Do Children Perceive Whether a Robotic Peer is Learning or Not?

Shruti Chandra[1,2], Raul Paradeda[2,3], Hang Yin[1,2], Pierre Dillenbourg[1], Rui Prada[2] and Ana Paiva[2]
[1]École Polytechnique Fédérale de Lausanne, Switzerland.
firstname.lastname@epfl.ch
[2]INESC-ID & Instituto Superior Técnico, University of Lisbon, Portugal.
firstname.lastname@tecnico.ulisboa.pt; ana.paiva@inesc-id.pt
[3]State University of Rio Grande do Norte, Brazil.

ABSTRACT

Social robots are being used to create better educational scenarios, thereby fostering children's learning. In the work presented here, we describe an autonomous social robot that was designed to enhance children's handwriting skills. Exploiting the benefits of the *learning-by-teaching* method, the system provides a scenario in which a child acts as a teacher and corrects the handwriting difficulties of the robotic agent. To explore the children's perception towards this social robot and the effect on their learning, we have conducted a multi-session study with children that compared two contrasting competencies in the robot: *'learning'vs 'non-learning'* and presented as two conditions in the study. The results suggest that the children learned more in the learning condition compared with the non-learning condition and their learning gains seem to be affected by their perception of the robot. The results did not lead to any significant differences in the children's perception of the robot in the first two weeks of interaction. However, by the end of the 4^{th} week, the results changed. The children in the learning condition gave significantly higher *writing ability* and *overall performance* scores to the robot compared with the non-learning condition. In addition, the change in the robot's learning capabilities did not show to affect their perceived intelligence, likability and friendliness towards it.

KEYWORDS

Social robotics, learning-by-teaching, multi-session studies

ACM Reference Format:
Shruti Chandra[1,2], Raul Paradeda[2,3], Hang Yin[1,2], Pierre Dillenbourg[1], Rui Prada[2] and Ana Paiva[2]. 2018. Do Children Perceive Whether a Robotic Peer is Learning or Not?. In *HRI '18: 2018 ACM/IEEE International Conference on Human-Robot Interaction, March 5–8, 2018, Chicago, IL, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3171221.3171274

1 INTRODUCTION

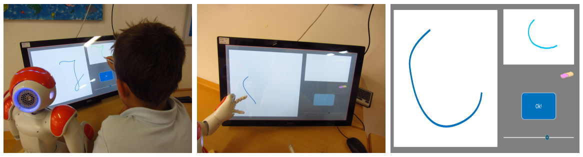
Our work focuses on the question of how a social robot could help children acquire handwriting skills. Over the past several years, robots have been introduced into educational contexts seeking different ways of fostering engagement in learning activities. A significant part of the work in educational robotics utilises robots as tutors; however, some researchers have started using the robots as tutees based on the peer-assisted educational method named as *learning-by-teaching* or *peer tutoring* [9, 24]. Besides academic benefits of the method, it is also believed to engage and motivate students [33, 37]. For example, Kanda et al., used Robovie, a humanoid robot, as an English peer-tutor for Japanese students [24] and concluded that the robot encouraged some of the students to improve their English and form relationships with them. Similarly, Hood et al. also found that the Nao robot[1] could improve children's writing skills and engage them successfully [21]. However, *how do the children perceive these robots? How do these perceptions change over multi-session interactions? Do the children's perceptions affect their learning?* The children's perception of the robots indeed seems to be relevant in child-robot interactions [23, 29]. In fact, T. N. Beran et al. asked questions about a robot to 198 children and found that a significant proportion of children were able to ascribe behavioural, cognitive characteristics to the robot [7]. One of the aims of this method is to maximise the learning gains of students. In the current work, we exploit this method with the goal of answering a question: *how would the learning capabilities of a tutee impact the learning of a tutor?* Therefore, in our current study, we design a child-robot scenario, where a child acts as a *teacher (tutor)* and a robot acts as a *learner (tutee)*. The use of a robot allows us to manipulate its capabilities and behaviour in a controlled way. To have a clear differentiation in the learning capability, we have chosen two extreme capabilities of the learner-robot: *'learning'vs 'non-learning'*.

To test how a social robot could help children to acquire handwriting skills: a system can be built to investigate an educational scenario that includes different aspects such as *modes of child-robot interactions*; *children's perceptions of the robot's capabilities and behaviour*; and *handwriting difficulties.* In the work described here, we developed a robot with autonomous social behaviour that provides an educational scenario for children to improve their handwriting skills. The system relies on the *learning-by-teaching* method in which a child helps the robot improve its handwriting skills by providing the *corrective feedback* [11] on the robot's writing. Further, we aimed to understand how children perceive the agent's behaviour and capabilities since they can impact the child-robot learning interaction. Aspects such as children's learning gains, engagement, and perception may be altered [16, 25]. Past research has

HRI '18, March 5–8, 2018, Chicago, IL, USA

ACM ISBN 978-1-4503-4953-6/18/03…$15.00
https://doi.org/10.1145/3171221.3171274

[1]Aldebaran robotics: `https://www.aldebaran.com/en`



(a) (b) (c)

Figure 1: Experimental setup: (a) When the child provided correction; (b) When the robot was writing; (c) Interactive features in the computer application.

shown that multiple sessions of handwriting practice are required to improve children's handwriting skills [22]. Also in the context of child-robot studies, it was shown that the children tend to change their attitude towards a social robot over time [24]. Thus, we tested the system by conducting a multi-session study in a school to explore children's perceptions towards a social robot exhibiting two contrast competencies: *learning and non-learning* and how these may affect children's learning gains.

2 RELATED WORK

The *learning-by-teaching* method provides an unambiguous and consistent distinction between the teaching and learning roles among the students [12] and is more effective when it includes *peer assessment* such as corrective feedback, which in turn benefits the tutor in achieving the clarity, self-concept and organisation of the domain knowledge [10, 12, 37]. According to Hattie *et al.* [20], *"feedback is powerful in its effect when it is addressed to a learning context and it is most powerful when it addresses faulty interpretations"*. In the presented work, we used corrective feedback as a peer assessment approach, where the tutor child gave feedback on the robot's handwriting. Furthermore, the learning-by-teaching approach is also explored based on *Role Theory* [35]. According to the role theory, the enactment of a role such as a teacher or a learner affects the behaviour, attitudes and perceptions of the participants consistent with role expectations [8]. Tutor-children attribute prestige, authority, self-esteem [1] and their perceived *self-efficacy* towards tutoring and this has also been associated with academic achievement [4, 5, 40]. Besides, the performance of the tutee may also affect tutor's attributions towards the tutee [2].

In child-robot scenarios, children's perception towards a robotic agent is related to several aspects such as the robot's role, capabilities and physical or nonphysical behaviour [16, 25]. Oliveira et al. [3] found that the children's perceptions of a tutor robot's role changed over time. In the present work, we further investigate the aspects of role-theory, children's self-efficacy towards tutoring and the children's perception of the robot's role.

Handwriting is a complex blend of motor and cognition skills. Ineffective motor skills are difficult to change, so it is important to sharpen the skills of preschoolers from the beginning [28, 36]. Recent studies have also explored the learning-by-teaching method to improve children's writing capabilities [21, 27]. Shizuko et al. performed a study using the method to enhance children's knowledge of English words, in which children drew some shapes associated with the words. The results suggested that the teleoperated robot capable of learning helped children learn unknown English words for the shapes [27]. Following a similar line of work, this paper makes three contributions: first, by the creation of an autonomous behaviour of a social robot designed to provide a learning scenario in which a child learns by teaching the robot, in a natural setting, by sitting next to each other; second, the understanding of the children's perception of the robot's capabilities (particularly the learning capability), which will help in the design of better educational scenarios with social robots; and third, by exploring the effects of children's enactment as a tutor and their perception of the robot on their learning.

(a) (b) (c)

Figure 2: Handwriting issues used for generating the deformed letters: (a) Proportion; (b) Breaks; (c) Alignment

3 A ROBOT THAT LEARNS FROM A CHILD

3.1 Interaction Design

Our experimental setup involves a child sitting side-by-side with an autonomous Nao robot and a touchscreen used by both to write: the learner-robot writes a deformed letter on the screen and asks the teacher-child to correct it (Fig. 1(a)). The teacher-child then corrects the letter and demonstrates a sample of the same letter by writing on the other side of the screen. In our design, the turn-taking occurs when the robot prompts the child for correction once it finishes writing (using sentences like 'could you show me the correct shape of this letter?') and when the child presses a button on the screen after finishing the correction.

3.2 Handling deformed letters

Generating the deformed letters for the robot and demonstrating the correct letter (from the child side) are key points. The most common handwriting issues faced by children are related to overall neatness, letter formation, and alignment [18]. Instead of generating random deformations in the letters, we targeted three common handwriting issues as described below:

Proportion: the issue relates to distortion in a sub-part or sub-parts in the letter's trajectory. As shown in Fig.2(a), the sub-part (the first part of the trajectory of the letter) is not in proportion with respect to the rest of the trajectory.

Breaks: the issue arises when there is a 'hole 'present in the letter's trajectory, resulting in a discontinuous shape (Fig.2(b)).

Alignment: the issue relates to nonalignment of the letter, for example, rotation of the letter at some angle (Fig. 2(c)).

The choice of the three handwriting issues is due to their prevalence in children's handwriting [18, 19]. For handling the *proportion* and the *breaks* issues, we used an algorithm proposed in [39], which estimates motion parameters associated with stroke deformities in an imitation learning problem. The choice of using the algorithm is motivated by two reasons: first, it has the power to synthesize and learn the multiple-mode motion trajectories, with the integration of rapid extraction and representation. Second, we wanted to generate deformed letters for the robot that closely resembles children's handwriting. The algorithm incorporates human movement inspired features and generalises the synthesis to generate poor or good written samples of a letter. The input of the algorithm

Algorithm 1 Learning with curvature-based features for modelling handwriting motion

Require: $\mathcal{D} = \{\varsigma_i\}$, M
Ensure: $\mathcal{D}^m_{k=1:K_m}$, $\hat{\theta}^m_k = \{{\mu_f}^m, {\Sigma_f}^m\}$, $m = 1, \ldots, M$
  $\mathcal{D}^m_{k=1:K_m} \leftarrow$ Partitioning($\mathcal{D}$, M) ▷ Build locally similar dataset $\mathcal{D}$ through partitioning
  for all m in $1{:}M$ **do**
    $${\mu_\varsigma}^m, {\Sigma_\varsigma}^m \leftarrow \underset{\theta}{\operatorname{argmax}} \sum_{i=1}^{|\mathcal{D}^m_k|} \log \mathcal{N}(\varsigma^*_i|\theta)$$
    ${\mu_f}^m \leftarrow \mathrm{RXZERO}({\mu_\varsigma}^m)$
    ${\Sigma_f}^m \leftarrow J(\mu^m_\varsigma)^T \Sigma^m_\varsigma J(\mu^m_\varsigma)$
  end for

Figure 3: The blue trajectory denotes the letter input to the algorithm while the red denotes the output. The letter 'S'shows the *proportion*; the letter 'B'(right) represents the *break* and the left one represents the *proportion* and *break* handwriting issues.

is a planar handwriting data set $\mathcal{D}$ (see Algorithm 1). We created the dataset $\mathcal{D}$ in advance and used the algorithm to generate the deformed letters based on children's handwriting errors collected in previous studies [13–15]. For generating similar instances of a letter from the dataset, the Gaussian statistics of trajectories are extracted. The parameters received from the statistics are further converted to a curvature-based representation with nonlinear feature embedding. The mean $\mu^m_f = \{A_j, z^j_0, \mu_j, \sigma_j, \alpha^j_s, \alpha^j_e\}_{j=1:N}$ is obtained by an RXZero process [30], which effectively solves a nonlinear optimization to match the trajectories through Equation (1) and (2). The two equations determine the stroke velocity $|v|$ and the path angular direction ϕ_f on the time phase z and are sufficient to reconstruct letter trajectories. The covariances are transformed through local linear projections, with $J(\cdot)$ denoting the Jacobian of the reconstruction. It is easy to find that some of the extracted parameters are correlated to the motion path in an explicit manner.

$$|v(z)| = \sum_{j=1}^{N} \frac{A_j}{\sqrt{2\pi}\sigma_j(z - z^j_o)} \exp\left(-\frac{(\ln(z - z^j_0) - \mu_j)^2}{2\sigma^2_j}\right) \tag{1}$$

$$\phi_j(z) = \alpha^j_s + \frac{\alpha^j_e - \alpha^j_s}{2}\left(1 + \mathrm{erf}\left(\frac{\ln(z - z^j_0) - \mu_j}{2\sigma_j}\right)\right) \tag{2}$$

For instance, in Equation 1, A_j is proportional to the velocity magnitude such that the size of specific letter components can be modulated without influencing the other components. Equation 2 shows that α^j_s and α^j_e anchor the start and end angular position. Hence, the two parameters allow the stroke orientation or the curvature to be adjusted; see Fig. 3. Overall, the algorithm generates the full letter and capable of improving or deforming the shape of the letter. For the *alignment* issue, we developed an algorithm to produce the rotation effect with different angles. As a result, this provides a digital approach to control the letter deformities, hence allowing the robot agent to exhibit a spectrum of levels of handwriting skills.

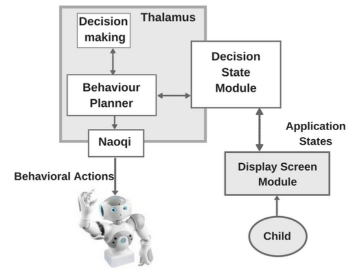


Figure 4: System architecture

3.3 Writing Application

Because the interaction between the robot and a child involves writing/correcting a deformed letter on the screen, we developed two applications for the writing activity. The first application, named *computer application*, provided the interactive capability to the teacher-child to correct the shapes of deformed letters written by the learner-robot. By using the interactive features of the application, the child could write, erase and correct the shapes drawn on the screen. As shown in Fig. 1(c), the application interface includes a screen showing a large white box for the robot to write in (left side of the screen), a small white box (right side) for the child to write in, a small button for erasing the shapes, a blue button to move to the next screen and a slider. The slider allowed children to alter the shape of a letter since it was linked to the algorithms, the movement in the slider produced a change in the algorithm parameters resulting in a change in the letter's shape. In addition to these features, the application was connected to the robot to notify the current state of the writing activity. The second application, called *tutorial application*, was used by children before interacting with the robot to become familiar with the application features. It has the same features as the first application but without integrating the connection with the robot. An Android application called *test application* was developed specifically for the pre-/post-test, which aims to evaluate the learning gains of the children (details in Section 4.1). Each screen would display three different shapes of a letter, space for writing the same letter, an eraser button and a next button to move to the next letter.

3.4 System Architecture

Fig. 4 shows the overall system architecture, which was distributed in several modules: a *display screen module* contained the computer application including all algorithms, letter trajectories, *decision state module*, *the Thalamus* [31], the Naoqi API (provided by the Aldebaran) and the Nao robot. The display screen module and the robot were connected through a high-level integration framework, called *Thalamus*, which can accommodate the social robots with a multimedia application [32]. In addition, we used a semi-autonomous behaviour planner called *Skene* [32] to provide the robot's behavioural actions such as gazing, pointing, speech and gestures. The *decision state module* was developed to provide the communication among the display screen module, the Thalamus, the Skene and the Noaqi API. The essential functionality of the module was to inform the robot about the current states of the computer application through the Thalamus. Each interactive feature (letter screen, button, small writing box, slider; see Fig.1(c)) in the computer application refers to a state. When the child interacts with any of these states, this module determines the action for the robot to perform. The decisions include whether to perform an action, the type of the action and when the action should be performed. The robot's actions cover waiting, performing gestures without interacting with the child, and interacting with the child using gestures and dialogues. By receiving the action (according to the current state), the robot would know the current working status of the child.

To provide timely interaction, the decision module would check the application state and add the waiting timer in behavioural actions for the robot to react. Because we wanted all the children to perform a correction on the same number of letters, we partially facilitated the children to work at their own pace depending on the application state. For example, when the child finishes the correction of a letter, s/he can move to another letter without waiting, and when the robot is writing a letter, the child must wait until the robot finishes writing and for the correction process. Moreover, if the application state changes quickly depending on the child's speed, the system will cancel the behavioural actions of the robot to provide timely interaction between the child and the robot. The hand-drawing movements of the robot were animated for writing all the letters on the screen and timely synchronised with the computer application (when the letter appears point-by-point on the screen). Hence, the robot is fully autonomous in the interaction with the child.

4 THE STUDY

Our study consists of a between-subjects design with two conditions: *learning & non-learning*. The scenario involves a learner-robot which seeks help from the teacher-child for correcting the shape of a few letters. In the learning condition, the robot shows progress after each interaction (that is, it is actually becoming competent in learning how to write). In the non-learning condition, the robot exhibits a consistent yet still social behaviour, throughout the study, but it does not learn. Each child interacted with the robot four times with an interaction gap of 4-5 days. We state the research questions and hypotheses as follows:

Q1-Can the children respond differently to the robot's contrast learning capabilities (related to the robot's overall performance, writing capability and handwriting improvement)?
Because the robot shows two contrasting competencies, we hypothesise that the children would be able to recognise the learning capabilities of the robot in each condition.

Q2-Would children in the learning condition consider themselves a better teacher than in the non-learning condition?
In the study, children's perceived self-efficacy towards tutoring the robot means how good children consider themselves as a 'teacher'. Thus, we hypothesised that by the end of the study, the children in the learning condition would perceive themselves better teachers compared to the other condition because they may believe that their teaching abilities improved the robot's writing skills.

Q3- Would the two different competencies (learning and non-learning) of the robot affect the children's perceived likability and friendliness towards it?
We hypothesise that if the children could differentiate the two competencies, they would like the robot more in the learning condition because they would have more positive interactions with the robot.

Q4- Would the two competencies affect children's learning gains? Which competency would affect more?
We hypothesise that the children would benefit in both conditions regarding handwriting skills. However, the children in the learning condition may benefit more by seeing the robot's continuous improvement.

Material & Participants: the material in the study included a computer with a touchscreen, stylus, tablet (for pre- and post-test), video camera, microphone and Nao robot (only torso part). As shown in the past research, several letters in the English alphabet (lowercase - (b, q, d, k, g, p, z, u, j, n, a), uppercase - (K, Y, Z, W, R, M, F, D)) are problematic for primary grade children [18, 19]; we targeted letters from this category only. For each interaction, we prepared a different set of deformed letters, avoiding the conflict of using the same letters with the same handwriting issues. We conducted the study in a local private school, 'Colégio da Fonte'in Porto Salvo, Portugal. Twenty-five children participated from the 7- to 9-year-old age group (1st and 2nd grade) over a period of 1 month. Thirteen children (M=7.92; SD= 0.82; 5 male and 8 female) participated in the *learning condition*, and 12 children (M=8.08; SD=.75 years old; 7 male and 5 female) participated in the *non-learning condition*. Only children who assented and whose parents signed a consent form participated in the study.

4.1 Protocol

The children were randomly assigned to one of the two conditions to perform the collaborative writing activity with the robot. Each session was performed with a child and a robot and lasted approximately 13-15 minutes. The study was set up in a computer classroom in the school, and the children were comfortable with the classroom settings. The robot was named *Michael* to provide friendly interaction. During the introduction of the study, the researcher explained to the participants that they were going to interact with the robot 4 times (once per week) for the writing activity because the robot needed help in improving its handwriting skills. The researcher also explained the role of the robot as a learner and the participants' role as a teacher. Moreover, the participants were also informed that if they did not wish to continue, they could leave the study anytime they wanted. However, no participant left the study in between. The study was organized into following steps:

Tutorial: in this step, a researcher would bring a child into the classroom and explain the features of the tutorial application. The child then performed the tutorial activity on the monitor in the presence of the researcher. This step was crucial because all participants were young and needed the understanding to use the application. All children performed this step only before the first interaction.

Pre-test: the researcher would ask the child to perform the pre-test on the tablet. The pre-test application is developed to test the child's knowledge about the shape of a letter. This knowledge was examined based on two legibility factors: first, how well they could recognise the most correct shape of the letter; second how well they could demonstrate the letter by rewriting it. Thus, 3 different shapes of the same letter were displayed on the screen, and the children had to choose the most correct letter and write it on the same screen. The process was repeated for the remaining 8 letters.

Teaching activity with the robot: this phase involved the child-robot interaction. The researcher would bring the child into the classroom and leave him/her alone with the robot.

-Welcome greeting: the robot greeted the child and expressed its writing difficulties while showing its poor grades (fictitious grades it received from its fictitious teacher). This step was important to create an environment in which the child would find the robot in need of help.

-Correction period: the robot wrote a letter on the screen (Fig. 1(b)) and asked the child to correct it by using a slider. After that, the robot asked the child to demonstrate the same letter in the small white box (right side) (Fig. 1(a)). After finishing the correction and writing, the child advanced to the next screen and repeated the process for the remaining 15 letters.

-Goodbye greeting: after the correction period was completed, the robot would thank the child for the help. Moreover, the robot also informed the child that it would practice these letters until the next interaction.

Post-test: the researcher would ask the children to perform a post-test, identical to the pre-test.

Interview Questions: after finishing the teaching activity with the robot, the researcher would perform an interview with the child by asking a few questions for 10-12 min. As shown in Table 1 and 2, the questions were divided into five categories and were based on categorical and 5-point Likert scale respectively. Before interviewing the children, the researcher would emphasise that honest answers are important for the robot's learning, and the answers would not be disclosed to the robot.Then, the researcher would present the questions to the children on a sheet of paper in a child-friendly manner. For example, for the Q1 (3^{rd} category, Table 2), the values of the Likert scale (1,2,3,4,5) were represented with stars (*, **, ***, ****, *****). Furthermore, these questions were asked at three-time intervals, immediately after the second, third and fourth interactions with the robot.

Learning vs Non-learning capability: in the learning condition, the robot would show its progression in two forms: by writing the improved letter and by showing improved grades. The robot's handwriting ability and grades would improve at a constant rate from low to high. It would show the grades as: session1, grade = 2/5; . . . session4, grade = 5/5. In the non-learning condition, the robot would display consistently unimproved letters and constant grades throughout the sessions (session 1 to 4, grade = 2/5). We chose the two parameters (the handwriting and the grades) to show the robot's performance explicitly to the children. The robot would show its grades at the beginning of each session and then perform the writing activity. Besides, the robot showed same social behaviour and interaction in both conditions.

5 RESULTS

We collected the data from the questionnaires and logs of the tablets. The results of the Shapiro-Wilk test did not show the normal distribution of the data, so we analysed the data by using the non-parametric tests. We used different tests for the data and describe as follows: *Man-Whitney U Test-* (Table 2, 3^{rd} category) and learning gains; *correlations-* (Table 2, 4^{th} category) and (Table 2, 5^{th} category); *Sign test-* pre-and post-test scores; and *Chi-square test-*(Table 1, 2^{nd} category).

Table 1: Interview Questions in 1^{st} & 2^{nd} category

Robot's Grades (1^{st} category)†	Options
Q1. Have you observed Michael's grades? How well did it score?	Yes/No, score(value)
Q2. Was the grades improving or not?	Improved/Not Improved
Children's Perceived role of the robot (2^{nd} category) [3]	**Options**
Q1. What do you think Michael writes like a?	Child (younger than you) your friend (same age) your parents you/your teacher
Q2. How do you consider Michael as a?	Brother/Classmate Stranger/Relative Friend/Parent/Teacher

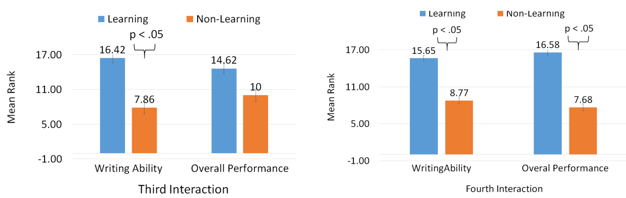
†**Based on study's research questions**

5.1 Robot's Grades

The questions assessed in the 1^{st} category correspond to the children's awareness of the robot's grades (Table 1, 1^{st} category). We found that all children were aware of the robot's grades in terms of the score (Q1) and the improvement made (Q2) since they answered both questions accurately according to each condition. The result suggests that the children indeed paid attention to the robot's grades and learning skills during the interaction and were aware of this knowledge while answering the other questions.

5.2 Children's perceived capabilities of the robot

The presented questions in the 3^{rd} category such as Q1 (overall performance), Q2 (writing capability) and Q3 (improvement) correspond directly to the robot's learning capabilities (see Table2, 3^{rd} category). Despite showing no differences after the first two interactions with the robot, over repeated interactions, the perception of two of these capabilities showed significant differences over time. For example, after the second interaction, the results of the Man-Whitney test showed no significant difference between the learning (mean rank = 9.44) and the non-learning condition (mean rank = 7.56), $p > .05$. However, after the third interaction, children in the learning condition (mean rank = 16.42) gave significantly higher *writing capability scores* to the robot compared to the non-learning condition (mean rank = 7.86), $U = 20.5$, $z = 86.5$, $p = .002$ (see Fig. 5(a)). Furthermore, after the last (4^{th}) interaction, *overall performance* scores and *writing capability scores* both showed significant results. Children in the learning condition (mean rank = 16.58) gave significantly higher *overall performance* scores to the robot compared to the non-learning condition (mean rank = 7.68), $U = 18.5$, $z = -3.366$, $p = .001$. Besides, for the *writing capability*, they (mean rank = 15.65) gave higher scores to the robot compared to the non-learning condition (mean rank = 8.77), $U = 30.50$, $z = -2.67$, $p = .015$ (Fig. 5(b)). The results suggest that the children noticed the difference in the robot's capabilities regarding its writing capability and the overall performance only after the third and final interactions. Regarding the improvement in the robot's handwriting (Q3), we did not find any significant difference



(a) **(b)**

Figure 5: Results of the robot’s learning scores in both conditions: (a) Writing capability scores after the third interaction; (b) Writing capability and overall performance scores after the fourth interaction.

between the two conditions, which is interesting given that in one of the conditions, the robot was not learning. Indeed, after the last interaction, most of the children in both conditions (learning - 92%, non-learning - 75%) provided high scores to the robot’s handwriting improvement (see Table 2, 3rd category). This result suggests that 75% of the children in the non-learning condition, being aware of the robot’s fictitious grades (these grades were not improving in the non-learning condition) believed that the robot’s handwriting was improving (although it was not). We believe this may be due to two reasons: the children actually thought that the robot was improving, because they were teaching it; and the introduction of different sets of letters in each interaction made it more difficult for them to realize that the robot was not improving. Thus, the hypothesis that the children would be able to perceive the difference in the robot’s learning capabilities is only partially proved.

5.3 Children’s self-efficacy towards tutoring

One aim of the *learning-by-teaching* method is to motivate learners to teach, leading to a rewarding feeling as the tutee improves and learns. To assess this feeling, we evaluated children’s perceived self-efficacy towards tutoring. Perceived self-efficacy is defined as “*people’s judgement of their capabilities to execute actions required to attain designated types of performance*" [5]. In the context of our study, it is defined as the children’s self-perception of their capability of tutoring the robot. To measure this, we have included a specific question- Q3 (Table 2, 4th category) that the children answered after each interaction with the robot. The other questions in this category correspond to the children’s current and future likeness towards tutoring the robot (Table2, 4th category). Despite the differences in competencies of the robot, in all questions, we found no significant difference between the two conditions at any point. Besides, after the last (4th) interaction, we found some strong correlations in both conditions. In the learning condition, we found two strong correlations: one between the perceived writing capability of the robot and the children’s fondness towards teaching the robot, $rs(11) = .736$, $p = .004$; and the other between the perceived writing capability of the robot and the grades children gave themselves for tutoring, $rs(11) = .693$, $p = .009$. These results suggest that as the robot progressed, the children’s fondness towards teaching the robot was also increasing, and they were considering themselves better teachers after each interaction. In the non-learning condition, 92% of the children gave high scores to all questions (Table2, 4th category) with strong correlations between the children’s willingness towards future tutoring and the robot’s capability to help

Table 2: Percentage of children who gave high scores to the robot for the questions in the 3rd, 4th and 5th category after the last interaction.

Children’s perceived capabilities of the robot (3rd category)†	L(%)	NL(%)
Q1. How many stars would you like to give for Michael’s overall performance?	85	67
Q2. How many stars would you like to give for Michael’s writing capabilities?	92	58
Q3. Do you think Michael’s handwriting is improving?	92	75
Children’s self-evaluation towards tutoring (4th category)†	**L(%)**	**NL(%)**
Q1. Do you like teaching Michael?	100	92
Q2. Would you like to teach Michael in the future?	92	92
Q3. How good were you as a Michael’s teacher?	92	92
Children’s perceived impression of the robot (5th category) [6]	**L(%)**	**NL(%)**
Q1. What grade would you like to give for Michael’s intelligence?	100	67
Q2. Do you like Michael?	100	92
Q3. Do you find Michael friendly? How much?	100	92

***Learning(L), Non-Learning(NL); †Based on study’s research questions; The % of children (right side) who chose either 4 or 5 points on the Likert scale; The scale values are considered as scores in the analysis; High scores: combined Scores of 4th and 5th point in the 5-point Likert scale (1st-point = Lowest, 5th-point = highest)**

others ($rs(11) = .69$, $p = .01$), the robot learned from the children ($rs(11) = .695$, $p = .01$) and improvement in the robot’s handwriting ($rs(11) = .72$, $p = .01$). This suggests that despite of being aware of the robot’s incapacity to learn, the children still wanted to teach the robot in the future. Our third hypothesis that the children in the learning condition would consider themselves better teachers compared with the non-learning condition was thus not proved.

5.4 Children’s perceived impressions of the robot

The questions in this category are inspired by the Godspeed questionnaire [6] and correspond to children’s perceived intelligence, likability and friendliness towards the robot, see Table 2 (5th category). Regarding all questions, no significant differences were found between the conditions. We observed after the last interaction that more than 92% of the children gave high scores for the fondness and friendliness scale (see Table 2 (5th category, Q2-Q3)). In addition, in the learning condition, we found a correlation between the likability and the overall performance, $rs(13) = .567$, $p = .043$, and in the non-learning condition, between the friendliness and the overall performance, $rs(11) = .606$, $p = .04$. The results suggest that the children’s social behaviour such as fondness towards the robot did not show to be affected by its writing capabilities. Regarding the perceived intelligence, 100% of the children in the learning condition and 67% in the non-learning condition gave high scores (see Table 2, 5th category, Q1). Furthermore, in both conditions, we found multiple strong correlations. The results show that after the

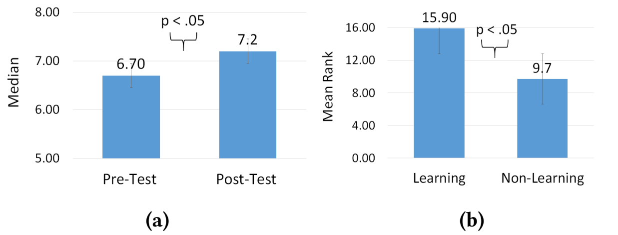


Figure 6: Results of the children's learning scores: (a) Pre- and Post-test scores in the learning condition; (b) Learning gains between the learning and non-learning condition.

last interaction in the learning condition, all the children found the robot intelligent which can be directly related to the robot's learning progression. However, it was more interesting to see the other capabilities of the robot due to which the children perceive the robot to be intelligent. As shown in Table 3, the children in both condition perceived the robot's intelligence because of different perceived abilities. Thus, our third hypothesis is not proved because we did not find a significant difference between the two conditions.

5.5 Children's learning gains

The pre- and post-test consisted of 9 letters (see Section 4.1, pre-test). For every letter, the child was asked to perform two actions on a tablet: (1) *letter selection*- to select the best-shaped letter among 3 presented samples. Five points were given to the correct letter and 0 points for the incorrect letter; and (2) *letter writing*- to write the correct sample of the letter. Two independent coders analysed each letter by comparing it with the correct sample of the letter. The coders rated each letter based on its legibility in the scale of 1 to 5: 1 = unreadable; 2 = difficult to read; 3 = readable with multiple errors such as shakiness, missing stroke; 4 = readable with one error, and 5 = readable with no error. The reliability of the scores using the Cohen's kappa showed good agreement between the two coders, $\kappa = .75, p < .05$. We scaled the above action scores to match the data range and obtained the averaged combined score. Children's post-test scores were relatively compared with their pre-test scores and grouped by the condition. As shown in Fig. 6(a), the results of the Sign test suggest that the children in the learning condition showed significant improvement between pre- (median = 6.7) and post-test scores (median = 7.2), $Z = -2.8$, $p = .004$, effect size $(r) = -0.55$. Out of 13 children, ten elicited an improvement and three remained with no improvement (equal pre-post scores). On contrary, all the children in the non-learning condition showed improvement but no significant difference was observed $p > .05$, effect size $(r) = -0.088$. Moreover, the results of Mann-Whitney test showed significantly higher learning gains (normalised) in the learning condition (mean rank = 15.96) compared to the non-learning condition (mean rank = 9.7), $U = 39.5$, $z = -2.1$, $p = .035$, effect size $(r) = -0.42$ (see Fig. 6(b)). These results validate our fourth hypothesis that the children would benefit more in the learning condition compared with the non-learning condition.

5.6 Children's perceived role of the robot

The questions under this category correspond to the perceived role of the robot by the children regarding its writing capability and the social relationship, and only Q1 (2^{nd} category, Table1) showed a

Table 3: Correlations between perceived intelligence and the other capabilities in the learning & non-learning condition.

Learning condition	Perceived Intelligence
Perceived likeness	rs(13) = .77, p = .00
Perceived overall performance	rs(13) = .77, p = .00
Likeness towards tutoring	rs(13) = .67, p = .01
Likeness towards future tutoring	rs(13) = .67, p = .01
Non-learning condition	**Perceived Intelligence**
Perceived overall performance	rs(11) = .68, p = .02
Future teaching	rs(11) = .67, p = .02
Handwriting Improvement	rs(11) = .65, p = .02

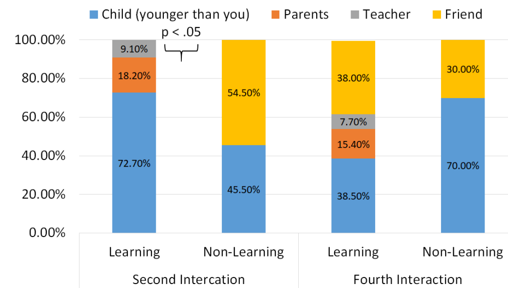


Figure 7: Perceived robot's role (in terms of writing capability) by children after the second and fourth interactions

significant result over time. After the second interaction, we found a statistically significant association between the roles assigned by the children to the robot between the two conditions, $\chi^2(4) = 11.60$, $p < .021$ (Fig 7). In the learning condition, after the second interaction, 72.7% of the children perceived the robot as a *younger child*, and no child perceived the robot as a *friend*; however, after the last interaction, an equal number of the children perceived the robot as a *friend* and a *younger child* (Fig.7). In the non-learning condition, after the second interaction, 54.5% of the children perceived the robot as a *friend*, and 45.5% of the children perceived it as a *younger child*. Nevertheless, after the last interaction, 70% of the children perceived the robot as a *younger child*, while 30% of the children perceived it as a friend (Fig. 7). After the third interaction, no significant association was found between the roles assigned in both conditions, $\chi^2(4) = 3.5$, $p > .05$. These results in the learning condition suggest that the increase in the perception of the robot as a friend may be due to its continuous learning, allowing the children to see it as a peer. In the non-learning condition, the increase in the perception of the robot as a younger child is due to the robot's proof of poor learning.

6 DISCUSSION

The results provided partial support for the first hypothesis; by the end of the last interaction, the children were able to differentiate the two out of three robot's competencies. We did not find support for the second hypothesis that the children would like more the learning robot compared to the non-learning robot. Also, the third hypothesis was not proved as the children's self-efficacy towards tutoring did not show the difference between the conditions. However, the children showed improvement with the learning robot compared to the non-learning robot, hence, providing support to our fourth hypothesis.

Effect of multi-session study on children's perception: *"longitudinal studies are extremely useful to investigate changes in user behaviour and experiences over time." [26]*, In fact, children tend to change their attitude and behaviour towards a social robot over time in multi-session studies [17, 24, 34]. In the current work, conducting a longitudinal study with 4 consecutive sessions was a key factor in finding some of the most relevant outcomes, because some aspects became significant only over time. The findings of children's perception regarding the robot's *writing capability, overall performance* and *role* changed after the two or three interactions which could not be possible to explore in a single session study. However, we also observed an unexpected but a relevant outcome that the children's perceived likability, friendliness and intelligence towards the robot did not get affected by its learning capabilities.

Children's role and self-efficacy: more than 92% of children in both conditions liked tutoring the robot, wanted to teach the robot in the future and, in fact, rated themselves as good teachers (see Table 2, 4^{th} category). The results of children's self-efficacy towards tutoring are consistent with Bierman et al. findings, who concluded that children when acting as a tutor tend to form positive attitudes regarding perceived competence on the task [8]. In our study, the impact of 'being a teacher' influenced all the children towards self-evaluation as teachers. As a result, we did not find the difference between the two conditions. Moreover, Allen et al. also found the similar results in peer tutoring scenario and concluded that the differential performance of tutees did not affect tutor's self-evaluation of their own teaching [2]. The study of Berninger and Allen involved students (human); however, the findings of the current study gives the evidence that the similar effects can also be present in human-robot interaction studies.

Link between children's perceptions of the robot and their learning gains: despite obtaining similar pre-test scores, $p > .05$, effect size $(r) = -0.076$, the children in the learning condition improved significantly compared with the non-learning condition. We believe that the children's role and overall perceptions towards the robot such as perceived robot's capabilities, role, social behaviour and self-efficacy towards tutoring are not only linked [2] but also influenced their learning outcomes. We speculate that there could be at least two variables: *motivation* and *knowledge construction* that may account for more learning in the learning condition. In the learning condition, due to the continuous improvement of the tutee-robot in each consecutive session, the tutor-children might have experienced it as a rewarding feeling because they could see the results of their teaching efforts. Consequently, they put extra efforts and attention on the robots writing which eventually improved their learning. Also, the corrective feedback which they provided during tutoring propelled through motivation encouraged them to heed in writing the letters. In fact, Topping [38] described that the act of tutoring involves the construction of new knowledge and improvement in self-concept . On the contrary, the children with the non-learning robot lacked enough motivation. In spite of their ongoing efforts, in each consecutive interaction, they found the robot is struggling to acquire the handwriting skills, performance and better grades. And therefore, they might not have put extra efforts and attention which consequently improved their handwriting skills but insignificantly.

7 CONCLUSIONS & FUTURE WORK

In the presented work, we describe a social robot that autonomously interacts with children to foster their handwriting skills. Two versions of the robot were developed: one where the robot actually improves its performance and one where the robot does not change its capability of writing. Using an algorithm that incorporates human-inspired movements we could reproduce childlike errors in the writing and improve them over the weeks of interaction. We tested the system by conducting a longitudinal study with 4 sessions in a school and found that the children's writing skills improved with the learning robot compared to the non-learning robot. We explored the children's perception towards the 'learning'vs 'non-learning'capabilities in the robot and the effect of these perceptions on their learning. One should note that only one session is not sufficient to assess the perception of certain skills in a social robot and the longitudinal study was essential to see the evolution both of perception and learning. In general, we found that the system was well accepted by the children for two reasons: first, we were able to conduct multi-session study successfully, and the children really engaged with the robot over a period of 4 weeks; and second, we found that most of the children wanted to teach the robot in the future. The sample size used in the study is relatively small and the results may not apply to broader populations. However, the current results (based on the sample size of 25) gave insight into how children's critical perception can affect their learning; and may motivate other researchers in the future in the design of educational child-robot studies based on two competence models of the robot. We believe the existing system can be exploited by school teachers to improve children's handwriting skills. Shortly, we plan to address some issues we encountered in the present study. We measured the children's perception of self-efficacy towards tutoring but did not measure the perceptions related to their efforts in tutoring and attention to the robot's writing. Therefore, we believe that the efforts and attention paid by the children in both conditions lack evidence. Moreover, the system used offline models of the letter trajectories, and the Nao robot used an animated handwriting movement while writing. We plan to address some of these issues by testing the system with online model trajectories and using a different framework and server connections that would allow us to avoid the use of animations.

ACKNOWLEDGMENTS

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013 and through project AMIGOS (PTDC/EEISII/7174/2014). The first and the third author acknowledge grants with ref. SFRH/BD/51935/2012 and SFRH/BD/51933/2012 funded by the FCT. The second author acknowledges the program "Science without Borders" with process number: 201833/2014-0 under the National Council for Scientific and Technological Development (CNPq) and the State University of Rio Grande do Norte - Brazil. The authors show their gratitude to "Colégio da Fonte" in Porto Salvo, Portugal and its school principal, teachers and students for their participation in this research.

REFERENCES

[1] Vernon L Allen. 1976. Children as Teachers; Theory and Research on Tutoring. (1976).

[2] Vernon L Allen and Robert S Feldman. 1974. Tutor attributions and attitude as a function of tutee performance. *Journal of Applied Social Psychology* 4, 4 (1974), 311–320.

[3] Patrıcia Alves-Oliveira, Pedro Sequeira, and Ana Paiva. 2014. The Role that an Educational Robot Plays. *Proceedings of Robot and Human Interactive Communication (RO-MAN)* (August 2014).

[4] Albert Bandura. 1977. Self-efficacy: toward a unifying theory of behavioral change. *Psychological review* 84, 2 (1977), 191.

[5] Albert Bandura. 1986. Social foundation of thought and action: A social-cognitive view. *Englewood Cliffs* (1986).

[6] Christoph Bartneck, Elizabeth Croft, and Dana Kulic. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics* 1, 1 (2009), 71–81. https://doi.org/10.1007/s12369-008-0001-3

[7] Tanya N. Beran, Alejandro Ramirez-Serrano, Roman Kuzyk, Meghann Fior, and Sarah Nugent. 2011. Understanding How Children Understand Robots: Perceived Animism in Child-robot Interaction. *Int. J. Hum.-Comput. Stud.* 69, 7-8 (July 2011), 539–550. https://doi.org/10.1016/j.ijhcs.2011.04.003

[8] Karen L Bierman and Wyndol Furman. 1981. Effects of role and assignment rationale on attitudes formed during peer tutoring. *Journal of Educational Psychology* 73, 1 (1981), 33.

[9] Gautam Biswas, Krittaya Leelawong, Kadira Belynne, Karun Viswanath, Nancy Vye, Daniel Schwartz, and Joan Davis. 2004. Incorporating self regulated learning techniques into learning by teaching environments. In *The 26th Annual Meeting of the Cognitive Science Society*. 120–125.

[10] Gautam Biswas, Krittaya Leelawong, Daniel Schwartz, Nancy Vye, and The Teachable Agents Group at Vanderbilt. 2005. Learning by teaching: A new agent paradigm for educational software. *Applied Artificial Intelligence* 19, 3-4 (2005), 363–392. http://www.tandfonline.com/doi/abs/10.1080/08839510590910200

[11] John Bitchener. 2008. Evidence in support of written corrective feedback. *Journal of Second Language Writing* 17, 2 (2008), 102 – 118.

[12] David Boud, Ruth Cohen, and Jane Sampson. 2001. Peer learning and assessment. *Peer Learning in Higher Education: Learning from & with Each Other* (2001), 67.

[13] Shruti Chandra, P. A.Oliveira, Séverin Lemaignan, Pedro Sequeira, Ana Paiva, and Pierre Dillenbourg. 2015. Can a child feel responsible for another in the presence of a robot in a collaborative learning activity?. In *Robot and Human Interactive Comm.(RO-MAN), 2015 24th IEEE Int. Symposium on*. IEEE, 167–172.

[14] Shruti Chandra, P. A.Oliveira, Séverin Lemaignan, Pedro Sequeira, Ana Paiva, and Pierre Dillenbourg. 2016. Children's peer assessment and self-disclosure in the presence of an educational robot. In *Robot and Human Interactive Comm.(RO-MAN), 2016 25th IEEE Int. Symposium on*. IEEE, 539–544.

[15] Shruti Chandra, Pierre Dillenbourg, and Ana Paiva. 2017. Classification of Children's Handwriting Errors for the Design of an Educational Co-writer Robotic Peer. In *Proceedings of the 2017 Conference on Interaction Design and Children*. ACM, 215–225.

[16] A. Deshmukh, S. Janarthanam, H. Hastie, M. Y. Lim, R. Aylett, and G. Castellano. 2016. How expressiveness of a robotic tutor is perceived by children in a learning environment. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 423–424. https://doi.org/10.1109/HRI.2016.7451787

[17] Rachel Gockley, Allison Bruce, Jodi Forlizzi, Marek Michalowski, Anne Mundell, Stephanie Rosenthal, Brennan Sellner, Reid Simmons, Kevin Snipes, Alan C Schultz, et al. 2005. Designing robots for long-term social interaction. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*. IEEE, 1338–1343.

[18] Steve Graham, Karen R Harris, Linda Mason, Barbara Fink-Chorzempa, Susan Moran, and Bruce Saddler. 2008. How do primary grade teachers teach handwriting? A national survey. *Reading and Writing* 21, 1-2 (2008), 49–69.

[19] Steve Graham, Naomi Weintraub, and Virginia Berninger. 2001. Which manuscript letters do primary grade children write legibly? *Journal of Educational Psychology* 93, 3 (2001), 488.

[20] John Hattie and Helen Timperley. 2007. The power of feedback. *Review of educational research* 77, 1 (2007), 81–112.

[21] Deanna Hood, Séverin Lemaignan, and Pierre Dillenbourg. 2015. When children teach a robot to write: An autonomous teachable humanoid which uses simulated handwriting. In *Proceedings of the Tenth Annual ACM/IEEE Int. Conf. on Human-Robot Interaction*. ACM, 83–90.

[22] Monica MP Hoy, Mary Y Egan, and Katya P Feder. 2011. A systematic review of interventions to improve handwriting. *Canadian Journal of Occupational Therapy* 78, 1 (2011), 13–25.

[23] Peter H Kahn Jr, Takayuki Kanda, Hiroshi Ishiguro, Nathan G Freier, Rachel L Severson, Brian T Gill, Jolina H Ruckert, and Solace Shen. 2012. "Robovie, you'll have to go into the closet now": Children's social and moral relationships with a humanoid robot. *Developmental psychology* 48, 2 (2012), 303.

[24] Takayuki Kanda, Takayuki Hirano, Daniel Eaton, and Hiroshi Ishiguro. 2004. Interactive Robots As Social Partners and Peer Tutors for Children: A Field Trial. *Human-Computer Interaction* 19, 1 (June 2004), 61–84.

[25] James Kennedy, Paul Baxter, and Tony Belpaeme. 2015. The Robot Who Tried Too Hard: Social Behaviour of a Robot Tutor Can Negatively Affect Child Learning. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (HRI '15)*. ACM, New York, NY, USA, 67–74. https://doi.org/10.1145/2696454.2696457

[26] Iolanda Leite, Carlos Martinho, and Ana Paiva. 2013. Social robots for long-term interaction: a survey. *International Journal of Social Robotics* 5, 2 (2013), 291–308.

[27] Shizuko Matsuzoe and Fumihide Tanaka. 2012. How smartly should robots behave?: Comparative investigation on the learning ability of a care-receiving robot. In *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*. IEEE, 339–344.

[28] Jane Medwell and David Wray. 2008. HandwritingâĂŞa forgotten language skill? *Language and Education* 22, 1 (2008), 34–47. http://www.tandfonline.com/doi/abs/10.2167/le722.0

[29] Sandra Y Okita, Daniel L Schwartz, Takanori Shibata, and Hideyuki Tokuda. 2005. Exploring young children's attributions through entertainment robots. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*. IEEE, 390–395.

[30] Christian O'Reilly and Réjean Plamondon. 2009. Development of a Sigma–Lognormal representation for on-line signatures. *Pattern Recognition* 42, 12 (2009), 3324–3337.

[31] Tiago Ribeiro, Eugenio Di Tullio, Lee J Corrigan, Aidan Jones, Fotios Papadopoulos, Ruth Aylett, Ginevra Castellano, and Ana Paiva. 2014. Developing interactive embodied characters using the thalamus framework: a collaborative approach. In *Intelligent Virtual Agents*. Springer, 364–373.

[32] Tiago Ribeiro, André Pereira, Eugenio Di Tullio, Patrıcia Alves-Oliveira, and Ana Paiva. 2014. From Thalamus to Skene: High-level behaviour planning and managing for mixed-reality characters. In *Proceedings of the IVA 2014 Workshop on Architectures and Standards for IVAs*.

[33] Cynthia A Rohrbeck, Marika D Ginsburg-Block, John W Fantuzzo, and Traci R Miller. 2003. Peer-assisted learning interventions with elementary school students: A meta-analytic review. *Journal of Educational Psychology* 95, 2 (2003), 240.

[34] Tamie Salter, Kerstin Dautenhahn, and R Bockhorst. 2004. Robots moving out of the laboratory-detecting interaction levels and human contact in noisy school environments. In *Robot and Human Interactive Communication, 2004. ROMAN 2004. 13th IEEE International Workshop on*. IEEE, 563–568.

[35] Theodore R Sarbin. 1966. *Role Theory: Concepts and Research*. John Wiley & Sons , Inc. 1966.

[36] Rosemary Sassoon, Ian Nimmo-Smith, and Alan M. Wing. 1986. An analysis of children's penholds. *Advances in Psychology* 37 (1986), 93–106.

[37] Fumihide Tanaka and Shizuko Matsuzoe. 2012. Children teach a care-receiving robot to promote their learning: Field experiments in a classroom for vocabulary learning. *Journal of Human-Robot Interaction* 1, 1 (2012).

[38] Keith J Topping. 1996. The effectiveness of peer tutoring in further and higher education: A typology and review of the literature. *Higher education* 32, 3 (1996), 321–345.

[39] H. Yin, P. Alves-Olivera, F. S. Melo, A. Billard, and A. Paiva. 2016. Synthesizing Robotic Handwriting Motion by Learning from Human Demonstrations. In Proceedings of International Joint Conference on Artificial Intelligence (IJCAI). (2016).

[40] Barry J Zimmerman. 2000. Self-efficacy: An essential motive to learn. *Contemporary educational psychology* 25, 1 (2000), 82–91.