to teach handwriting to children (Hood et al., 2015; Lemaignan et al., 2016). Drawing on the Kazakhstani case, our study takes an interdisciplinary approach by bringing together smart solutions from robotics, computer vision areas, and educational frameworks, language, and cognitive studies that will benefit diverse groups of stakeholders. In this study, a human-robot interaction application is designed to help primary school children learn both a newly-adopted script and also its handwriting system. The setup involved an experiment with 62 children between the ages of 7–9 years old, across three conditions: a robot and a tablet, a tablet only, and a teacher. Based on the paradigm—learning by teaching—the study showed that children improved their knowledge of the Latin script by interacting with a robot. Findings reported that children gained similar knowledge of a new script in all three conditions without gender effect. In addition, children's likeability ratings and positive mood change scores demonstrate significant benefits favoring the robot over a traditional teacher and tablet only approaches.

Keywords: human-robot interaction, child learning, language learning, social robot, cognitive learning theory, learning by teaching, interdisciplinary

OPEN ACCESS

Edited by:
Agnese Augello, Institute for High Performance Computing and Networking (ICAR), Italy

Reviewed by:
Michelle Neumann, Griffith University, Australia
Goren Gordon, Tel Aviv University, Israel

****Correspondence:***
Anara Sandygulova
anara.sandygulova@nu.edu.kz

Specialty section:
This article was submitted to Human-Robot Interaction, a section of the journal Frontiers in Robotics and AI

Received: *23 April 2020*
Accepted: *17 June 2020*
Published: *07 October 2020*

Citation:
Zhexenova Z, Amirova A, Abdikarimova M, Kudaibergenov K, Baimakhan N, Tleubayev B, Asselborn T, Johal W, Dillenbourg P, CohenMiller A and Sandygulova A (2020) A Comparison of Social Robot to Tablet and Teacher in a New Script Learning Context. Front. Robot. AI 7:99. doi: 10.3389/frobt.2020.00099

1. INTRODUCTION

The gradual transition of the Kazakh alphabet from Cyrillic to the Latin script was first introduced by the Kazakhstani government in late October 2017 (Altynsarina, 2018). This stage-by-stage transfer to Latin is expected to be fully implemented by 2025 (Presidential decree, 2017). Considering the explicitly formulated rationales and objectives for this reform, it is essential to pay attention to teaching all populations literacy skills in the Latin script. Even though it is thought that learning a new script will be effortless owing to the knowledge of English or other linguae mundi, there are various threats when facing the transition, such as decreased motivation to develop basic literacy skills in the Latin-based Kazakh among youth and elderly populations (Kadirova, 2018).

Since most language reforms are grounded with a certain purpose in mind, their resultant impact on literacy (Crisp, 1990), identity (Hatcher, 2008), and education in general, need to be taken into account to ease the change. Following the transfer to the Latin-based Kazakh alphabet and subsequent necessity for acquiring knowledge of the script, innovative approaches and instruments can facilitate a smooth Latin switch-over for teaching and learning. Many innovative solutions are being implemented for the purposes of educational applicability (Mubin et al., 2013) for early language and literacy learning (Neumann, 2020), handwriting learning (Hood et al., 2015), or foreign language acquisition (Balkibekov et al., 2016). For instance, Sysoev et al. (2017) presented SpeechBlocks, which is an application assisting young learners in their pursuit of mastering spelling strategies through listening to the differently positioned letters in a word. The use of this application accelerates children's engagement, self-confidence, and autonomy in learning. Furthermore, Dewi et al. (2018) developed a Javanese script learning application for Indonesian elementary age children, which made script learning easy to understand and engage with. Similarly, Yanikoglu et al. (2017) revealed that tablet-based learning supplemented by handwriting recognition and automatic evaluation was more preferred among first-graders compared to paper-based learning.

Furthermore, in recent years, research has provided a huge space for the area of language acquisition deploying social robots (Tazhigaliyeva et al., 2016; Belpaeme et al., 2018b) and this, in turn, was an impetus to the rise of human-robot interaction (HRI) as a promising research field (Mubin et al., 2013). It has created new opportunities for the integration of social robots into educational settings. To date, one of the original approaches to language acquisition and new language scripts is the Swiss-based CoWriter project. It has a clear target to assist children to learn handwriting with a social robot on the basis of learning by using a teaching approach (LbT) (Hood et al., 2015; Jacq et al., 2016; Lemaignan et al., 2016). Since the development of these studies, others have effectively employed the robot-assisted LbT approach to other fields of inquiry (Jamet et al., 2018; Yadollahi et al., 2018).

Central to our study is the CoWriting Kazakh system, which integrates a humanoid NAO robot and a tablet with a digital pen. In this scenario, the robot interacts with learners as a social partner. It is programmed to show enthusiasm to learn Kazakh language. Moreover, since the robot is programmed to speak English, the child needs to translate basic phrases from English into Kazakh (e.g., "hello—sálem"). In this way, the child takes on the role of the robot's teacher of the Kazakh language. As the children engage with the robot, they show, or "teach" the robot how to write the words in Latin-based Kazakh script. In other words, the child is recognized as a "more knowledgeable other" who leads the learning process as a teacher and peer (Vygotsky, 1980; Huong, 2007). Thus, their interaction includes a child-robot cooperation in writing words in turns where the robot's spelling of the Latin-based Kazakh words is programmed to always be correct. While not an expert, the child's expertise in comparison to the programmed robot provides an avenue for learning through teaching.

In order to investigate whether a social robot is important in the CoWriting Kazakh system, this paper aims to contribute to the literature on human-computer/robot interaction by comparing different learning aids, such as robot and tablet, tablet only, and a traditional teacher to see which teaching method is the most effective in terms of new script learning gains. We believe that by purposefully integrating an interdisciplinary lens involved in the system, inspired by pedagogy, cognitive science, and linguistics, will enhance an understanding of the research and the associated learning gains. In this sense, we have to seek out and discuss other perspectives and theories in order to offer effective learning scenarios that might increase children's learning outcomes. Thus, this paper also deals with current theories from different research fields to embed them into the CoWriter Kazakh system, evaluated by their effectiveness on children's learning experiences. This interdisciplinary nature of the study allows us to expand our understanding of a complex issue from different angles (Klein, 1990; CohenMiller and Pate, 2019). Using the human-robot interaction framework, the CoWriting Kazakh learning scenario will reduce the boundaries between various disciplinary fields and contribute to the area of new literacy studies.

2. RELATED WORK

2.1. Transliteration and Script Learning

In an increasingly globalized world, an English-related writing system is gaining popularity for use across languages. One language may use more than one writing system, such as the Kazakh language written both in Cyrillic and in Latin-based alphabets (see **Figure 1** for a comparison between the scripts). This phenomenon is known as "digraphia," which comprises English-related Latin (or Roman) script to constitute another language (e.g., Kazakh) (Rivlina, 2016). Roman-Cyrillic script-alternation is an example of "biscriptal" practices that are used to associate transliterated written language. For instance, the Kazakh word for naming "door" can be written either "ecik" in Cyrillic or "esik" in Latin. Rivlina (2016) broadly discussed the sociolinguistic phenomenon of employing Latin script alongside Cyrillic script to represent Russian written discourse. Building on the results of a web scraping analysis, the authors reached a conclusion that digraphic practices are used to visually draw people's attention to the written texts and to strengthen recognition and memorability by playing with words. It is also emphasized that digraphia produces translingual effects that can eliminate boundaries in terms of linguistic, national, cultural, and domain aspects.

Another study performed by Al-Azami et al. (2010) examines the effectiveness of the script conversion (i.e., transliteration) as a learning tool for writing in Bengali. In schools in London, this method is adopted to teach British-Bangladeshi students between the ages of 7–11. This Bengali-Roman biscriptal switch converts speech into text, helping children to communicate with parents and teachers, and importantly to practice a new method of increasing bilingual skills. To illustrate, if students do not recognize the correct spelling of a certain word, transliteration allows them to visualize the word, and students could grasp

English	A	-	B	D	E	F	G	-	H	I	-	J	K	L	M	N	-	O	-	P	Q	R	S	T	U	-	V	Y	-	Z	Sh	Ch
Latin Kazakh	A	Á	B	D	E	F	G	Ǵ	H	I	I	J	K	L	M	N	Ń	O	Ó	P	Q	R	S	T	U	Ú	V	Y	Ý	Z	Sh	Ch
Cyrillic Kazakh	А	Ә	Б	Д	Е	Ф	Г	Ғ	Һ	І	И	Ж	К	Л	М	Н	Ң	О	Ө	П	Қ	Р	С	Т	Ұ	Ү	В	Ы	У	З	Ш	Ч

Shared Similar Unique

FIGURE 1 | Comparison of a new Latin-based Kazakh alphabet to English and Cyrillic-based Kazakh.

the meaning of the spoken word and develop their cognitive abilities. The study also showed that the use of English phonemes and converting them into a Bengali (Sylheti) script caused rapid learning. A key point in this research is that transliteration serves as a practical tool for teachers to increase students' attention span by expanding their linguistic capacities and to stimulate them to develop bilingual skills in more than one script.

Previous studies touch upon digraphia and biscriptal practices that generate social influences, however, they generally do not take into consideration an educational approach toward addressing the issues of the new script's introduction into the educational domain. So far, some methods were proposed to introduce the new alphabet to students. For instance, Gonzalez et al. (2011) experimented with two methods of tracing or copying to learn handwritten character patterns using a tablet with a stylus. It was found that two methods had differing advantages relying on short or long-term learning measures: short-term retention was better when tracing, while long-term performances had no significant difference when both methods were used. Consistent with these two methods, our study also attempts to investigate the impact of these on learning the newly-introduced Latin-based Kazakh alphabet.

2.2. Robot-Assisted Learning

Recent research efforts within the HRI field have shown that social robots are increasingly deployed in robot-assisted learning and education (Neumann, 2020). Robots are generally welcomed by students who view them as learning partners or companions in an optimistic way (Kennedy et al., 2016; Charisi et al., 2020). Rosenberg-Kima et al. (2019) found that the physical presence of robots brought positive changes for university students because of the technical functionality, social, and psychological activity. Namely, students pointed out the benefits as follows: "accessible to multiple people," "immediate feedback," "he is not judgmental like human beings," "pleasant and motivating." Some research has targeted specific skills required for language learning: reading (Gordon and Breazeal, 2015; Michaelis and Mutlu, 2018; Yadollahi et al., 2018), grammar (Belpaeme et al., 2018b), or vocabulary learning (Balkibekov et al., 2016). Other research demonstrated that learners cultivate favorable impressions toward robots as learning companions and the child-robot interaction may lead to increased self-confidence and better task performance requiring creativity (Dennis et al., 2016; Alves-Oliveira et al., 2017) and problem-solving (Liu and Chang, 2008). Other studies (Kanda et al., 2007; Sharkey, 2016) explored long-term learning between robots and children to better understand this type of HRI in a real-world environment.

Since 2014, the CoWriter project has investigated how robots can provide a learning environment for children in order to improve handwriting skills based on the LbT paradigm (Hood et al., 2015; Jacq et al., 2016; Lemaignan et al., 2016). This autonomous approach allows children to act as a teacher, or a tutor, who is responsible for the robot's learning. Therefore, the children, committed to the learning success of a robot, become a central actor in handwriting practices along with a social robot. In the field of pedagogy, researchers dubbed this type of process as the Protége effect in reference to Seneca's famous saying "while we teach, we learn." In this regard, previous studies have addressed the potential benefits of LbT for learner's motivation (Jacq et al., 2016), task commitment, increased self-esteem, and mental activity (Jamet et al., 2018). In addition, Lubold et al. (2018) suggested a set of design propositions to adjust dialog strategies, revealing that individual characteristics affect the LbT outcome. Motivated by this paradigm, the CoWriting Kazakh project aims to increase children's self-confidence and motivation to learn the Latin-based Kazakh alphabet and its orthography. In view of the recent language reform in Kazakhstan, this paper investigates whether the CoWriting Kazakh project addresses challenges of teaching and motivating young learners to learn a new Latin-based Kazakh alphabet. Such findings are particularly timely as they can inform future research and practice to promote remote learning, such as required as a result of the recent COVID-19 pandemic.

2.3. Prior Work on CoWriting Kazakh

The CoWriting Kazakh system was previously deployed in two separate HRI studies within the novel context of learning the new Latin-based Kazakh script: an exploratory study with 48 children (Kim et al., 2019) and a follow-up study with 67 children (Sandygulova et al., 2020). Participants were asked to teach a humanoid NAO robot how to write Kazakh words using one of the scripts, Latin or Cyrillic. We hypothesized that a scenario in which the child is asked to mentally convert the word to Latin would be more effective than having the robot perform conversion itself. Two conditions were implemented that differed in who performed the conversion: Latin-to-Latin (L2L) and Cyrillic-to-Latin (C2L) conditions. In L2L, the child heard the word to be written and had to write it directly in a new Latin script. Then the robot wrote the word in Latin as corrective feedback. From this demonstration, the child is given an opportunity to see the error-free spelling in the Latin script, and importantly to learn from the robot's correct spelling via the error analysis (Jobeen et al., 2015). In C2L, the child heard the word and wrote it in a familiar Cyrillic script. Then the robot performed the script conversion by writing the same word using the Latin-based Kazakh alphabet. Results demonstrated a gender bias with the L2L strategy being more effective for girls. In contrast, boys learned significantly more when they spelled the words using Cyrillic and only observed the robot's correct spelling of the Latin-based Kazakh words. The study presented in this paper employs the L2L version of the system in order to compare what learning aid would result in greater learning gains of a new script.

2.4. Human Teacher vs. Robot Interaction

The shortage of teachers has become a topic for discussion across many contexts (Edwards and Cheok, 2018; Garcia and Weiss, 2019) and continues in a time where innovative technology becomes more of an imperative. Therefore, the demand for school teachers has increased exponentially and it has resulted in a necessity to recruit almost 69 million teachers to provide quality education (SDG 4) by 2030 (United Nations, 2015). This problem has led to the development of AI in education (AIEd) tools and Intelligent Tutoring Systems (ITS), which are likely to scaffold teachers in flexible and personalized ways (Luckin et al., 2016). These transformations include social robots that may be embedded into a classroom to serve the role of teacher's assistants (e.g., PaPeRo Tung, 2016, iROBI Han and Kim, 2009) by helping students to stay engaged and motivated. With ever-increasing technological advancements in education, future human teachers should focus on developing students' critical and productive thinking skills and robot assistants can minimize a teacher's workload by scaffolding the learning environment in a digitized way (Newton and Newton, 2019). This characteristic of robots is considered an asset for human teachers who may focus more on content delivery and creative instruction.

To date, robots and teachers are rarely investigated to compare their effectiveness in the classroom. Sharkey (2016) stressed that robots can act in tandem with and supplement a human teacher, but it seemed unimaginable that fully-fledged robots can be in charge of the whole learning process by themselves. Evaluating a teacher condition with and without a robot presence, Alemi et al. (2014) found children in the teacher-robot condition learned significantly more than only with a human teacher. What is worth noting here is that children can learn similarly well when the instruction is delivered either by a robot or by a teacher (van den Berghe et al., 2019). Central to the LbT approach (Lemaignan et al., 2016) employed in the present study, children tend to take on the responsibility to commit themselves for robot learning. Therefore, children's task commitment may increase in a robot condition similar to how teachers invest their time and knowledge in children's learning (Chase et al., 2009). Thus, this phenomenon is clearly important to consider as an effective approach to increase children's learning curve which needs to be supported by convincing studies in the HRI field. As of today, it seems obvious that robots can not replace teachers in classroom settings but rather act as a helpful assistant to human teachers to effectively deliver instruction. We suggest that the complementary nature of robot-assisted teaching can change ensuing dynamic technological solutions in educational settings.

2.5. Robots vs. Other Learning Aids

In comparison to current traditional technologies, using robots in language classrooms is stimulating, relying on their (non)verbal and social characteristics (Meghdari et al., 2013; Neumann, 2020). Unlike other computing technologies, such as tablets and laptops, the use of social robots may yield significant benefits for learning in three ways (Belpaeme et al., 2018a). First, as most learning and teaching processes happen in the classroom, robots seem a feasible option to fit the physical world and thus facilitate classroom engagement. It is highlighted that the physical embodiment of robots has a huge impact on people seeing them as more human-like, sociable, and more creative than a tablet (Li, 2015; van den Berghe et al., 2019). To illustrate, students exposed to the robot condition perceived it to be more comfortable for learning compared to the tablet condition (Rosenberg-Kima et al., 2019). Second, the presence of robots enables more social behaviors from people whose learning is not a mere task-based type of learning. For instance, Westlund et al. (2015) compared the effectiveness of three learning scenarios (human, robot, and tablet) with regard to children's rapid word learning. It was revealed that young learners strongly preferred robots despite similar word learning outcomes in three learning scenarios. No significant differences in vocabulary learning were found when robots were put on par with computers (Hyun et al., 2008). Finally, learners are more motivated and interested to learn due to the interactive communication with robots, leading to further result in an increased task commitment. Li (2015) came to the conclusion that the physical presence of a robot improves a learner's task performance compared to other learning aids.

3. HRI SYSTEM

This section details the CoWriting Kazakh system and its scenario.

3.1. Software and Hardware Components

The hardware components of the system include the Wacom Cintiq Pro tablet and a humanoid robot NAO. The tablet is used as the second monitor when connected to a laptop. It is coupled with a stylus with an 8.192° of pressure sensitivity and tilting recognition. This allows us to acquire the trajectory of handwriting and the pressure and tilt at each point (Sandygulova et al., 2020). The humanoid robot NAO is an autonomous and programmable robot manufactured by SoftBank Robotics. It is the mostly used humanoid robot in HRI research for robot-assisted educational and healthcare applications. The height of the robot is 58 cm which makes it easy to transport, and its human-like appearance also attracts children. It also has 25 degrees of freedom and seven tactile sensors. In fact, CoWriting Kazakh is an extended version of the CoWriter system[1]. In comparison with the original CoWriter's LbT paradigm in which

[1]https://github.com/chili-epfl/cowriter

the robot's handwriting gradually improves throughout several demonstrations by the child, the CoWriting Kazakh does not include a handwriting improvement component. In the proposed system, the child-robot cooperative learning is 2-fold: (1) the robot learns new Kazakh words from the child; (2) the child learns the Latin-based Kazakh script from the robot. Their interaction takes the form of turn-taking in writing words in Kazakh (see **Figure 2**; Sandygulova et al., 2020).

Regarding the software aspect of the CoWriting Kazakh system, it is designed to recognize learner's handwriting in Cyrillic and transliterate it into Latin script. We developed the handwriting recognition based upon our collection of the Cyrillic-MNIST dataset (Sandygulova et al., 2020).

The interaction was implemented using NAO's English text-to-speech and face recognition engines. Across the communicative interaction, the robot demonstrated a set of animations along with hand gestures and head movements. Moreover, the robot was able to generate non-verbal social behaviors, such as recognizing the child's physical presence with eye contact. When the robot "writes," it looks down at the tablet and moves its right hand mirroring the letters' trajectory in the air as they appear on the tablet next to child's writing. Children usually watch this "writing" motion closely, being attracted and interested by the fact that the robot can write without holding a pen or touching the surface of the tablet. The demonstration is available at the link: tiny.cc/iektpz.

3.2. Scenario

The scenario involves a robot taking the role of a peer. As it is introduced to a child, the peer robot is put into the position of a native English speaker who wants to learn Kazakh. Since the only alphabet known to the robot is Latin, the child is asked to show it how to write Kazakh words in the Latin-based Kazakh alphabet. The child is motivated to try their best to listen to the robot carefully in order to understand the robot's speech. It was crucial to create basic robot spoken utterances for the children's English-level appropriateness that was verified by children's English teachers.

On average, the child-robot interaction lasted 20-30 min according to how much time children take to write. The robot's list of speech utterances are as follows:

NAO: -Hello. I am a robot. My name is Mimi. [Waves his hand]
Child: -...
NAO: -I study Kazakh language. Can you help me?
Child: -...
NAO: -How do you say "Hello" in Kazakh?
Child: -Sálem
NAO: -How do you write it? Please write it using Latin letters so that I can read it.
Child: -[Writes on a tablet using Latin-based Kazakh]
NAO: -Let me try to write it too [gesticulates]. This is a correct writing using Latin letters.

... repeated for another 12 words
NAO: -You are a great teacher. Thank you very much! Goodbye! [waves].

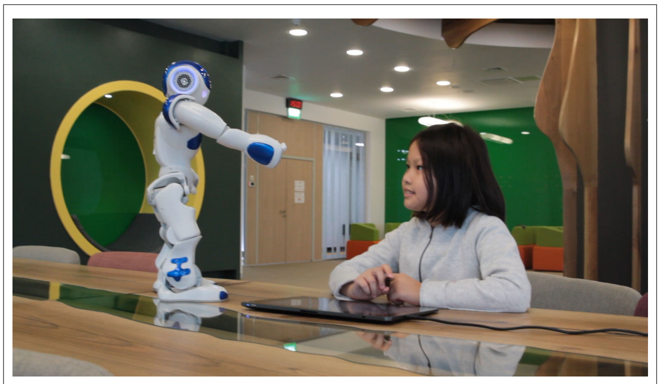


FIGURE 2 | Experimental setup.

4. EXPERIMENT

The methodology of the present study was developed and then aligned with the previous work (Kim et al., 2019; Sandygulova et al., 2020).

4.1. Method

The experiment was carried out at a primary school in Kazakhstan's capital city. It included a one-to-one interaction for each child participant. The participants were introduced to a condition in a between-subject design, with a learning aid type as the between-subject variable.

Each child interacted with a randomly selected learning aid condition for ∼20–30 min. A third of the children interacted with the robot + tablet in a *Robot* condition, another third of the children interacted with a version of the CoWriting Kazakh using only a tablet in a *Tablet* condition, while the other third of the children interacted with a teacher in a *Teacher* condition using pen and paper for demonstrations. Counterbalancing was also applied in terms of gender and year group so that each condition had a balanced number of boys and girls. Assignment to each of the conditions was otherwise random for any particular child. It should be noted that when the whole experimental procedure was over (i.e., after the post-test), children from the two non-robot conditions were offered the opportunity to interact with a robot. The majority of children expressed their desire to interact with a robot. Thus, their post-test score and interview results were not affected by this interaction.

4.2. Recruitment

The present research project was granted approval by the Nazarbayev University Institutional Research Ethics Committee. To conduct the experiment, informed consent forms were obtained from all participants and their parents. It is supplemented by including an assent form for children and an informed consent form for their parents or legal guardians. Children were provided with an overview of the study's purpose and the data collection process. With the presence of their teachers, assent and informed consent forms were distributed to

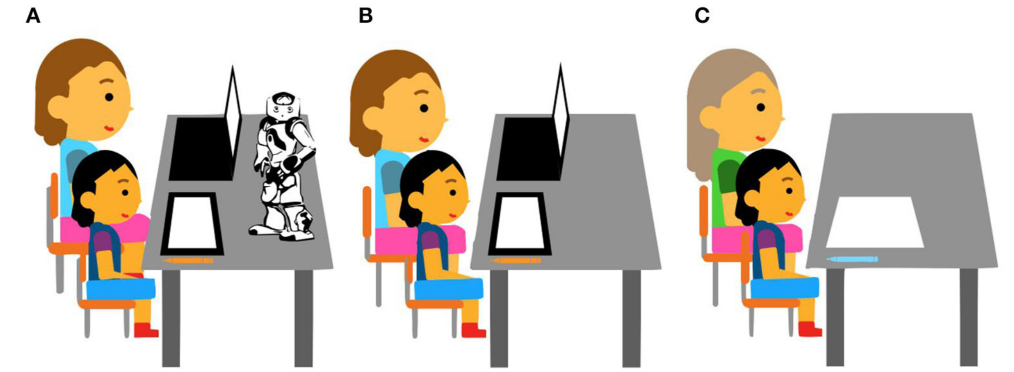


FIGURE 3 | Schematic overview of the setup by conditions: **(A)** robot + tablet, **(B)** tablet only, **(C)** teacher + pen and paper.

children in a classroom. Afterwards, they were asked to show the documents to their parents, and with their permission to submit them to their teachers who collected all the documents for us.

4.3. Participants

In total, the study recruited an equal number of 62 male and female children aged 7–9 years old. Children were assigned randomly to either a robot condition ($N = 21$), tablet condition ($N = 21$), or teacher condition ($N = 20$). The children represented different socio-economic backgrounds and all of them were native or fluent Kazakh language speakers. According to their writing experiences, second-graders had spent about 16 months writing in Cyrillic, and third-graders had spent about 28 months writing in Cyrillic before the experiment. The children learned handwriting for 6 h on a weekly basis, ranging from simple shapes to the Cyrillic alphabet after nearly 6 weeks in the first grade. In addition, they had 2 h of weekly English lessons in which they also practiced the English alphabet starting from grade one. In other words, they had spent 16 months of handwriting in English. However, the children had not been taught to write in a Latin-based Kazakh alphabet (revised version of the English alphabet with 6 distinctive letters) and its associated writing system. Therefore, compared to the Cyrillic script, all children had no learning experience in the Latin-based Kazakh alphabet.

4.4. Hypotheses and Conditions

Based on the CoWriting Kazakh system explained above, we examined whether it is more effective for a child to perform the mental conversion and see correctly written Latin words given by the robot. To that end, the main hypotheses are formulated as follows:

- H1: The CoWriting Kazakh will provide an effective learning scenario that will significantly improve the amount of learned letters, which, in turn, suggests that the proposed intervention contributes to learning a new script.
- H2: Girls will outperform boys in letter learning, as in our previous work we observed such gender effect in a Latin-to-Latin condition when children performed mental script conversion (Sandygulova et al., 2020).
- H3: Children will learn more letters when learning from a robot and a tablet than from a teacher and tablet only.
- H4: Children will enjoy the robot condition more in comparison to the tablet and teacher conditions, as it was reported in Li (2015), Westlund et al. (2015), and Rosenberg-Kima et al. (2019) that robots are of great advantage due to their physical presence and human-like appearance.

To test these hypotheses, three conditions are distinguished with respect to the type of learning aid:

- Robot condition: the child hears the word to be written pronounced by the robot in English and has to translate it to Kazakh and write it directly in Latin on the Wacom tablet using its stylus. Then, the robot simulates the writing while the letters are written on the tablet in Latin as corrective feedback. The video demonstration is available at the link: tiny.cc/iektpz. **Figure 3A** presents a schematic overview of the Robot condition where a researcher controls the system launch on their computer.
- Tablet condition: the child is presented with a pop-up window on the tablet with instructions to first translate and then write the words in Latin-based Kazakh. The vocabulary is the same and 13 words are in the same order as in the Robot condition. When its time for corrective feedback, the correct spelling of the words appear in the same way on the tablet as in the Robot condition. **Figure 3B** presents a schematic overview of the Tablet condition where a researcher controls the launch of the system on their computer.
- Teacher condition: the teacher speaks Kazakh language and asks children to write the words in Latin-based Kazakh. The vocabulary is the same and 13 words are in the same order as in the other conditions. When it is time for corrective feedback, the teacher then shows a correctly written spelling in Latin-based Kazakh. They use a pen and paper. **Figure 3C** presents a schematic overview of the Teacher condition.

In all three conditions, children had to mentally perform the script conversion without help. We did not assist them in their writing process unless they did not comprehend or recognize the robot's speech. **Figure 3** shows the setup of each condition.

4.5. Procedure

The procedure of the experiment included a survey, a pre-test, a learning activity, an interview, and a post-test. The whole process for each child took about 30–40 min.

Each child was invited from a class and accompanied by the first researcher to a place where the experiment was conducted. Before reaching the place, the first researcher began with an icebreaker to put the child at ease: "My name is Zhanel. And what is your name?" "When I was your age, I was fond of Mathematics, and what about your most liked subject?" When they entered the room, children were given a seat at the table with surveys and responded to a couple of demographic questions (i.e., age, gender) and what their mood was. Afterwards, children were sat alongside the second researcher to take a pre-test that evaluated children's existing knowledge of the Latin-based Kazakh alphabet. As the surveys and pre-tests were completed, children changed their seats and sat at the table with their learning condition (robot, tablet only, or teacher). Following the interaction, children participated in a structured interview with the first researcher who asked how they perceived their corresponding learning aid. At the end of the experiment, children were distributed a post-test analogous to the pre-test to obtain the measure of their knowledge of Latin-based Kazakh script. Similarly, the stage-by-stage procedure was followed when the first researcher accompanied the child back to the class and invited the next child.

4.5.1. Survey

A mini-questionnaire was conducted by the first researcher who documented the child's demographic profile and before-the-experiment mood using a 5-point Likert scale.

4.5.2. Pre-test

The pre-test was the next stage, where each child was introduced to a table of 23 Cyrillic-based Kazakh alphabet letters to complete the task by converting each letter in Cyrillic to an equivalent in the Latin-based Kazakh alphabet. This allowed us to identify the child's knowledge of Latin script before the experiment.

4.5.3. Learning Activity

When the child completed the pre-test, the researcher asked the child to sit in front of the robot, tablet, or teacher. The activity would come to an end either by the child or after all 13 words were trained. As mentioned before, the words were selected in accordance with the children's level of English, which were previously approved by their English instructor. It should be noted that all 33 Latin-based letters were present in the chosen 13 words with a minimum of one letter occurrence.

4.5.4. Interview

As the interaction with a robot was completed, the child took a seat along with the first researcher who then carried out a structured interview which involved the following questions from our previous studies (Kim et al., 2019; Sandygulova et al., 2020):

1. How is your mood? (5-point Likert scale)
2. Funometer scale (Markopoulos et al., 2008) was described to a child by providing an example of how it operated: the winter has the coldest weather (at the lowest level of the meter) while the summer is the sunniest season (at the highest level of the meter). How would you rate today's weather? Afterwards, the following example showed an enjoyable measurement: "imagine that you are having a birthday party and you receive many gifts, you enjoy your time very much (rate your mood at the top of the meter), or in reverse when you feel bored with waiting for a bus (rate your mood at the bottom of the meter). Similarly, how would you rate your learning activity?" (The rating was scaled from 0 to 100).
3. Sorting task: The researcher illustrated this task to a child by demonstrating five items they considered the most and least interesting. In an activity, five small paper items were presented: a book, a tablet, a NAO robot, a computer, and a teacher. (The sorted position of the child's learning aid was recorded using a 5-point Likert scale).
4. Likewise, the researcher asked the children to sort the five items with regard to what/who is the least/most effective for learning? (The sorted position of the child's learning aid was recorded using a 5-point Likert scale).
5. Children also performed a sorting task with the five items (a book, a tablet, a robot, a computer, and a teacher) responding to the question what/whom they preferred the least/most?
6. In closing, children sorted the five items based on what/who is the easiest way to learn with/from?

These questions helped to reveal how children feel about the interactions. We applied different techniques to explain the procedures explicitly and ask easy to follow questions. For instance, the Funometer scale and a paper version of the learning aids were printed for children to manually move the paper and situate it on a scale. This was an appropriate option compared to pictorial five-level Likert items by providing more detailed responses. Most children placed their ratings on a Funometer scale near 70–90 out of 100. In addition, the children were asked to rate their mood after the interaction, to compare whether their mood changed or not. Finally, we performed a group of sorting tasks in which children were asked to sort first their corresponding learning aid (e.g., teacher) and then to sort a robot as well.

4.5.5. Post-test

The post-test was the final stage in the experiment. At this stage, children were introduced to the same table of 23 Cyrillic letters as distributed in the pre-test. Similarly, children were asked to write Latin-based Kazakh letters. This stage was important to evaluate the children's learning gains by comparing the number of learned letters in pre- and post-tests. Children were given a book for participation after the completion of the post-test. Children that did not get to interact with the robot were offered the opportunity to repeat the learning activity but with the robot this time. Their performance in the tests and responses to interview questions were not affected by this activity with the robot.

TABLE 1 | Pre- and post-test descriptives.

Gender	Condition	Age	*N*	Pre-test	Post-test	Learned letters
Boys	Robot	8.30	8	$M = 11.63$, $SD = 4.75$	$M = 15.50$, $SD = 5.24$	$M = 3.88$, $SD = 2.48$
	Tablet	8.40	10	$M = 12.00$, $SD = 5.06$	$M = 15.20$, $SD = 4.16$	$M = 3.20$, $SD = 3.12$
	Teacher	8.36	11	$M = 10.55$, $SD = 4.41$	$M = 14.82$, $SD = 5.17$	$M = 5.18$, $SD = 2.27$
	Overall	8.36	29	$M = 11.34$, $SD = 4.61$	$M = 15.14$, $SD = 4.69$	$M = 4.14$, $SD = 2.69$
Girls	Robot	8.64	11	$M = 11.64$, $SD = 5.59$	$M = 15.00$, $SD = 4.92$	$M = 3.36$, $SD = 1.43$
	Tablet	8.36	10	$M = 12.10$, $SD = 4.46$	$M = 14.70$, $SD = 3.59$	$M = 2.60$, $SD = 2.37$
	Teacher	8.67	8	$M = 10.63$, $SD = 4.93$	$M = 14.38$, $SD = 5.55$	$M = 3.75$, $SD = 1.91$
	Overall	8.55	29	$M = 11.51$, $SD = 4.90$	$M = 14.72$, $SD = 4.53$	$M = 3.21$, $SD = 1.92$

5. RESULTS

A series of Kolmogorov-Smirnov and Shapiro-Wilk tests were conducted on all dependent variables overall and within groups (i.e., gender and condition) to check the assumption of normality. Since some scores were significantly non-normal, non-parametric tests were used for the statistical data analysis presented in some of the following sections.

5.1. Learned Letters

Four children did not complete their post-tests, thus this analysis was conducted on data from 58 children (see **Table 1** for demographics of participants for every condition). The number of learned letters was calculated to identify the difference between letters known in the post-test and the pre-test (e.g., if 18 correct letters were marked in the post-test and 10 correct letters were marked in the pre-test, the number of learned letters is 8). As a result of the learning activity, children improved their knowledge of the Latin-based Kazakh alphabet. The average number of learned letters was 3.67 (SD = 2.37, Max = 9, Min = 0).

To test H1, we conducted a paired samples t-test on pre- and post-tests which revealed that children had a statistically significant improvement in their Latin alphabet knowledge from 11.48 ± 4.64 to 14.68 ± 4.62: $t_{(57)} = -10.5, p < 0.0005$. **Table 1** presents pre- and post-test descriptives.

A two-way ANOVA was conducted examining the effect of gender and condition on a number of learned letters. We did not find a statistically significant interaction between the effects of gender and condition, $F_{(2, 52)} = 0.225, p = 0.799$. Boys and girls learned the most letters in the Teacher condition: boys learned 5.18 ± 2.27 while girls learned 3.75 ± 1.91. The robot condition was the second most effective learning aid where boys learned 3.88 ± 2.48 and girls learned 3.36 ± 1.43 letters. The tablet condition was the least effective for both gender groups (3.2 ± 3.11 vs. 2.6 ± 2.36), though not significant. These results are presented in **Figure 4**.

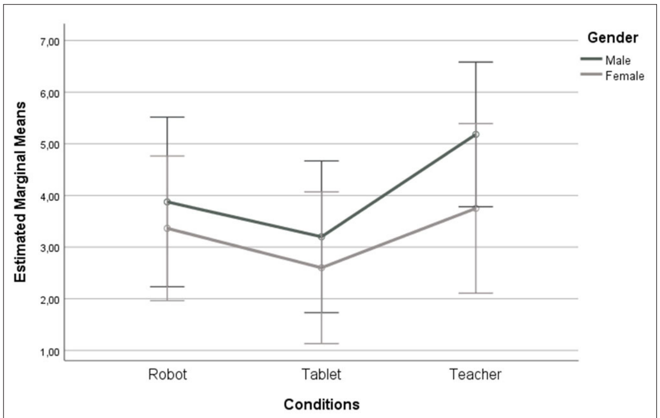


FIGURE 4 | Average number of learned letters for boys and girls by the conditions. Error bars show 95% Confidence Interval.

To test H2, a Welch's ANOVA was conducted to examine whether there is a significant gender difference in the number of learned letters: $F_{(1, 50.54)} = 2.299, p = 0.136$. Boys learned 4.14 ± 2.69 while girls learned 3.21 ± 1.92 letters. Girls scored slightly better in a pre-test (11.51 ± 4.90 vs. 11.34 ± 4.61), but in a post-test boys outperformed girls (15.14 ± 4.7 vs. 14.72 ± 4.53), though not significantly. Then, a series of separate one-way ANOVAs was conducted to find gender differences for each condition: Teacher, Robot, and Tablet. However, the differences in learning gains were not significant between boys and girls when analyzed separately either. This finding rejects our H2, suggesting that boys and girls learned more-or-less equally, which contradicts our previous finding that the Latin-to-Latin approach was more effective for girls who learned more, as it was previously found in Sandygulova et al. (2020).

Finally, to test H3, we examined whether there is a significant difference in the number of learned letters between the three conditions. The assumption of normality was met by all three groups. Levene's test revealed that population variances of learned letters for the three types of conditions are equal, $F_{(2, 55)} = 1.3, p = 0.28$. As all the assumptions were met, we proceeded with a one-way ANOVA which revealed that there is no statistically significant difference in the number of learned letters between conditions: $F_{(2, 55)} = 2.618, p = 0.082$. Children learned slightly more letters in the Teacher condition (4.58 ± 2.19), followed by the Robot (3.58 ± 1.89), and Tablet conditions (2.9 ± 2.71), though without significance. This finding rejects our H3, suggesting that Robot, Tablet, and Teacher conditions did not lead to significantly different learning gains.

In addition, in order to find out if the three conditions were equally as effective as a learning aid, we performed an equivalence analysis TOST (two one-sided tests) test (Rusticus and Lovato, 2011; Lakens et al., 2018) setting equivalence bounds Δ_L and Δ_U to SESOI which is equal to $\pm\ d_{critial}$. The critical effect size was calculated using the following formula, $d_{crital} = t_{critical}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$, that was proposed by Lakens et al. (2018). As a result, it did not show a significant equivalence: Robot vs. Tablet: $d = 0.30$, 95% CI for Cohen's d: $[-0.34, 0.94]$, $\Delta_L = -0.33$, $\Delta_U = 0.33$, $t_{(37)} = 0.35, p = 0.636$, 90% CI for mean difference [−0.26, 1.62]; Tablet vs. Teacher: $d = 0.713$, 95% CI for Cohen's d: $[-0.71, 2.35]$, $\Delta_L = -0.418$, $\Delta_U = 0.418$, $t_{(37)} = -1.599, p = 0.94$, 90% CI for mean difference $[-1.36, -0.06]$; Robot vs. Teacher: $d = -0.535$, 95% CI for Cohen's d: $[-1.18, 0.11]$, $\Delta_L = -0.423, \Delta_U = 0.423$, $t_{(36)} = -0.867, p = 0.804$, 90% CI for mean difference $[-1.79, -0.21]$. These findings suggest that the three conditions were neither significantly different nor significantly equivalent in their facilitation of learning gains. This result is due to our sample size being quite small, leading to not having sufficient power to reject either null hypothesis.

5.2. Mood Change

The mood change variable was calculated as the difference between reported pre- and post-interaction ratings of children's mood on a 5-point Likert scale.

A series of Mann-Whitney U tests was conducted that revealed that there is a statistically significant difference in Mood Change score between the Robot (0.45 ± 0.68) and Teacher (-0.05 ± 0.52) and Tablet (-0.05 ± 0.83) conditions: $U = 122, W = 312, Z = -2.35, p = 0.019$. This finding supports our H4, in that the Robot condition was more enjoyed in comparison to the other two conditions.

A Mann-Whitney U test was conducted to check gender differences in children's Mood Change values, however it was not significant: $U = 370.5, W = 805.5, Z = -1.184, p = 0.236$.

Apart from the numerical value of the Mood Change variable, we also categorized it as either Increased, Decreased, or Unchanged. A series of chi-square tests of independence was conducted to examine the effect of categorical variables (gender or condition) on children's Mood Change. We did not find any statistically significant results between boys and girls for these measurements.

There were no significant differences between conditions in how children responded to Mood Change: χ^2 (4, $N = 59$) $= 5.932, p = 0.204$. **Figure 5** presents that although there is a similar number of children who did not have their mood changed in all conditions, children in the Robot condition were more likely (7) to have their mood increased in comparison to Tablet (4) and Teacher (2) conditions. And none of the children in Robot condition had their mood decreased in contrast to three children in Tablet and Teacher conditions each.

5.3. Funometer

Children were asked to rate how much they enjoyed their corresponding learning activity ranging from 0 to 100 on a Funometer scale (Markopoulos et al., 2008). An average rating for all children was 78.22 ± 17.61 ($Mdn = 75, Max = 100, Min = 40$).

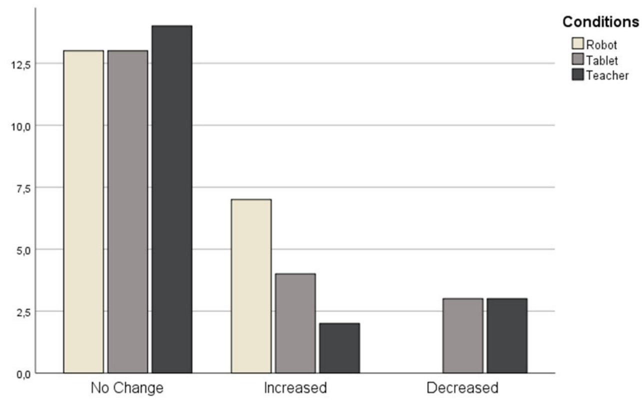


FIGURE 5 | Number of children from three conditions grouped by their mood change.

A Kruskal-Wallis test revealed that there is no significant difference in children's ratings between three conditions: $\chi^2_{(2)} = 0.849, p = 0.654$. Children in the Robot condition rated their experience as slightly higher ($M = 80.75, SD = 17.71, Mdn = 77.5, Max = 100, Min = 50$) than those children in the Tablet ($M = 77.5, SD = 18.17, Mdn = 75, Max = 100, Min = 40$) and Teacher conditions ($M = 76.32, SD = 17.55, Mdn = 75, Max = 100, Min = 45$), though not significantly.

We conducted a Mann-Whitney U Test to compare children's ratings between gender groups showing that boys rated their interaction as slightly better than girls did, even though not significantly: $U = 319.5, W = 754.5, Z = -1.784, p = 0.074$. Boys' rating was 82.17 ± 15.69 ($Mdn = 80, Max = 100, Min = 50$) while girls rated their experience as 74.14 ± 18.81 ($Mdn = 75, Max = 100, Min = 40$).

5.4. Sorting of Learning Aids

When asked to position children's corresponding learning aid according to its effectiveness to teach, easiness to learn from, being interesting and enjoyable in comparison with a robot, book, tablet, computer, and a teacher, children's ratings were recorded and analyzed against each other.

5.4.1. Effectiveness Rate

A Kruskal-Wallis test revealed that there was a significant difference in this rating between conditions: $\chi^2_{(2)} = 7.36, p = 0.025$. A series of Mann-Whitney U tests was conducted to check between Teacher and Robot ratings: $U = 88, W = 298, Z = -2.761, p = 0.006$. The Teacher was rated as 3.9 ± 1.4 which was significantly higher than the Robot's rating (2.9 ± 1.1). Tablet was rated as 3.05 ± 1.47 which did not have significant difference with the Robot rating, but was significantly different with the Teacher rating: $U = 88, W = 298, Z = -2.761, p = 0.006$. No gender differences were found for this rating.

We also noted the sorting position of the robot for all children. We found that children who interacted with the tablet rated the effectiveness of the robot as 2.76 ± 1.00, while children who were taught by the teacher rated the robot as 2.79 ± 0.86. Participants in a robot condition rated it slightly higher at 2.9 ± 1.1. A series of Mann-Whitney U tests did not find a statistically significant difference in this rating, neither between conditions nor between gender groups.

5.4.2. Easiness Rate

No statistically significant differences were found for this rating between different learning aids. Girls rated the learning activity significantly easier (3.82 ± 1.28) than boys did (3 ± 1.39) according to a Mann-Whitney U test: $U = 285.5, W = 781.5, Z = -2.313, p = 0.021$.

Participants from the tablet condition rated the robot as slightly more difficult (2.81 ± 1.07) than those participants that interacted with a teacher (2.79 ± 1.36) and robot conditions (3.3 ± 1.42), though not significantly. Gender groups did not rate the robot as significantly different for this rating.

5.4.3. Likeability Rate

A Kruskal-Wallis test was conducted to see which learning aid was rated the most likable and showed that there was a statistically significant difference in this rating: $\chi^2_{(2)} = 12, p = 0.002$. A series of Mann-Whitney tests showed that the robot was rated as statistically significantly higher (4.35 ± 1.04) than both the tablet (3.57 ± 1.29): $U = 127.5, W = 358.5, Z = -2.284, p = 0.022$, and the teacher ($3.06 \pm 1.21$): $U = 70.5$, $W = 241.5, Z = -3.329, p = 0.001$. No gender differences were found for this rating.

A series of Mann-Whitney U tests showed that those who interacted with the robot liked it significantly more and rated the robot as 4.35 ± 1.04 than those who interacted with the tablet only (3.57 ± 1.36): $U = 137.5, W = 368.5, Z = -2.013, p = 0.044$. Children from the teacher condition rated the robot as 4.11 ± 1.29, though it was not significant.

5.4.4. Interest Rate

A series of Mann-Whitney U tests revealed that children rated the Robot as significantly more interesting than the Teacher: $U = 114.5, W = 285.5, Z = -2, p = 0.045$. The robot was rated as 3.9 ± 1.12, while the teacher was rated as 3.17 ± 0.99. No statistically significant differences were found between other learning aids as well as between gender groups.

A series of Mann-Whitney U tests revealed that children in the robot condition rated it as significantly more interesting (3.9 ± 1.12) than those in the tablet condition (3.09 ± 1.22): $U = 135.5, W = 366.5, Z = -2.042, p = 0.041$. No statistically significant differences were found between participants' ratings in Teacher and Robot conditions, as well as between gender groups.

6. DISCUSSION AND LIMITATIONS

Since all participants attended the same school, we cannot generalize our results or confidently state that the findings will be workable for other Kazakhstani schools.

However, as found in the analysis of the results, we can claim that H1 is valid, supporting that the intervention with the system was effective on the children's performance in both pre- and post-tests at a high statistically significant level ($p < 0.001$). We found similar findings in our previous studies which allow us to declare the effectiveness of the proposed learning approach of teaching in a single session. The children were able to learn from the approach when they first attempt to convert the words to Latin themselves and then observe the corrective feedback.

6.1. Gender Differences

Given the non-significant differences between gender groups in the presented study, we can interpret that boys and girls learned more-or-less similarly in all conditions. It contradicts our H2 and previous study's results (Sandygulova et al., 2020) in which we found a gender imbalance in the performance of boys and girls with respect to the learning gain results. Girls performed better in the Latin-to-Latin condition and learned significantly more letters. As distinct from it, boys learned more letters when following the Cyrillic-to-Latin condition. Since this study only offered the Latin-to-Latin condition, this mismatch is an unexpected turn but might be due to the different set of words that was selected for this study. This time, most of the words that the child had to show to the robot had a maximum of four letters in contrast to Sandygulova et al. (2020)'s selected words. This should be carefully accounted for in our future studies.

6.2. Robot vs. Human Teacher

The study revealed neither a statistically significant difference nor statistically significant equivalence in the number of learned letters when taught by a robot or by a teacher. This result is due to our sample size being quite small leading to an insufficient power to reject either null hypothesis. This resonates with the previous works that found no significant differences in the number of learned words (Westlund et al., 2015), and test-scores in mathematics with either a robot or human teacher (Mubin et al., 2019). In the meantime, significant benefits of peer robots over traditional teacher-to-student interactions and advantages of robot-assisted classes in contrast to only a teacher-led classes have been discussed so far (Alemi et al., 2014; Belpaeme et al., 2018a). In a similar vein, Rosenberg-Kima et al. (2019) also indicated that robots successfully assisted the learning experience of students, and in some cases even more effective interactions were reported in comparison with human teachers. Importantly, they also stressed the idea of Human-Robot-Collaboration (HRC) that provides a space for a human teacher and a social robot to work in tandem. Robots do have essential skills to act in the capacities of tutors and teacher's assistants, bearing in mind that human teachers cannot be fully replaced in a classroom. Considering that the comparison of the effectiveness of a human and a robot intervention is rarely explored, this study needs further refinement in a larger sample size and with longer interactions.

6.3. Robot + Tablet vs. Tablet Only

Similarly, the results from these two conditions fail to reject the standard null hypothesis, while failing to reject the equivalence null hypothesis, which leads us to conclude both "not different" and "not equivalent." The non-significant difference between the two conditions can be explained by the fact that letter learning is a simple task and any exposure to this task leads to learning gains. In addition, since our learning scenario does not rely on the main advantage of social robots over the tablet, i.e., their ability to provide verbal and non-verbal cues, this might have caused the tablet only version to provide more-or-less similar alphabet learning gains. Thus, it can be noted that touch-screen tablets are a relevant option for learning a new script in line with robots. These results are reminiscent of the large-scale study (Vogt et al., 2019) which indicates that the success of learning L2 words cannot be accomplished merely with the robot condition. As a result of these findings, we can deduce that robots combined with and assisted by tablets are considered preferable rather than just the robot or tablet. For instance, instead of using them separately, Park and Howard (2013) proposed the HRI toolkit that enables the use of tablets as mediators between humans and robots. By comparing them, however, an increasing number of studies (Li, 2015; Westlund et al., 2015; Rosenberg-Kima et al., 2019) have reported that robots are of a great advantage due to their physical presence and human-like appearance compared to portable tablets. These socially-situated features of robots seem essential to the learning process compared to the passive and virtual interaction with tablets. Future work should examine the effectiveness of a robot only, a tablet, and tablet and robot conditions on the children's learning outcomes.

6.4. Children's Perception

Interestingly, children's self-reported ratings of their mood were different for Robot and Teacher conditions, where children's mood was increased on average by 0.45 on a 5-point Likert scale after the Robot condition, while it was decreased on average by 0.05 points after the Teacher condition. On the other hand, children rated the teacher as more effective for learning in comparison to both the robot and the tablet aids.

Aligned with the Mood Change findings, children in the Robot condition rated their Likeability sorting of their learning aid type much higher than those who interacted in the Tablet and Teacher conditions. In addition, the robot's rating for being interesting was higher than this rating for the teacher. These results favoring the robot are important, since one of our main goals is to motivate and encourage children to learn the new script. Research has shown that affective responses, such as emotion and mood, are interwoven with learning and cognition, and it is hypothesized that positive mood leads to pleasant and open-minded cognitive experiences framed within "mood-dependent cognitive styles" (Hascher, 2010). Prior work (Bryan and Bryan, 1991; Bryan et al., 1996) has shown that children in a positive mood condition performed significantly better than children in a control group. Thus, we can assume that positive mood as an affective reaction might create a favorable learning environment, resulting in the enhancement of divergent thinking and task engagement (Pekrun, 1992; Efklides and Chryssoula, 2005). In HRI, researchers have started to investigate how social robot could benefit in making learning more efficient and more enjoyable (Movellan et al., 2009; Tozadore et al., 2017; Sandygulova and O'Hare, 2018; van den Berghe et al., 2019; Chen et al., 2020). (Johal, 2020) found that more than half of the recent studies in social robots for education evaluate the affective outcomes of the robot-learner interaction; and about 30% report both cognitive and affective outcomes. More generally, humanoid robots are suggested to provide positive peer-like interaction with children, broadly promoting enjoyment through the interaction. Our study shows that children's likeability and positive mood change bring significant benefits compared to other teaching approaches. However, the relationship between enjoyment and learning outcomes is still not clear (i.e. a causality, a correlation or a more complex relationship) (Girard et al., 2013). As such, investigations are needed to assess the added value provided by robot-assisted learning (which other teaching approaches otherwise lack) as well as a follow-up longitudinal study allowing to evaluate retention outcomes. In such future research, the effect of mood should also be integrated as related to students' learning outcomes.

6.5. Task Difficulty

Indeed, this experiment has brought up some questions of identifying effective learning scenarios and tools for learning a new script. Future studies can focus on vocabulary choice as it might benefit children to use their foreign language vocabulary resources to improve foreign script learning (e.g., Latin). Apart from this strategy, the use of unfamiliar linguistic items in the experiment might bring more promising results in order to not misinterpret children's existing knowledge. Consistent with what was investigated in this study, we are encouraged to make use of other strategies that might build a cognitive learning scenario with the presence of a social robot. We believe that interactions with the robot can involve several modes (verbal, visual, tactile) and be integrated in relation to all perceptual modalities, together with events on the tablet, its stylus data, and children's feedback.

6.6. Handwriting Recognition

To measure the children's handwriting performances, we developed handwriting recognition for the Cyrillic alphabet. The accuracy rate on a validation set using state-of-the-art algorithms, i.e., 784-500-500-2000 reported in Hinton and Salakhutdinov (2006) and CNN similar to Le-Net-5 (LeCun et al., 1998) with custom parameters is 98% on the Cyrillic-MNIST data set. However, the recognition of children's handwriting data was only 38%. This is reflected in other works that use adult datasets with child data: state-of-the-art speech recognition technologies (Kennedy et al., 2017) did not perform well with child speech, while age and gender determination did not perform well on children's faces (Sandygulova et al., 2014). As noted by Asselborn et al. (2018), the quality of handwriting performance can only be evaluated when considering the age and gender of children. The collection of a dataset on children's Cyrillic handwriting will, subsequently, allow us to adequately evaluate the quality of Cyrillic handwriting in real-time.

7. CONCLUSION

In this paper, the CoWriting Kazakh system and its proposed learning scenario were discussed in relation to the script conversion task in Kazakh, from Cyrillic to Latin in three conditions (a robot, tablet, and a teacher). Contributing to the HRI field, the main findings that can be drawn from this interdisciplinary study are: (1) tablets only and tablets along with robots have the potential to provide more-or-less similar learning gains as a teacher in the script learning scenario with children, since the three conditions did not show significant differences, however (2) robots are advantageous based on the significant positive mood change and children's responses that they liked the robot significantly more and considered it as significantly more interesting than other learning aids in the present study and in conclusions reached by previous studies (Park and Howard, 2013; Westlund et al., 2015; Vogt et al., 2019), (3) an open question remains as to whether gender difference is significant in regard to learning outcomes. Our study could not reach definitive conclusions since there were several limitations such as single-session intervention, relatively small sample size and the lack of only robot condition. However, there is an overall lack of such studies in the field of HRI that compare effectiveness of robotic systems as opposed to other learning aids. In essence, social robots can significantly impact children's learning as they tend to cultivate a responsive and friendly interaction (Belpaeme et al., 2018a; Kanero et al., 2018). Considering all the above, our future studies should aim for longitudinal interaction and further investigate gender difference, differentiated learning, the refinement of learning scenarios related to word choice, and adaptations for remote and online learning. We hope this study will increase scholarly attention towards the use of robots for script learning and handwriting practice. In order to effectively teach the Latin-based Kazakh alphabet, our essential purpose is to develop an adaptive system relying on differentiated learning strategies relevant to various learning scenarios and individuals.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Nazarbayev University Institutional Research Ethics Committee. Written informed consent to participate in this study was provided by the participants' legal guardian/next of kin. Written informed consent was obtained from the individual and minor's legal guardian for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

ZZ: investigation, data curation, visualization, and writing. AA: writing. MA, KK, and NB: investigation and data curation. BT and TA: software. WJ and PD: conceptualization, supervision, and funding acquisition. AC: supervision, funding acquisition, and writing. AS: conceptualization, supervision, funding acquisition, writing, investigation, and data curation. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the Nazarbayev University Collaborative Research Program grant, CoWriting Kazakh: Learning a New Script with a Robot (Award number is 091019CRP2107). This research was also partly funded by the Swiss State Secretariat for Education, Research and Innovation (SERI) as parts of the CIS Region Seeding Grants (Project SFG 1038), and the NCCR Robotics (SNSF).

ACKNOWLEDGMENTS

We would like to thank the children and their parents for participating in this research. The authors would also like to express great appreciation to the members of staff of the participating primary school for their time and help with the experiment.

REFERENCES

Al-Azami, S., Kenner, C., Ruby, M., and Gregory, E. (2010). Transliteration as a bridge to learning for bilingual children. *Int. J. Bilingual Educ. Bilingualism* 13, 683–700. doi: 10.1080/13670050903406335

Alemi, M., Meghdari, A., and Ghazisaedy, M. (2014). Employing humanoid robots for teaching english language in iranian junior high-schools. *Int. J. Humanoid Robot*. 11:1450022. doi: 10.1142/S0219843614500224

Altynsarina, E. (2018). *Kazakhstan Adopts New Version of Latin-Based Kazakh Alphabet*. Nur-Sultan: The Astana Times.

Alves-Oliveira, P., Arriaga, P., Paiva, A., and Hoffman, G. (2017). "Yolo, a robot for creativity: a co-design study with children," in *Proceedings of the 2017 Conference on Interaction Design and Children* (Stanford), 423–429. doi: 10.1145/3078072.3084304

Asselborn, T., Gargot, T., Kidziński, Ł., Johal, W., Cohen, D., Jolly, C., et al. (2018). Automated human-level diagnosis of dysgraphia using a consumer tablet. *NPJ Digit. Med*. 1, 1–9. doi: 10.1038/s41746-018-0049-x

Balkibekov, K., Meiirbekov, S., Tazhigaliyeva, N., and Sandygulova, A. (2016). "Should robots win or lose? Robot's losing playing strategy positively affects child learning," in *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (New York, NY: IEEE), 706–711. doi: 10.1109/ROMAN.2016.7745196

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. (2018a). Social robots for education: a review. *Sci. Robot*. 3:eaat5954. doi: 10.1126/scirobotics.aat5954

Belpaeme, T., Vogt, P., Van den Berghe, R., Bergmann, K., Göksun, T., De Haas, M., et al. (2018b). Guidelines for designing social robots as second language tutors. *Int. J. Soc. Robot*. 10, 325–341. doi: 10.1007/s12369-018-0467-6

Bryan, T. and Bryan, J. (1991). Positive mood and math performance. *J. Learn. Disabil*. 24, 490–494. doi: 10.1177/002221949102400808

Bryan, T., Mathur, S., and Sullivan, K. (1996). The impact of positive mood on learning. *Learn. Disabil. Q*. 19, 153–162. doi: 10.2307/1511058

Charisi, V., Gomez, E., Mier, G., Merino, L., and Gomez, R. (2020). Child-robot collaborative problem-solving and the importance of child's voluntary interaction: a developmental perspective. *Front. Robot. AI* 7:15. doi: 10.3389/frobt.2020.00015

Chase, C. C., Chin, D. B., Oppezzo, M. A., and Schwartz, D. L. (2009). Teachable agents and the protégé effect: increasing the effort towards learning. *J. Sci. Educ. Technol*. 18, 334–352. doi: 10.1007/s10956-009-9180-4

Chen, H., Park, H. W., Zhang, X., and Breazeal, C. (2020). "Impact of interaction context on the student affect-learning relationship in child-robot interaction," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge, UK: Association for Computing Machinery), 20, 389–397. doi: 10.1145/3319502.3374822

CohenMiller, A., and Pate, P. E. (2019). A model for developing interdisciplinary research theoretical frameworks. *Qualit. Rep*. 24, 1211–1226. Available online at: https://nsuworks.nova.edu/tqr/vol24/iss6/2

Crisp, S. (1990). "Soviet language planning 1917–53," in *Language Planning in the Soviet Union*, ed M. Kirkwood (London: Springer), 23–45. doi: 10.1007/978-1-349-20301-7_2

Dennis, M., Masthoff, J., and Mellish, C. (2016). Adapting progress feedback and emotional support to learner personality. *Int. J. Artif. Intell. Educ*. 26, 877–931. doi: 10.1007/s40593-015-0059-7

Dewi, R. K., Priandani, N. D., Brata, K. C., and Fanani, L. (2018). Usability evaluation of mobile-based application for javanese script learning media. *J. Inform. Technol. Comput. Sci*. 3, 88–93. doi: 10.25126/jitecs.201 83146

Edwards, B. I., and Cheok, A. D. (2018). Why not robot teachers: artificial intelligence for addressing teacher shortage. *Appl. Artif. Intell*. 32, 345–360. doi: 10.1080/08839514.2018.1464286

Efklides, A. and Chryssoula, P. (2005). Effects of mood on students' metacognitive experience. *Learn. Instr*. 15, 415–431. doi: 10.1016/j.learninstruc.2005.07.010

Garcia, E., and Weiss, E. (2019). *The Teacher Shortage Is Real, Large and Growing, and Worse Than We Thought. The First Report in "The Perfect Storm in the Teacher Labor Market" Series*. Washington: Economic Policy Institute.

Girard, C., Ecalle, J., and Magnan, A. (2013). Serious games as new educational tools: how effective are they? a meta-analysis of recent studies. *J. Comput. Assist. Learn*. 29, 207–219. doi: 10.1111/j.1365-2729.2012.00489.x

Gonzalez, C., Anderson, J., Culmer, P., Burke, M., Mon-Williams, M., and Wilkie, R. (2011). Is tracing or copying better when learning to reproduce a pattern? *Exp. Brain Res*. 208, 459–465. doi: 10.1007/s00221-010-2482-1

Gordon, G., and Breazeal, C. (2015). "Bayesian active learning-based robot tutor for children's word-reading skills," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*. Austin.

Han, J., and Kim, D. (2009). "R-learning services for elementary school students with a teaching assistant robot," in *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (La Jolla: IEEE), 255–256. doi: 10.1145/1514095.1514163

Hascher, T. (2010). Learning and emotion: Perspectives for theory and research. *Eur. Educ. Res. J*. 9, 13–28. doi: 10.2304/eerj.2010.9.1.13

Hatcher, L. (2008). Script change in Azerbaijan: acts of identity. *Int. J. Sociol. Lang*. 2008, 105–116. doi: 10.1515/IJSL.2008.038

Hinton, G. E., and Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science* 313, 504–507. doi: 10.1126/science.1127647

Hood, D., Lemaignan, S., and Dillenbourg, P. (2015). "The cowriter project: teaching a robot how to write," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts* (Prtland), 269. doi: 10.1145/2701973.2702091

Huong, L. P. H. (2007). The more knowledgeable peer, target language use, and group participation. *Can. Mod. Lang. Rev*. 64, 329–350. doi: 10.3138/cmlr.64.2.329

Hyun, E., Kim, S., Jang, S., and Park, S. (2008). "Comparative study of effects of language instruction program using intelligence robot and multimedia on linguistic ability of young children," in *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication* (Munich: IEEE), 187–192. doi: 10.1109/ROMAN.2008.4600664

Jacq, A., Lemaignan, S., Garcia, F., Dillenbourg, P., and Paiva, A. (2016). "Building successful long child-robot interactions in a learning context," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Christchurch: IEEE), 239–246. doi: 10.1109/HRI.2016.7451758

Jamet, F., Masson, O., Jacquet, B., Stilgenbauer, J.-L., and Baratgin, J. (2018). Learning by teaching with humanoid robot: a new powerful experimental tool to improve children's learning ability. *J. Robot*. 2018:4578762. doi: 10.1155/2018/4578762

Jobeen, A., Kazemian, B., and Shahbaz, M. (2015). The role of error analysis in teaching and learning of second and foreign language. *Educ. Linguist. Res*. 1, 52–62. doi: 10.5296/elr.v1i2.8189

Johal, W. (2020). Research Trends in Social Robots for Learning. *Curr. Robot. Rep*. 1–9. doi: 10.1007/s43154-020-00008-3

Kadirova, R.-B. (2018). *The Sociolinguistic Attitudes of Kazakhs Towards the Latin Alphabet and Orthography Reform in Kazakh*. North Carolina: University of North Carolina at Chapel Hill Graduate School.

Kanda, T., Sato, R., Saiwaki, N., and Ishiguro, H. (2007). A two-month field trial in an elementary school for long-term human-robot interaction. *IEEE Trans. Robot*. 23, 962–971. doi: 10.1109/TRO.2007.904904

Kanero, J., Geckin, V., Oran?., C., Mamus, E., Küntay, A. C., and Goksun, T. (2018). Social robots for early language learning: Current evidence and future directions. *Child Dev. Perspect*. 12, 146–151. doi: 10.1111/cdep.12277

Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2016). "Social robot tutoring for child second language learning," in *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Christchurch: IEEE), 231–238. doi: 10.1109/HRI.2016.7451757

Kennedy, J., Lemaignan, S., Montassier, C., Lavalade, P., Irfan, B., Papadopoulos, F., et al. (2017). "Child speech recognition in human-robot interaction: evaluations and recommendations," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (Vienna), 82–90. doi: 10.1145/2909824.3020229

Kim, A., Omarova, M., Zhaksylyk, A., Asselborn, T., Johal, W., Dillenbourg, P., et al. (2019). "Cowriting Kazakh: transitioning to a new latin script using social robots," in *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (New Delhi: IEEE), 1–6. doi: 10.1109/RO-MAN46459.2019.8956471

Klein, J. T. (1990). *Interdisciplinarity: History, Theory, and Practice*. Detroit: Wayne State University Press.

Lakens, D., Scheel, A. M., and Isager, P. M. (2018). Equivalence testing for psychological research: a tutorial. *Adv. Methods Pract. Psychol. Sci*. 1, 259–269. doi: 10.1177/2515245918770963

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 2278–2324. doi: 10.1109/5.726791

Lemaignan, S., Jacq, A., Hood, D., Garcia, F., Paiva, A., and Dillenbourg, P. (2016). Learning by teaching a robot: the case of handwriting. *IEEE Robot. Autom. Mag*. 23, 56–66. doi: 10.1109/MRA.2016.25 46700

Li, J. (2015). The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents. *Int. J. Hum. Comput. Stud*. 77, 23–37. doi: 10.1016/j.ijhcs.2015.01.001

Liu, T.-C., and Chang, M. (2008). "Human-robot interaction research issues of educational robots," in *2008 Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning* (Banff: IEEE), 209–210. doi: 10.1109/DIGITEL.2008.31

Lubold, N., Walker, E., Pon-Barry, H., Flores, Y., and Ogan, A. (2018). *Using Iterative Design to Create Efficacy-Building Social Experiences With a Teachable Robot*. London: International Society of the Learning Sciences, Inc. [ISLS].

Luckin, R., Holmes, W., Griffiths, M., and Forcier, L. B. (2016). *Intelligence Unleashed: An Argument for AI in Education*. London: Pearson.

Markopoulos, P., Read, J. C., MacFarlane, S., and Hoysniemi, J. (2008). *Evaluating Children's Interactive Products: Principles and Practices for Interaction Designers*. San Francisco: Elsevier.

Meghdari, A., Alemi, M., Ghazisaedy, M., Taheri, A., Karimian, A., and Zandvakili, M. (2013). "Applying robots as teaching assistant in efl classes at Iranian middle-schools," in *Proceedings of the International Conference on Education and Modern Educational Technologies (EMET-2013)* (Venice).

Michaelis, J. E., and Mutlu, B. (2018). Reading socially: transforming the in-home reading experience with a learning-companion robot. *Sci. Robot*. 3:eaat5999. doi: 10.1126/scirobotics.aat5999

Movellan, J. R., Eckhardt, M., Virnes, M., and Rodriguez, A. (2009). "Sociable robot improves toddler vocabulary skills," in *2009 4th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (La Jolla, CA: ACM). 307–308. doi: 10.1145/1514095.1514189

Mubin, O., Alhashmi, M., Baroud, R., and Alnajjar, F. S. (2019). "Humanoid robots as teaching assistants in an Arab school," in *Proceedings of the*

31st Australian Conference on Human-Computer-Interaction (Perth/Fremantl), 462–466. doi: 10.1145/3369457.3369517

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., and Dong, J.-J. (2013). A review of the applicability of robots in education. *J. Technol. Educ. Learn*. 1:13. doi: 10.2316/Journal.209.2013.1.209-0015

Neumann, M. M. (2020). Social robots and young children's early language and literacy learning. *Early Child. Educ. J*. 48, 157–170. doi: 10.1007/s10643-019-00997-7

Newton, D., and Newton, L. (2019). Humanoid robots as teachers and a proposed code of practice. *Front. Educ*. 4:125. doi: 10.3389/feduc.2019.00125

Park, H. W., and Howard, A. (2013). "Providing tablets as collaborative-task workspace for human-robot interaction," in *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Tokyo: IEEE), 207–208. doi: 10.1109/HRI.2013.6483574

Pekrun, R. (1992). The impact of emotions on learning and achievement: Towards a theory of cognitive/motivational mediators. *Appl. Psychol*. 41, 359–376. doi: 10.1111/j.1464-0597.1992.tb00712.x

Rivlina, A. (2016). Global english-related digraphia and Roman-Cyrillic biscriptal practices. *Proc. Soc. Behav. Sci*. 236, 207–212. doi: 10.1016/j.sbspro.2016.12.006

Rosenberg-Kima, R., Koren, Y., Yachini, M., and Gordon, G. (2019). "Human-robot-collaboration (HRC): social robots as teaching assistants for training activities in small groups," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Daegu: IEEE), 522–523. doi: 10.1109/HRI.2019.8673103

Rusticus, S., and Lovato, C. (2011). Applying tests of equivalence for multiple group comparisons: demonstration of the confidence interval approach. *Pract. Assess. Res. Eval*. 16, 1–6. doi: 10.7275/d5wf-5p77

Sandygulova, A., Dragone, M., and O'Hare, G. M. (2014). "Real-time adaptive child-robot interaction: age and gender determination of children based on 3d body metrics," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication* (Edinburgh: IEEE), 826–831. doi: 10.1109/ROMAN.2014.6926355

Sandygulova, A., Johal, W., Zhexenova, Z., Tleubayev, B., Zhanatkyzy, A., Turarova, A., et al. (2020). "Cowriting Kazakh: learning a new script with a robot," in *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (Cambridge), 113–120. doi: 10.1145/3319502.3374813

Sandygulova, A. and O'Hare, G. M. (2018). Ageand gender-based differences in children's interactions with a gender-matching robot. *Int. J. Soc. Robot*. 10, 687–700. doi: 10.1007/s12369-018-0472-9

Sharkey, A. J. (2016). Should we welcome robot teachers? *Ethics Inform. Technol*. 18, 283–297. doi: 10.1007/s10676-016-9387-z

Sysoev, I., Hershman, A., Fine, S., Traweek, C., and Roy, D. (2017). "Speechblocks: a constructionist early literacy app," in *Proceedings of the 2017 Conference on Interaction Design and Children* (Stanford), 248–257. doi: 10.1145/3078072.3079720

Tazhigaliyeva, N., Diyas, Y., Brakk, D., Aimambetov, Y., and Sandygulova, A. (2016). "Learning with or from the robot: exploring robot roles in educational context with children," in *International Conference on Social Robotics* (Kansas City: Springer), 650–659. doi: 10.1007/978-3-319-47437-3_64

Tozadore, D. C., Pinto, A. H. M., Ranieri, C. M.,Batista, M. R., and Romero, R. A. F. (2017). "Tablets and humanoid robots as engaging platformsfor teaching languages," in *2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR)* (Curitiba: IEEE). 1–6. doi: 10.1109/SBR-LARS-R.2017.8215290

Tung, F.-W. (2016). Child perception of humanoid robot appearance and behavior. *Int. J. Hum. Comput. Interact*. 32, 493–502. doi: 10.1080/10447318.2016.1172808

United Nations (UN). (2015). *Transforming Our World: The 2030 Agenda for Sustainable Development*. New York, NY: General Assembley 70 session.

van den Berghe, R., Verhagen, J., Oudgenoeg-Paz, O., van der Ven, S., and Leseman, P. (2019). Social robots for language learning: a review. *Rev. Educ. Res*. 89, 259–295. doi: 10.3102/0034654318821286

Vogt, P., van den Berghe, R., de Haas, M., Hoffman, L., Kanero, J., Mamus, E., et al. (2019). "Second language tutoring using social robots: a large-scale study," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (Daegu: IEEE), 497–505. doi: 10.1109/HRI.2019.8673077

Vygotsky, L. S. (1980). *Mind in Society: The Development of Higher Psychological Processes*. Cambridge: Harvard University Press.

Westlund, J. K., Dickens, L., Jeong, S., Harris, P., DeSteno, D., and Breazeal, C. (2015). "A comparison of children learning new words from robots, tablets, & people," in *Proceedings of the 1st International Conference on Social Robots in Therapy and Education* (Almere).

Yadollahi, E., Johal, W., Paiva, A., and Dillenbourg, P. (2018). "When deictic gestures in a robot can harm child-robot collaboration," in *Proceedings of the 17th ACM Conference on Interaction Design and Children* (Trondheim), 195–206. doi: 10.1145/3202185.3202743

Yanikoglu, B., Gogus, A., and Inal, E. (2017). Use of handwriting recognition technologies in tablet-based learning modules for first grade education. *Educ. Technol. Res. Dev*. 65, 1369–1388. doi: 10.1007/s11423-017-9532-3

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Advantages of publishing in Frontiers

OPEN ACCESS
Articles are free to read for greatest visibility and readership

FAST PUBLICATION
Around 90 days from submission to decision

HIGH QUALITY PEER-REVIEW
Rigorous, collaborative, and constructive peer-review

TRANSPARENT PEER-REVIEW
Editors and reviewers acknowledged by name on published articles

Frontiers
Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org
Contact us: info@frontiersin.org | +41 21 510 17 00

REPRODUCIBILITY OF RESEARCH
Support open data and methods to enhance research reproducibility

DIGITAL PUBLISHING
Articles designed for optimal readership across devices

FOLLOW US
@frontiersin

IMPACT METRICS
Advanced article metrics track visibility across digital media

EXTENSIVE PROMOTION
Marketing and promotion of impactful research

LOOP RESEARCH NETWORK
Our network increases your article's readership