Learning By Collaborative Teaching :
An Engaging Multi-Party CoWriter Activity

Laila El Hamamsy[1], Wafa Johal[1,2], Thibault Asselborn[1], Jauwairia Nasir[1], Pierre Dillenbourg[1]

Abstract— This paper presents the design of a novel and engaging collaborative learning activity for handwriting where a group of participants simultaneously tutor a Nao robot. This activity was intended to take advantage of both collaborative learning and the learning by teaching paradigm to improve children's meta-cognition (perception of their own skills). Multiple engagement probes were integrated into the activity as a first step towards fostering long term interactions. As a lot of research targets social interactions, the goal here was to determine whether an engagement strategy focused on the task could be as, or more efficient than one focused on social interactions and participants' introspection. To that effect, two engagement strategies were implemented. They differed in content but used the same multi-modal design in order to increase participants' meta-cognitive reflection, once on the task and performances, and once on participants' enjoyment and emotions. Both strategies were compared to a baseline by probing and assessing engagement at the individual and group level, along the behavioural, emotional and cognitive dimensions, in a between subject experiment with 12 groups of children. The experiments showed that the collaborative task pushed the children to adapt their manner of writing to the group, even though the adopted solution was not always correct. Furthermore, there was no significant difference between the strategies in terms of behaviour on task (behavioural engagement), satisfaction (emotional engagement) or performance (cognitive engagement) as the group dynamics had a stronger impact on the outcome of the collaborative teaching task. Therefore, the task and social engagement strategies can be considered as efficient in the context of collaboration.

I. INTRODUCTION

Despite the increased use of tablets and computers in schools, handwriting remains an important skill that must be acquired at a young age. As children are faced with increased cognitive loads over the years, those that experience difficulties in handwriting tasks see their overall progress in schools impeded [1]. That is why taking advantage of the increased interest to integrate technology in schools in order to develop tools that may help children master such skills is important. That is where the force of the CoWriter project [2], [3], [4] lies as it aims to help children acquire these skills that are generally tedious to learn. The idea was to invert the learning dynamics and have the child tutor a robot learner. This had many benefits of which engaging the child for extended periods of time [3], increasing self-esteem, making them feel responsible for the robot's success through the Protégé Effect [2], as well as having them reflect meta-cognitively on the robot's performance and by extension their own.

[1]CHILI Laboratory, EPFL, Switzerland
[2]Biorobotics Laboratory (BioRob), EPFL, Switzerland

Because handwriting is an individual task, up until now the children that have partaken in the CoWriter activity have interacted with the Nao robot in a dyadic setting. However, this is not scalable to classroom settings. Ideally multiple children would interact simultaneously with the robot. Therefore, developing an activity that permits group level interactions on a traditionally individual task would be a beneficial contribution which can be seen as a first step towards orchestrating classroom level activities. That is why the current paper focuses on the development and validation of a multi-party handwriting activity. This collaborative teaching activity retains the learning by teaching principle all the while taking advantage of the benefits of collaborative learning. Hence the term "learning by collaborative teaching". Furthermore, different engagement probes were integrated to the activity in order to sustain the learners' involvement for extended periods of time. Two main orientations were investigated based on the literature : a task oriented and a social oriented engagement strategy which are described in section III-C. As such, the goal was to determine how the implemented engagement strategies affected the groups' *behaviour on task* (behavioural engagement), *performance* (cognitive engagement) and *satisfaction* (emotional engagement).

II. RELATED WORK

Whereas many researchers are interested in human robot interactions, few target multi-party scenarios. Furthermore, in HRI, the focus is often split between engagement assessment and task performance assessment. Engagement is often evaluated at the social level for the different individuals separately, where many studies take place in dialogue settings [5], [6] focusing on intention detection [7]. Few consider engagement at the group level [8], [7], or with respect to the task [9], and even fewer targeting children [10], [11]. When considering learning tasks with social robots (as opposed to educational robotic kits), many play the role of a tutor in dyadic settings, contrary to the role of a facilitator in a context with more participants, for example when soft skills are the target [10]. In addition, the learning activity may not necessarily be collaborative and require the same level of engagement from all participants to achieve the task. Provided the vast body of literature on HRI, the focus was put on reviewing the literature related to 1) educational contexts with social robots, notably a tutoring robot which must probe the engagement of participants (Section II-A) 2) collaborative learning which requires the same level of engagement of all participants to succeed in a task (Section II-B) and finally 3) engagement assessment which can extend to both the individual and group level (Section II-C).

A. Social Robots in Tutoring Contexts

As the present study involves the interaction of participants with a robot learner, one must consider the role that social robots play in tutoring contexts. Physical robots are preferred to virtual agents when the context requires interacting with the physical world such as handwriting. However, it is important to have a robot that is believable and sociable in order to engage the child cognitively for extended periods of time [4]. When considering the sociability of the robot, "an aspect that seems to play an important role in sustaining long-term engagement with children is the complexity of the robot's social behaviour, and the amount of diverse behaviours that the robot can display over the interactions" [12]. Many studies have done so by incorporating more diverse dialogues, motions and non verbal cues [13], [14]. However, repetitive and predictive behaviour were shown to have a negative impact on interest and motivation due to the habituation effect [12]. That is why one cannot neglect the importance of adaptivity, novelty, personalisation and unexpected behaviours. This is why researchers have looked into the benefits they can offer in terms of engagement and learning gains [15], [12], [16], [17], [18].

B. Collaborative Learning

The second point to be addressed and which constitutes the core of the activity is the integration of collaborative learning in a learning task and the benefits it may have. It is important to start by distinguishing collaboration from cooperation as it does not rely on the "division of labour amongst the participants" but on having them "mutually engage in a coordinated effort to solve the problem" [19]. Collaborative learning would then be "a situation in which two or more participants learn or attempt to learn something together" [20]. Certain factors must be considered at the individual level in order to have successful collaboration. For instance, the participants must "have skills in the task, be able to reflect on their own performance, [and] be able to determine where the learner went wrong". Participants in the present study must teach the robot how to write together in an interaction that could typically last around one hour and be repeated over successive sessions to help improve their skills. That is why the type of learning considered is face to face, synchronous and relies on collaborative work and joint effort [19]. Therefore, one hypothesis considered was that collaborative learning, or in this case, learning by collaborative teaching, could have a positive effect on both engagement and learning through the reassessment of what was done with respect to the group and feeling responsible for the group's success. This is analogous to the Protégé effect when considering learning by teaching.

C. Engagement For Long Term Interactions

Finally, it was important to consider the participants' engagement since the general postulate is that high engagement leads to more beneficial interactions in terms of learning gain and performance [16]. This is because engagement is often considered to be a reflection of ones motivation which has been correlated to higher practice, which can increase learning gains [21]. However, engagement in itself is a complex construct that is difficult to assess as it depends highly on what it means to be engaged in the considered setting. This is due to the fact that expected behaviours and their interpretation vary depending on what we are doing, be it, for example, a handwriting task or a conversational setting. That is why it is important to define engagement in the context of the study and design the interactions appropriately in order to maximise participants' engagement with respect to the facet considered [22], without negatively impacting the interaction and learning gain [13].

The most prevalent definition of engagement is that of Sidner et al. as "the process by which two (or more) participants establish, maintain and end their perceived connection". However, a more specific definition is useful when participants partake in both a social and task context. Social here refers to the social interactions with the robot and other participants, and task to the resolution of the activity which is often linked to a personal "need for achievement" and being in a state of flow [23], [24]. The definition of engagement selected is therefore "all effort, task or social, put into achieving the goal" [8] (in this case teaching a robot to write collectively). Here, engagement is considered to be context dependent [25], [9], [8] and comprised of *behavioural* (what participants do), *emotional* (how participants feel) and *cognitive* (the amount of mental effort put into the task) engagement [22]. Certain studies have looked into probing and assessing engagement socially versus with respect to the task [16], [9], [23] but one cannot dissociate or eliminate the effect that a purely social or purely task oriented probe has on the other dimension. This is even more prominent in a collaborative task which can only be resolved through efficient interactions with other members of the group.

That is why the present study designed an engaging activity that *probed and assessed engagement through the behavioural, emotional and cognitive dimensions of engagement*. The content of said probes then varied with respect to the task and socially oriented engagement strategies, as described in section III-C. To *probe engagement*, different factors were taken into account. It was important to ensure that the task difficulty was adapted for cognitive engagement. This follows the principles of Flow Theory [24] that stipulate that an adapted difficulty is necessary to be immersed in the task and in a optimal mental state for learning [23]. Feedback was also integrated in order to obtain higher meta-cognitive reflection on the learners' part. For emotional engagement, avoiding boredom and repetition was considered by integrating variety and personalisation, for example, in the interaction [12]. In doing so, behavioural engagement tends to follow as participants are more likely to invest in the task, especially in the context of an interactive and collaborative activity where they must work together to solve the task. Similarly, to *assess engagement*, measurements along the behavioural, emotional and cognitive dimensions were considered for both strategies. *Behavioural* engagement

is generally assessed through gaze features, posture and motion. However, it can also be assessed through the task behaviour, which is to say, the actions done throughout the activity (e.g. number of words written in a session). *Emotional* engagement on the other hand is often assessed through audio using prosody, visual emotion recognition or a combination of both. Unfortunately, non commercial emotion recognition modules tend to be highly context dependent and suffer from limited performances[1]. *Cognitive* engagement is usually assessed using performance measures as the goal is to maximise learning. Nonetheless, certain studies are looking into the potential of using physiological sensors. These however require significant setups and are complex to decode and analyse which is why they are not frequently employed. Given the complexity of measuring emotional and cognitive engagement, one of the most common methods used is self-assessment questionnaires administered at the end of the activity. Although this runs the risk of being biased by subjectivity as well as primacy and recency memory effects, they are still considered by certain researchers as "the most valid measure of cognitive and emotional engagement [...] as [they] focus heavily on student's perceptions of their experience" [22] through questions related to challenge, concentration, interest, enjoyability, satisfaction and so forth.

D. Summary

None of the studies that were reviewed considered all of the following points simultaneously : 1) engagement probing and assessment at group and individual levels, both considering the same modalities of engagement; 2) in a multi party HRI settings with a social robot; 3) in a tutoring context; 4) with children; and 5) with the evaluation of performance for a hard skill. Therefore, the contributions of the present study can be summarised as: 1) A collaborative handwriting task which pushes for reassessment of performances; 2) A multi-party HRI setting in which children simultaneously tutor a robot; 3) The theoretical framework around the engagement probing and assessment, at individual and group levels, considering both the task and social interactions.

III. DESIGNING THE ENGAGING COLLABORATIVE HANDWRITING ACTIVITY

In this section, the core of the design of the collaborative teaching task is presented and is followed by a brief description of one round of the activity. The engagement probes are then described with respect to the engagement strategies.

A. Collaboration In A Generally Individual Task

Given the current state of the art, the objective was to integrate collaboration into the handwriting task all the while taking advantage of the learning by teaching paradigm. To fully benefit from collaborative learning with all participants at equal footing, the idea was to have them jointly play the role of the teacher with the robot as the sole learner. It was therefore important to design the activity so that

the three (or more) teacher-children may write at the same time. The combination of what they wrote would then serve as the demonstration for the robot. The quality of the combination would depend on both the quality of the individual handwritten letters and their similarity.

The combination of the letters produced by the teacher-children (demonstration letters) is obtained through a stroke[2] by stroke process using a weighted sum[3]. The algorithm considers that the demonstrations provided for a given letter could be made up of a different number of strokes. This is necessary as letters can be made up of multiple strokes (e.g. t, i, j) but they can also be written in different manners leading to a different number of strokes for the same letter (e.g. two strokes for a print f versus one stroke for a cursive $\int$. Examples obtained with letters exhibiting such constraints can be seen in Fig. 1. The selected algorithm penalises differences in terms of shape, direction and order with which the strokes are drawn, as can be seen in Fig. 2. This is important because teachers do not only expect kids to produce the correct letter shape but also to follow the same direction and same order of strokes. One difficulty in the activity is therefore to write the letters in the same way. Difficult letters are then those that generally tend to be written in different ways as opposed to easy letters that tend to be written similarly. Letter difficulty was therefore assessed using a database of handwritten letters by children of the considered age using a dissimilarity metric for multivariate time series [27], in order to estimate the task difficulty with respect to the word[3].

To aid the participants in the collaborative task and help them reflect on their own performances as well as to facilitate the identification of differences between the letters drawn, each stroke drawn started and ended in a different colour as illustrated in Fig. 2. Therefore, a discrepancy in direction can be identified by noting that the gradient is in the other direction (letter a in Fig. 2). A discrepancy in the order of lines drawn (letter t in Fig. 2) can be seen by remarking that the overall colour of the strokes being compared are different. Using such a visualisation helped participants review the dynamics of what they wrote with respect to one another and helped identify what needed to be changed. Nonetheless, the collaborative activity remains challenging as it relies on the coordination and success of all the participants rather than having "low achievers progressively become passive when collaborating with high achievers" [19].

B. The Engaging Collaborative Teaching Activity Pipeline

Once the manner with which collaboration would occur was defined, the original CoWriter activity pipeline was adapted to integrate both collaboration and a set of engaging features. After the robot presents itself and the activity, one round of the activity goes as follows :

1) The selected word appears on the children's tablet, considering an adapted difficulty to avoid disengagement.

---

[1]Visual Deep Learning solutions struggle to reach 80% accuracy, and traditional methods 70% accuracy, both in controlled environments [26]

[2]A stroke here is defined as all the points that make up a line between the moment the pen is placed upon the tablet and then subsequently removed.

[3]Letter merging and letter difficulty computation can be found at `https://github.com/chili-epfl/allograph`
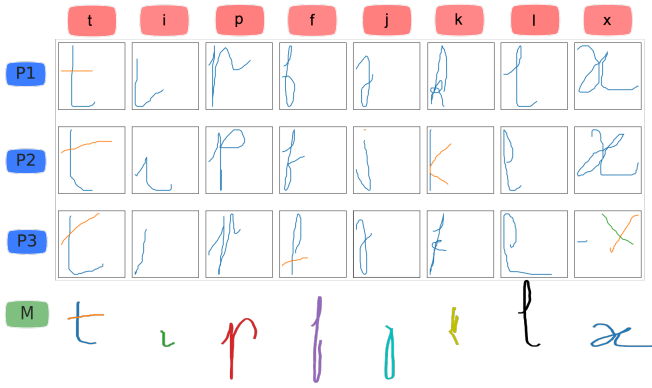
Fig. 1: Letter Merging Example. A column represents a letter, a row P_i a demonstration by a participant and row M the combination of the individual demonstrations.
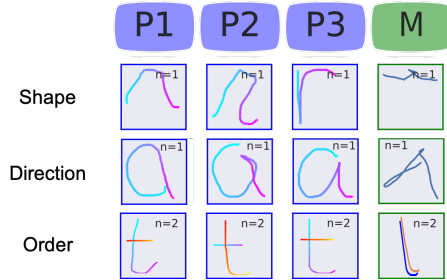

Fig. 2: Differences Penalised By The Algorithm. Examples From The Experiment. A column P_i shows letters drawn by a participant, and column M the combined demonstration.

2) The robot writes the word for the first time.
3) The robot asks for feedback to have participants meta-cognitively reflect on what is going on. An added benefit to this is that participants feel like their opinion matters which could in turn lead to better engagement on all fronts : emotional because they feel like their opinion matters, cognitive because they have to reflect on the question, and behavioural because they are required to respond in order to pursue the activity.
4) The participants write the word on their individual tablets. This step was accompanied by group negotiations to decide how to write (e.g. print versus cursive).
5) The combined demonstration is shown and the group is asked whether they are satisfied. This step is instrumental in the collaborative activity as they are shown their combined work and must unanimously decide whether they want to start over. To move forward, they all had to agree to keep the demonstration which sometimes lead to disagreements and even conflicts. If that was not the case, they were allowed to re-write the word without any limitation in number of attempts.
6) When the combined demonstration is approved, the robot learns from it and demonstrates what it learnt.
7) The robot then asks for feedback as in step 3.

As one of the objectives was to design a task which was engaging, the robotic intervention was designed to be sociable. Both the verbal and non verbal cues (gestures[4], breathing motions, linguistic fillers and exhibiting certain

basic emotions) were therefore adapted with an increase in the quantity and expressivity when in the engaging scenarios. This was done all the while making sure to limit repetitiveness by integrating novelty through random selection of the cues at each iteration from a predefined pool containing multiple variants of each. Note that the participants were also regularly addressed by their first names in order to solicit them and peak their interest.

C. Task and Social Engagement Strategies

All the engagement probes that were implemented were designed to vary according to the considered engagement strategy. Here, two engagement strategies, that probe along the different engagement components, are taken into account and compared to a baseline without any particular form of engagement (i.e. which contains the strict minimum elements necessary to conduct the activity). The first strategy is the task engagement strategy which focuses on the task at hand and having children meta-cognitively reflect on the robot's performance (and by extension on their own). The second is the social engagement strategy which focuses on their enjoyment and having them meta-cognitively reflect on their emotions with respect to the activity and the group interactions. The strategies are probed and assessed using the same modalities but with varying probe content, as can be seen in Table I which compares both to the baseline [5].

IV. EXPERIMENTAL EVALUATION

The objective of the experiment was to assess the collaborative activity and evaluate the effect of the task and social engagement strategies on the task behaviour (behavioural engagement), satisfaction (emotional engagement) and performance (cognitive engagement). To that effect, the collaborative activity was conducted with 35 seven-eight year old children (22 boys and 13 girls) from a private school in Switzerland, in a between subject experiment, that was conducted in January 2019 over the course of four days. The children were randomly put into groups of three with the exception of one pair. Each team then participated in two 40 minute sessions that were two days apart. During a session, each child was provided a tablet, and the combined demonstration (used to teach the Nao robot) was shown on the tablet placed in front of them, as can be seen in Fig. 3.
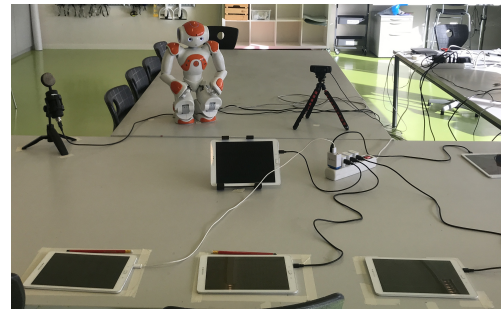

Fig. 3: Experimental setup for three participants

TABLE I: Engagement Probing and Measurements With Respect to The Engagement Strategy.

Engagement dimension	Probe Design	Probe Example Per Engagement Strategy			Measurement
		Baseline	Social	Task	
Behavioural	Collaborative Nature of The Activity	Constant in all strategies			Task behaviour (see Fig. 4)
Emotional	Dialogue (verbal and non verbal cues, of which the gestures and linguistic fillers) which vary in content with respect to the engagement strategy	Probe Description			Questions on satisfaction from the self assessment questionnaire administered at the end of the activity.
		Minimal dialogue and motions to carry out the activity	Dialogue focused on the participant's enjoyment	Dialogue focused on the importance of the task and performing well	
		Probe Example			
		Robot says : "There you go. I'm done."	Robot says : "There you go. I'm done. I am really enjoying writing with all of you."	Robot says : "There you go. I'm done. I am sure you can show me how to improve."	
Cognitive	Feedback	Probe Description			Group handwriting performance
		Asking participants whether they are ready	Meta-cognitively reflect about their emotions (questions about the interactions with the robot, group, enjoyment, emotions...)	Meta-cognitively reflect about the robot's performance and by extension their own	
		Probe Example			
		Robot asks : "Are you ready?"	Robot asks : "Do you like what we are doing?"	Robot asks: "Do you think I wrote "word" better?"	
	Task Difficulty	Progressively increasing in all scenarios (see Table II for selected words and corresponding difficulty)			

Each session started off by having the children copy a pangram to help acclimate to the Android tablets and tactile pens provided. The children then participated in the collaborative activity pipeline described in section III-B with a pre-defined set of words. These words were selected to have an incremental increase in difficulty within a session, all the while being constant between sessions, using the metric described in section III-A. The list of words selected for the two sessions and their corresponding difficulty are presented in Table II. This was important to avoid having participants disengage due to boredom and repetition.

TABLE II: Selected Words per Session and Difficulty Metric Score. Word difficulty = Σ letter difficulties. As all letters must be written well simultaneously, a longer word is harder.

Session 1 Words	Session 2 Words	Session 1 Word Difficulty	Session 1 - Session 2 Word Difficulty
1. cry	1. day	1.52	1.96×10^{-4}
2. wet	2. pig	1.72	1.30×10^{-4}
3. last	3. salt	1.96	0
4. aunt	4. tuna	2.05	0
5. shoe	5. hose	2.28	0
6. lawn	6. jump	2.47	1.74×10^{-4}

At the end, a short self-assessment questionnaire, adapted from exisiting literature [23], which employed a 5-point Likert scale, to gain insight into emotional and cognitive engagement was administered. As this suffers from certain limitations presented in Section II-C, these were complemented by observations made by the 6 experimenters present in pairs during the session. Cognitive engagement was also assessed based on the quality of the groups' combined demonstrations. Behavioural engagement on the other hand was evaluated through features extracted from the activity logs themselves, of which a few are illustrated in Fig. 4. The first session each group participated in was non engaging (baseline) in order to reduce the novelty effect and provide a baseline for the analysis. Over the course of the second session, half of the groups participated in the task engaging activity and the other half in the socially engaging activity, both with a more active and sociable robot. Non parametric tests were then used to assess the effects of the two strategies.

V. RESULTS AND DISCUSSION

The children had an overall positive experience in both sessions as noted in the responses to the self-assessment, with overall satisfaction scores of 4.3±1.1, 4.3±1.3 and 4.3±1.2 on the 5 point Likert scale (1 being the lowest and 5 the highest) in the baseline, social and task modes respectively. Many were invested and enjoyed the robot's presence. When fully engaged, some provided oral feedback and encouragements to the robot and even answered its questions out loud. Based on the experimenters' observations, many seemed to enjoy the increased sociability, attributed emotions to the robot, imitated some of its gestures and were even disappointed to learn that they would not participate again in the activity. All of this seems to indicate that the activity was well received.

A. Validation of Collaboration in a Handwriting Task

The groups did overall 6.0±1.7 words, each taking 04′54" ± 02′21", with 12.7±4.0 merges over all words during the 25 minutes dedicated towards the task itself in the session. They would, on average, attempt each word 2.3±0.9 times, taking around 02′17" ± 49" per attempt. These attempts included the interactions with the robot, the discussions between participants to determine with which manner to proceed and comparisons between the different letters, as observed by the experimenters present. This did not vary significantly between the different sessions and groups.

Although this would seem to point towards the effectiveness of the activity, it was important to determine whether integrating collaboration in the handwriting task effectively pushed participants to adapt their manner of writing to the group dynamics. That is why the evolution of what was written by the participants of a group was evaluated over all attempts using representations such as the one in Fig. 5. After the groups had grasped the underlying principle of the activity, and when provided a new word, they would at each round decide how to write the letters, as noted by the experimenters. This meant writing in either cursive or script, lower case or upper case, and direction and number of strokes when necessary, based on how they knew how to write the
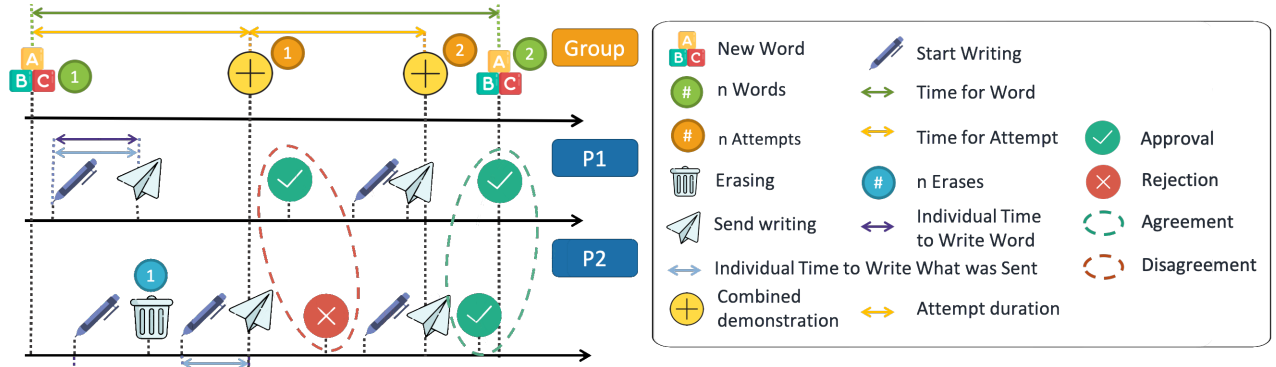
Fig. 4: Activity timeline example for two participants P_i with group level events illustrated in the first row. One can see a subset of the features used for the group and individual task behaviour assessment irrespective of the quality of the letters.

letters and what they thought would be the simplest solution. When they obtained the combined demonstration and if they were dis-satisfied with the results and wanted to try again, they would start over until they converged to a satisfactory output. These observations confirm that the collaborative task did push participants to meta-cognitively reflect on their performances and adapt to the group. This contrasts with what was noted during several dyadic CoWriter experiments where participants tended to write the same way until the robot perfectly imitated them. Furthermore, the original hypothesis was that a collective decision could lead to adopting the correct manner of writing a letter, as it was unlikely that two or more participants would write incorrectly and in the same manner. However, the outputs did not always correspond to what is taught in textbooks, notably in terms of direction (e.g. letter r Fig. 5). Groups converged to a shape score over 50% (see Section V-C for how scores are calculated) in 82% of approved cases, but for certain letters (a, d, i, p, r) would sometimes settle on an incorrect direction. In the cases with a shape score over 50%, and for the letter p, they converged to an incorrect direction 100% and 57% of the time for the task and social groups respectively, and 75% overall (see Table III). This highlights one of the limitations of the collaborative activity, although this concerns a limited number of letters and occurs over all letters only 4% of the time, in this activity.
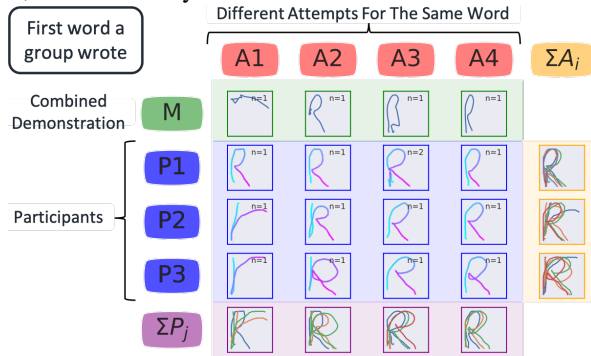

Fig. 5: Evolution of How the Letters Were Written by the Individual Participants When Collaborating.
Letter r written in four attempts the first time a group participated in the activity, i.e. without prior experience with the letter merging. Successive corrections were made to write similarly.

TABLE III: Percentage of letters that were approved by the group with a shape score over 50% (black), averaged over all annotators, and percentage of which had an error in direction (green). Letters highlighted in red (a, d, i, p, r) obtained high shape scores with errors in direction.

	Baseline Sessions with		Engaging Session		Overall
	Task Groups	Social Groups	Task	Social	
a	55%, 0%	100%, 0%	48%, 20%	86%, 0%	69%, 5%
c	88%, 0%	100%, 0%	0%, 0%	0%, 0%	95%, 0%
d	0%, 0%	0%, 0%	56%, 20%	90%,0%	74%, 7%
e	100%, 0%	100%, 0%	100%, 0%	100%, 0%	100%, 0%
g	0%, 0%	0%, 0%	67%, 20%	100%, 0%	82%, 0%
h	0%, 0%	0%, 0%	100%, 0%	0%, 0%	50%, 0%
i	0%, 0%	0%, 0%	44%, 0%	75%, 17%	59%, 10%
l	100%, 0%	100%, 0%	100%, 0%	100%, 0%	100%, 0%
n	50%, 0%	100%, 0%	100%, 0%	100%, 0%	92%, 0%
o	0%, 0%	0%, 0%	100%, 0%	100%, 0%	100%, 0%
p	0%, 0%	0%, 0%	56%, 100%	88%, 57%	71%, 75%
r	63%, 20%	58%, 0%	0%, 0%	0%, 0%	60%, 8%
s	33%, 0%	100%, 0%	44%, 0%	100%, 0%	67%, 0%
t	86%, 0%	85%, 0%	75%, 0%	82%, 0%	83%, 0%
u	50%, 0%	50%, 0%	75%, 0%	100%, 0%	75%, 0%
w	100%, 0%	100%, 0%	0%, 0%	0%, 0%	100%, 0%
y	100%, 0%	100%, 0%	100%, 0%	100%, 0%	100%, 0%
All	80%, 1%	91%, 0%	66%, 11%	90%, 5%	82%, 4%

B. Task Behaviour Assessment

No statistically significant difference was found with respect to the conditions and the metrics described in Fig. 4, in part due to the diversity of strategies employed to perform the task. Despite the decision regarding how to write the letters at first (print or cursive, lowercase or uppercase), the manner with which the attempt itself was conducted varied. Certain groups had a "leader" that the others would imitate whilst others wrote and compared only after obtaining the combined demonstration. Others compared before sending and would erase and write again until they were satisfied and then send to obtain the combined demonstration. Furthermore, if one participant was disruptive, this tended to push the others towards similar behaviours. Another phenomenon which was observed and was related to the nature of collaboration was when the participants did not have the same goals. This would sometimes lead to disagreements and conflicts which were not always beneficial to the interactions and learning process. As such, the manner with which the groups interacted and collaborated as well as the group dynamics

had a larger impact on how the activity went than the engagement strategies. Nonetheless, certain behaviours were correlated with the the self-assessment responses. The main ones are illustrated in Table IV. Generally, when participants spent a lot of time on a given word (high attempt duration or high number of attempts), this negatively impacted the satisfaction, perception of the activity and overall interactions.

TABLE IV: Significant Correlations (Spearmann R) Between Behaviours on Task and the Self Assessment. Only a few are shown due to the amount of features considered[6].

Task/Social Groups	Task Behaviours		
Self Assessment Responses[7]	Attempt duration	Time To Write What Was Sent	Number of attempts
Enjoy teaching		0.55*	−0.60*
Ease of teaching	−0.84***	0.54*	−0.73***, −0.62**
Helping the Group	−0.48*	0.71**	−0.70**
Concentration	−0.50*, −0.6**	0.77***	−0.70**
Wrote well		−0.55*	−0.51*
Satisfaction	−0.81***	0.79***	
ΔSatisfaction Between Sessions			−0.74***, −0.79***

C. Group Performance Analysis

Group performance was assessed on a scale of 0 to 1 through manual annotations by five annotators as it is considered to be "a more valid assessment [of the effects of collaborative learning]" than individual task performance measures [20]. The evaluation was based on the three criteria that had to be respected to obtain satisfactory results : shape, direction and order with which the strokes are drawn. Overall performance in Fig. 6 shows the interaction between the sessions and the experimental conditions as there is a decrease in performance which is higher in the task condition (unpaired Wilcoxon test, $p = 0.028$). As the words were selected to be of equivalent difficulty and as the children had already interacted with the system, the reason behind the decrease in performance was investigated by looking into the individual letter scores with respect to the session and the condition the groups were placed in. It seemed there was significant difficulty letter d from day and p from pig (first words of the second session), as well as a from $last/salt$ and $aunt/tuna$, the 3rd and 4th words of both sessions. This was notably due to issues in terms of direction. Individual issues in direction lead to an overall bad output shape (e.g. a and d) and when at the group level, lead to a good shape with bad direction (e.g. p). For a recap on group level shape and direction scores, refer to Table III. This lead to a high number of attempts and therefore frustration for many groups which was worse for the groups of the task condition that had struggled in the first session and were already scoring lower. One reason could be that the social condition was more enjoyable and motivating than the task condition or that the task condition put more pressure on the children which negatively impacted performances. Unfortunately, no definite conclusion can be drawn without considering the participants' engagement throughout the activity with respect to the experimental conditions.
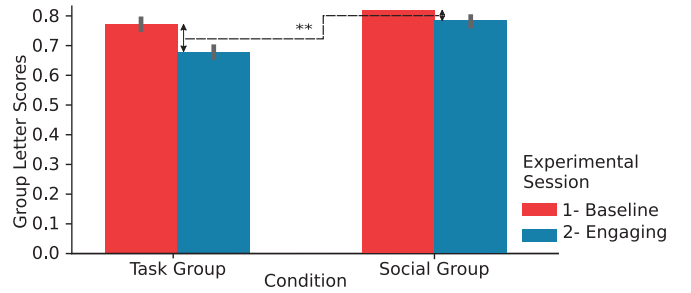

Fig. 6: Group Performance Evolution Over Sessions

VI. CONCLUSIONS AND FUTURE WORK

The collaborative task was well received by the children as well as the sociability of the robot, as noted in the questionnaire responses. No statistically significant difference was noted between the task and social conditions with respect to the groups' behaviours, overall performances and satisfaction. This is highly due to the diversity of behaviours observed as well as the impact of the group dynamics that were established by the collaborative nature of the task, and which were stronger than the implemented engagement conditions. This would seem to indicate that the task and social engagement strategies are as efficient in maintaining learners interest over the course of multiple interactions. Nonetheless, conducting the experiment over a larger number of sessions and groups would be interesting.

Two limitations exist in the current activity design. The first is that groups that succeed do not necessarily adopt the correct manner of writing certain letters in terms of direction (notably letter p) although they converge to a shape score over 50% in 82% of approved cases, with only 4% of those containing errors in direction. A solution must therefore be found for future iterations in order to push participants to adopt the correct direction all the while making sure the children still feel like they are the teachers. The second limitation is that the groups that conduct a large number of attempts for a given word tend to get frustrated and rapidly disengage even though they control the progress of the activity. In such cases, having an online performance assessment model trained on a dataset of handwritten cursive, print, uppercase and lowercase letters would be interesting to be able to assess online both group and individual performances. This would also help a teacher aid the groups in cases of frustration, provide targeted indications and guide the children to avoid disengagement or even alter the word to adapt the difficulty.

Furthermore, as we are mentioning frustration and disengagement, it would be interesting to see how the behavioural engagement evolved over the course of the activity. Currently only the behaviour on task was analysed, not considering the social interactions which are key to the collaborative activity. This can be done through annotation of individual and group engagement at the level of the task and the social interactions on the video streams acquired. A similar process can be

---

[6]For all correlations, refer to https://github.com/LailaHms/CollaborativeCowriter_Correlations

[7]For the full list of questions : https://github.com/LailaHms/Self-Assessment

considered for emotional engagement which was evaluated via the self-assessment at the end of the experiment. This would then provide an estimation of their affective states and valence throughout the activity, and as such, complement the former. All of this would provide insight into how the task behaviour, social behaviour, emotional engagement and performance relate to one another. Another utility to having such types of annotations would be to design a data driven online engagement prediction model for the task. Thus, having both online performance and engagement prediction models could be benefitial by integrating adaptation and triggering "correct behaviours and interventions to build and maintain engagement" [16] as it was noted that "performance improved when using an adaptive agent" [16].

Generally speaking, adaptivity could be integrated in the current framework with respect to: 1) the robot's behaviour by making it socially adaptive (e.g. by incorporating features from the domain of social signal processing [28] such as those pertaining to gaze, mutual gaze and joint attention) all the while considering the importance of being well timed and adapted to the context as social interactions are generally more complex than simply integrating more diverse dialogue, motions, verbal cues, emotions etc... ; 2) intelligent switching between the different engagement modes, for example, to have a balance between the social and task engagement and take advantage of both depending on the context [9] ; 3) task difficulty in order to ensure that the participants remain in a state of flow and avoid disengagement; 4) integrating smart task switching to have individual participants practice specific issues ; 5) providing insight to a teacher to help monitor and regulate the activity so they may act as a facilitator instead of a tutor [20] and intervene appropriately.

REFERENCES

[1] T. Asselborn, T. Gargot, Kidziński, W. Johal, D. Cohen, C. Jolly, and P. Dillenbourg, "Automated human-level diagnosis of dysgraphia using a consumer tablet," *npj Digital Medicine*, vol. 1, p. 42, 2018.

[2] D. Hood, S. Lemaignan, and P. Dillenbourg, "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*, HRI '15, 2015.

[3] A. Jacq, S. Lemaignan, F. Garcia, P. Dillenbourg, and A. Paiva, "Building successful long child-robot interactions in a learning context," in *HRI'16*, (Christchurch), pp. 239–246, IEEE, Mar. 2016.

[4] S. Lemaignan, A. Jacq, D. Hood, F. Garcia, A. Paiva, and P. Dillenbourg, "Learning by Teaching a Robot: The Case of Handwriting," *IEEE Robotics & Automation Magazine*, vol. 23, pp. 56–66, June 2016.

[5] S. M. Anzalone, S. Boucenna, S. Ivaldi, and M. Chetouani, "Evaluating the Engagement with Social Robots," *Int J of Soc Robotics*, vol. 7, pp. 465–478, Aug. 2015.

[6] H. Salam, O. Celiktutan, I. Hupont, H. Gunes, and M. Chetouani, "Fully Automatic Analysis of Engagement and Its Relationship to Personality in Human-Robot Interactions," *IEEE Access*, vol. 5, pp. 705–721, 2017.

[7] M. E. Foster, A. Gaschler, and M. Giuliani, "Automatically Classifying User Engagement for Dynamic Multi-party Human–Robot Interaction," *Int J of Soc Robotics*, vol. 9, pp. 659–674, Nov. 2017.

[8] H. Salam and M. Chetouani, "A multi-level context-based modeling of engagement in Human-Robot Interaction," in *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, (Ljubljana), pp. 1–6, IEEE, May 2015.

[9] L. J. Corrigan, C. E. Peters, and G. Castellano, "Social-Task Engagement : Striking a Balance between the Robot and the Task," in *Embodied Communication Goals Intentions Workshop ICSR*, vol. 13, pp. 1–7, 2013.

[10] B. Huskens, A. Palmen, M. Van der Werff, T. Lourens, and E. Barakova, "Improving Collaborative Play Between Children with Autism Spectrum Disorders and Their Siblings: The Effectiveness of a Robot-Mediated Intervention Based on Lego® Therapy," *J Autism Dev Disord*, vol. 45, pp. 3746–3755, Nov. 2015.

[11] J. Kim, K. Truong, and V. Evers, "Automatic detection of children's engagement using non-verbal features and ordinal learning," in *Workshop on Child Computer Interaction*, pp. 29–34, Sept. 2016.

[12] I. Leite, C. Martinho, and A. Paiva, "Social robots for long-term interaction: A survey," *International Journal of Social Robotics*, 2013.

[13] J. Kennedy, P. Baxter, and T. Belpaeme, "The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning," in *HRI '15*, (New York, NY, USA), pp. 67–74, ACM, 2015.

[14] T. Asselborn, W. Johal, and P. Dillenbourg, "Keep on moving! exploring anthropomorphic effects of motion during idle moments," in *27th edition of IEEE RO-MAN, 2017*.

[15] D. Leyzberg, S. Spaulding, M. Toneva, and B. Scassellati, "The Physical Presence of a Robot Tutor Increases Cognitive Learning Gains," *34th Annual Conference of the Cognitive Science Society*, no. 1, pp. 1882–1887, 2012.

[16] L. J. Corrigan, C. Peters, and G. Castellano, "Identifying task engagement: Towards personalised interactions with educational robots," in *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, pp. 655–658, Sep. 2013.

[17] D. Leyzberg, S. Spaulding, and B. Scassellati, "Personalizing robot tutors to individuals' learning differences," in *HRI '14*, 2014.

[18] P. Baxter, E. Ashurst, R. Read, J. Kennedy, and T. Belpaeme, "Robot education peers in a situated primary school study: Personalisation promotes child learning," *PLOS ONE*, vol. 12, p. e0178126, May 2017.

[19] P. Dillenbourg, M. J. Baker, A. Blaye, and C. O'Malley, "The evolution of research on collaborative learning," in *Learning in Humans and Machine: Towards an interdisciplinary learning science.* (Spada, E., Reiman, and P., eds.), pp. 189–211, Elsevier, Oxford, 1995.

[20] P. Dillenbourg, "What do you mean by collaborative learning?," *P. Dillenbourg (Ed) Collaborative-learning: Cognitive and Computational Approaches*, pp. 1–19, 1999.

[21] J. Hamari, D. J. Shernoff, E. Rowe, B. Coller, J. Asbell-Clarke, and T. Edwards, "Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning," *Computers in Human Behavior*, vol. 54, pp. 170–179, 2016.

[22] C. R. Henrie, L. R. Halverson, and C. R. Graham, "Measuring student engagement in technology-mediated learning: A review," *Computers & Education*, vol. 90, pp. 36–53, Dec. 2015.

[23] L. J. Corrigan, C. Basedow, D. Küster, A. Kappas, C. Peters, and G. Castellano, "Perception matters! engagement in task orientated social robotics," in *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 2015.

[24] M. Csikszentmihalyi, *Creativity: the psychology of discovery and invention.* modern classics, New York: Harperperennial, 1. ed ed., 1996. OCLC: 825754060.

[25] G. Castellano, I. Leite, A. Pereira, C. Martinho, A. Paiva, and P. W. McOwan, "Detecting engagement in hri: An exploration of social and task-based context," *2012 International Conference on Social Computing*.

[26] B. C. Ko, "A Brief Review of Facial Emotion Recognition Based on Visual Information," *Sensors (Basel, Switzerland)*, vol. 18, Jan. 2018.

[27] M. Kazemilari and A. Mohamadi, "Topological Network Analysis Based on Dissimilarity Measure of Multivariate Time Series Evolution in the Subprime Crisis," *International Journal of Financial Studies*, vol. 6, no. 2, p. 47, 2018.

[28] A. Vinciarelli, M. Pantic, D. Heylen, C. Pelachaud, I. Poggi, F. D'Errico, and M. Schroeder, "Bridging the Gap between Social Animal and Unsocial Machine: A Survey of Social Signal Processing," *IEEE Transactions on Affective Computing*, Jan. 2012.