The Effect of Repeating Tasks on Performance Levels in Mediated Child-Robot Interactions

Paul Baxter
Lincoln Centre for Autonomous Systems
School of Computer Science
University of Lincoln (U.K.)
Email: pbaxter@lincoln.ac.uk

James Kennedy, Emily Ashurst, Tony Belpaeme
Centre for Robotics and Neural Systems
The Cognition Institute
Plymouth University (U.K.)
Email: first.last@plymouth.ac.uk

***Abstract*—That "practice makes perfect" is a powerful heuristic for improving performance through repetition. This is widely used in educational contexts, and as such it provides a potentially useful feature for application to child-robot educational interactions. While this effect may intuitively appear to be present, we here describe data to provide evidence in support of this supposition. Conducting a descriptive analysis of data from a wider study, we specifically examine the effect on child performance of repeating a previously performed collaborative task with a peer robot (i.e. not an expert agent), if initial performance is low. The results generally indicate a positive effect on performance through repetition, and a number of other correlation effects that highlight the role of individual differences. This outcome provides evidence for the variable utility of repetition between individuals, but also indicates that this is driven by the individual, which can nevertheless result in performance improvements even in the context of peer-peer interactions with relatively sparse feedback.**

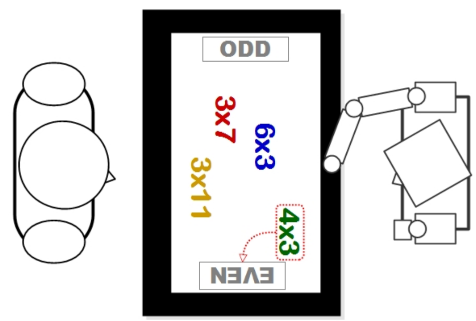

Fig. 1. Child (left) and robot (right) playing a collaborative maths sorting game (categorising the result of the multiplications as either odd or even in this example) on a large touchscreen located between them. Images on screen and dashed sample image path shown for illustration only; not to scale.

I. Introduction

The research-oriented application of social robots to educational contexts (in particular for children) has been rapidly increasing in recent years. The applications have spanned schools [1], healthcare [2], and extracurricular activity scenarios [3], and covered a wide range of subjects and skills, from nutrition [4] to handwriting [5].

Such work frequently attempts to bootstrap from contemporary learning theories. While methods such as learning by rote (effectively memorisation without explicit emphasis on understanding) were formerly stardard educational practice (based partly on behaviourist ideas [6]), more recently constructivist and related approaches have come to the fore [7] even though in practice the behaviourist approach frequently remains in place (the use of reinforcers and testing to name but two). Robotics applications have thus in principle typically followed this latter approach, emphasising concepts such as collaboration [8], social partership [9], guided discovery [10], and others.

Typically, given the as yet novel nature of such child-robot interaction work, these studies generally take a relatively high level perspective, emphasising metrics such as user preference/opinion and/or overall learning effects. However, as such applications mature, it will become necessary to perform more fine-grained analyses in order to establish the conditions under which children may maximise their learning with a robot, and the features (behavioural, morphological, etc) of the robot that best facilitate this learning process. The study described in this paper is presented in this context.

In this paper, we describe data that demonstrates the effect on child performance of repeating a collaborative task with a robot, if the performance of an initial attempt is low. Overall, the results do suggest that there is some benefit conferred by such repetition. However, one strong theme that further emerges from this analysis is the high variability of the effect, which indicates the importance of individual differences. First the study is described, highlighting the embedded nature of the experiment in a classroom and the collaborative nature of the task (section II), followed by results that explore overall phenomena and effects at an individual level (section III), before further discussing these effects at the end of the paper (section IV).

II. Study

The aim of the study presented in this paper is to assess the effect of repeating a task with a robot in which initial performance was low. To achieve this, we analyse a sub-set of data obtained from a larger study, which sought to assess the impact of embedded (i.e. present in the classroom itself with no experimenters present) personalised robot peers in a classroom on learning [11]. Using a two-condition setup (personalised intervention condition and non-personalised control condition), results indicate that personalisation supports additional learning, particularly in novel subjects [11], and that teachers will take advantage of a robot in their classroom for wider moivational purposes than just the task to be performed with the robot [12]. Since performance-based repetition was only present in the intervention condition, this is the data that we analyse below.

Experimental hypotheses are not ventured for the present paper due to its placement within the wider study. As such, we provide a descriptive analysis of the data obtained with respect to the effect of repetition, as a means of providing initial indications of effects that can be subsequently taken up in further studies in their own right.

A. Ethics

Approval for conducting this study was granted by the Plymouth University Faculty of Science and Technology Human Ethics Committee, as part of a thematic programme of research involving the robot and touchscreen setup, and children in local schools. An opt-out informed consent was obtained in writing from the parents/guardians of all participating children. It was made clear to all children that they could withdraw if and when they wished to.

B. Environment and Subjects

The study took place at a U.K. primary school towards the end of an academic year, in two matched age and ability classes, corresponding to the two conditions. A total of 59 children took part aged 7–8, 30 of whom were in the intervention condition of interest, and thus the primary focus of attention below (12 boys, 18 girls).

A robot and a 26” touchscreen (with supporting hardware) was placed in each of the classrooms permanently for a two week period. While in use during the school day there were no experimenters present; supervision was provided by class teacher, or teachers assistant. Interactions between a single child and the robot occured around the touchscreen (figure 1), which provides a mediator for the interaction – the context in which the child-robot interaction takes place.

C. Learning Task

The one-on-one interactions between the child and robot take place in the context of a broadly collaborative [8] sorting task, where the robot acts as a peer in attitude (e.g. informal, personalised, uses child’s name) and knowledge (e.g. makes mistakes to the same extent as the child). This sorting task is centred on a large touchscreen [13], on which there are a set of images that need to be sorted into one of two categories (see figure 1 for an example). Each such set of images is labelled an “image library”. In the present work, a total of 18 image libraries are used, with two subjects used (see table I): a familiar task for the children (maths times-tables), and an unfamiliar task (history – about the stone age). Each image library has an equal number of images for the two categories, with a total of 12 images for each maths image library and 14 images for the stone-age image libraries. The maths image libraries were organised such that there was a progressive

TABLE I

IMAGE LIBRARIES USED IN THE STUDY, SPLIT INTO FAMILIAR (MATHS TIMES TABLES) AND UNFAMILIAR (THE STONE AGE) TOPICS. NOTE THAT THE IMAGE LIBRARIES WERE INTERLEAVED DURING INTERACTIONS WITH THE ROBOT, AND THAT IN THE CASE OF THE FAMILIAR TOPIC, THE IMAGE LIBRARIES WERE ARRANGED IN INCREASING DIFFICULTY.

Maths (Familiar)		
Library	Contents	Task*
1	2x table	In/Out
3	10x table	In/Out
5	5x table	Odd/Even
7	2,10,5 div	Odd/Even
9	3x table	Odd/Even
10	4x table	In/Out
11	6x table	In/Out
12	3,4,6 div	Odd/Even
14	7x table	In/Out
15	8x table	In/Out
16	9x table	Odd/Even
17	11x & 12x	Odd/Even

Stone-Age (Unfamiliar)		
Library	Contents	Task*
2	SA lifestyle	Yes/No
4	SA animals	Yes/No
6	SA tools	Yes/No
8	SA art	Yes/No
13	SA mixed	Yes/No
18	SA mixed	Yes/No

* Task is a categorisation: the labels are the two categories shown on the screen

increase in difficulty. This arrangement was verified with the class teachers prior to the study.

In this collaborative game setting, both the child and the robot have the same interaction affordances; i.e. they are both able to select an image, drag it, and deposit it in one of the category locations (see figure 1 for an example). There are no turn-based constraints, and overlapping actions on the touchscreen is possible – although in actual interaction, a turn-based structure does nevertheless appear to emerge from the interaction [14], indicating that in this context, the robot can be seen as a (potentially) social agent by the child.

Further supporting the notion that the robot was a peer, feedback to image categorisation moves on the touchscreen was provided visually on the screen itself (green tick or red cross): from the perspective of the child, the robot thus had the same feedback on performance that they had. The robot did however comment on the child’s moves (e.g. “well done”, or “maybe you’ll do better on the next one”). No additional feedback information regarding individual images was provided: this is therefore a relatively sparse feedback regime. At the end of the image library (i.e. when all images had been sorted), and if the performance was below threshold, then the robot would make a brief comment (e.g. “oh dear, looks like the computer will make us do that one again”) to indicate that a repeat would occur.

The main feature of the learning task with respect to the present paper is the possibility for repeating an image library if performance of the child is low. Since both the child and the robot are able to make categorisation moves on the touchscreen, we consider only the child’s performance: i.e. only those moves made by the child on the touchscreen. Given that chance performance is 50% (two categories, equal number of members of each category), we consider acceptable performance to be at least 65% correct classifications (with a maximum number of three attempts). If the child’s performance falls below this, then the library is reset once completed (i.e. a rearrangement of the same images on the touchscreen), up to a maximum of three times, after which the next image library would be shown. If an image library was completed successfully, then the next

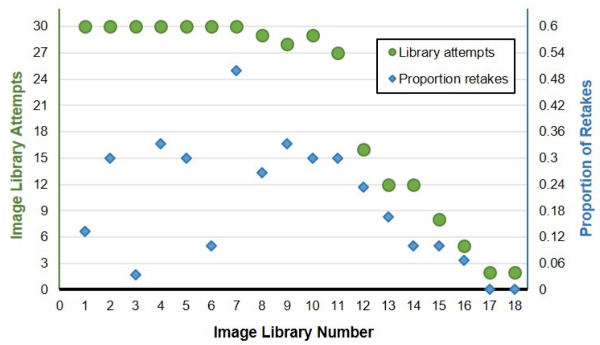

Fig. 2. Distribution of completions and proportion of repeats across image libraries for all children in the intervention condition.

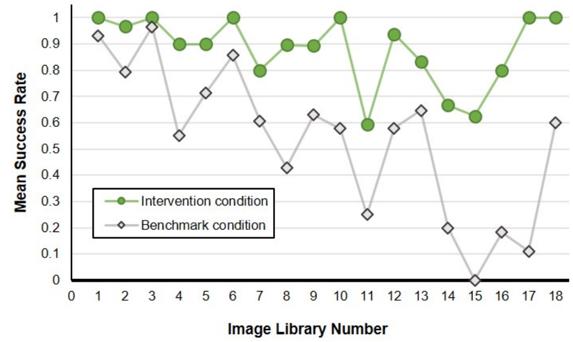

Fig. 3. Success rate for all image libraries in the intervention condition, and control condition. Success in a library is a child score of greater than 65% correct image classification.

image library (table I) would be automatically dispayed on the screen.

D. Procedure and Metrics

The hardware was set up in a corner of the respective classrooms at the start of the two-week experimental period, and remained in situ until the end. The system was started up each morning prior to the arrival of the children, and was shut down at the end of the school day after the children had left. No experimenters were present during the interactions of the children with the robot.

Over the course of the day, the teacher would nominate one child to interact with the robot at a time. This child would go over to the robot setup and interact while the rest of the class carried on with their normal activities. Each interaction would last five minutes (of interaction with the image libraries, not including introduction and closing procedures); over the course of the two week period, each child interacted with the robot on multiple occasions.

During each interaction, a range of information was collected. This included, for each child, the number of libraries completed, the child's score, and the number (and effect, in terms of score) of repeated image libraries. It is this data that is the primary subject of investigation below. In the wider study, a number of other metrics were recorded, including questionnaires, pre- and post-study knowledge tests, and video recordings – further details of these appear in [11].

III. Results

We reiterate at this point that the aim of this paper is to provide a descriptive analysis of data obtained that can be used as a basis for subsequent explorations, rather than as a hypothesis-led effort. Hence, while we make observations on a number of trends and relationships, we must leave further characterisation to future work. We further note that (unless otherwise stated), we focus on the results obtained in the intervention condition, i.e. the group of 30 children for whom there was the possibility of repeating image libraries.

A. Occurrence and Impact of Repeats

Not all of the libraries were completed by all 30 of the children over the course of the study (figure 2): after image library 11 (6x table), there is a sharp drop-off in completion rate. Considering the repeat rate for each library, it can be seen that there are a wide range of values. Where it may be expected that, for the maths libraries at least, increasing difficulty (seen in higher image library numbers) would result in a greater need for repeats, this is not evident from the data. A positive correlation is found here ($r = 0.914$, $n = 30$, $p < 0.001$), although this is likely to be due primarily to the drop-off in completion rate: those children likely to have progressed through more of the libraries may have been higher performing, hence requiring fewer repeats in the first place.

Repeating an image library does generally appear to confer an advantage in terms of score, when contrasted with a scenario in which no repetition is possible (figure 3). This provides an initial indication in support of the intuition that repetition of a task with a robot provides some advantage – however, due to the setup of the experiment, with a number of factors different between the conditions in addition to the possibility for repeats, this is not, on its own, conclusive.

In order to provide further insight, the impact of repeats per image library can be examined (figure 4). This shows that for most libraries where there are repeats, there is a score improvement from the first to the last attempt ($mean_{increase} = 0.218$, $n = 30$, 95% CI=[0.177,0.258]). The mean score change for the first image library seems to be an outlier here: it is likely to be due to uncertainty on the part of the four children as to what should be done; a shortfall quickly overcome on the second iteration. Indeed, each of these four individuals only had one repeat attempt.

B. Individual and Topic Differences

The overall difference in mean repeat rates between the maths libraries ($mean_{maths} = 0.424$, $n = 30$, 95% CI=[0.3,0.548]) and the stone age libraries ($mean_{SA} = 0.346$, $n = 30$, 95% CI=[0.217,0.475]) is small (with a large overlap in the 95% CIs). Examining the number of repeats per child across all image

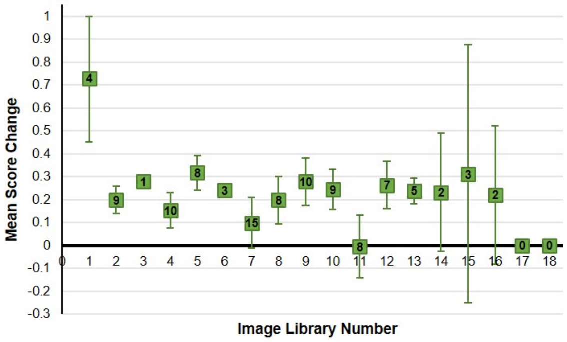

Fig. 4. Effect on score of repeat attempts, by image image library. Numbers in data points show number of repeats for that library across all children. Error bars show 95% CI.

libraries shows a high variability between children (figure 5). This seems to suggest that instead of looking at the group as a whole (i.e. is repetition generally a good strategy), it is necessary to consider the effects on individuals (i.e. under what circumstances and features of individuals does repetition confer a benefit to these individuals).

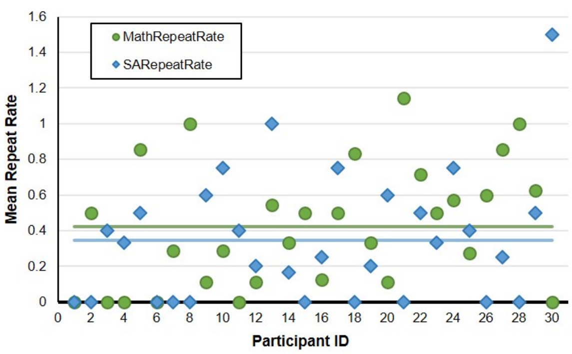

Fig. 5. Mean number of repeated attempts of image libraries per child, for the maths and stone age image libraries. Horizontal lines show mean for each image library subject.

This refocus on individual differences is further supported by considering the mean change in score achieved by each child (figure 6). While the difference in overall means is more pronounced between the image library subjects ($mean_{maths} = 0.132$, $n = 30$, 95% CI=[0.085,0.18]; $mean_{SA} = 0.076$, $n = 30$, 95% CI=[0.046,0.106]), a high degree of inter-subject variability is apparent[1]. Considering the relative performance increase for the two image library subjects, 18 children gained more from repeating maths image libraries, whereas only 10 individuals gained more from repeating the stone-age image libraries (two children did not repeat any image libraries).

[1]The mean values include values for those children who did not perform repeats. This is because repeat rate (or lack thereof) is a feature of the intersubject variability under examination; to exclude these instances would therefore be to skew the distribution under consideration.

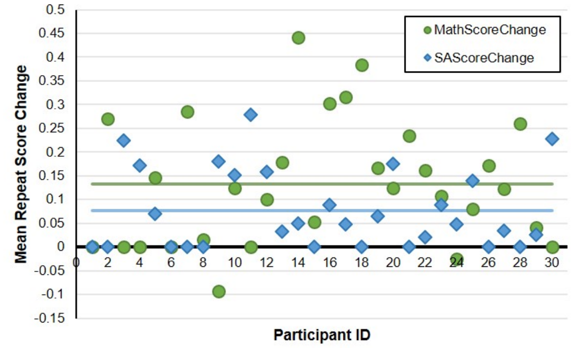

Fig. 6. Mean change in score after repeats per child, for the maths and stone age image libraries. Horizontal lines show mean for each image library subject.

TABLE II
CORRELATION MATRIX FOR MATHS TIMES TABLES (FAMILIAR) RESULTS. CELLS HIGHLIGHTED IN GREEN HAVE P<0.05, IN YELLOW IS P<0.1. N=30 FOR ALL CORRELATIONS. *Perf*: OVERALL CHILD CLASSIFICATION PERFORMANCE. *Rep rate*: MEAN REPETITION RATE. *Tot reps*: TOTAL NUMBER OF REPEAT ATTEMPTS. *ΔScore*: CHANGE IN SCORE, PRE- TO POST-REPEAT. *N_libs*: TOT NUMBER OF IMAGE LIBRARIES COMPLETED.

	Gender	Perf	Rep rate	Tot reps	ΔScore	N_libs
Gender	1					
Performance	-0.075	1				
Rep rate	0.124	-0.760	1			
Tot reps	0.100	-0.770	0.960	1		
ΔScore	0.278	-0.236	0.450	0.402	1	
N_libs	0.192	0.321	-0.245	-0.239	-0.125	1

TABLE III
CORRELATION MATRIX FOR STONE-AGE (UNFAMILIAR) RESULTS. CELLS HIGHLIGHTED IN GREEN HAVE P<0.05, IN YELLOW IS P<0.1. N=30 FOR ALL CORRELATIONS. LABELS AS FOR TABLE II

	Gender	Perf	Rep rate	Tot reps	ΔScore	N_libs
Gender	1					
Overall Perf	-0.075	1				
Rep rate	0.255	-0.112	1			
Tot reps	0.072	-0.182	0.931	1		
ΔScore	0.138	0.289	0.583	0.584	1	
N_libs	0.183	0.313	0.226	0.133	0.377	1

C. *Indications from Correlations*

In order to explore what individual influences there are on the effect of repeating image libraries on performance within the context of this study, we explore correlations between the various metrics recorded during the study. This form of analysis naturally does not provide proof of causality, but it can provide indications of trends, and relationships that could be explored further. First we break this down by image library subject (tables II and III), before considering the relationship between the two (table IV).

As would be expected, there is a strong (and significant) association between repeat rates and total number of repeats. Similarly, and in support of figure 4, there is a strong and statistically significant association between repeat rate (and total number of repeats) and score change, for both image library subjects: i.e. the greater the number of repeats, the greater the change in score.

However, one clear difference between the correlations for

TABLE IV

CORRELATION MATRIX COMPARING MATHS AND STONE-AGE RESULTS. CELLS HIGHLIGHTED IN GREEN HAVE P<0.05, IN YELLOW IS P<0.1. N=30 FOR ALL CORRELATIONS. *Math/SA-Libs*: TOTAL NUMBER OF IMAGE LIBRARIES ATTEMPTED OF RESPECTIVE SUBJECTS. *Maths/SA-Re*: NUMBER OF REATTEMPTS FOR EACH RESPECTIVE SUBJECT.

	Perf	SA-Libs	SA-Re	ΔSA	SA-Success
Perf	1	0.313	-0.182	0.289	0.214
Math-Libs	0.321	0.923	0.134	0.333	-0.075
Maths-Re	-0.770	-0.297	-0.261	-0.663	-0.120
ΔMath	-0.236	-0.083	-0.295	-0.445	-0.160
Maths-Success	0.644	0.320	0.067	0.323	-0.031

the two image library subjects is in the relationship between the repetition rate (and total repeats) and the overall image library performance (mean per child over the whole study period): for the maths times tables image libraries this is a strong negative correlation (significant), whereas this is only weak (non-significant) for the stone-age image libraries.

Considering the relationships between maths and stone-age image library-related behaviour provides some further insight into individual differences (table IV). Firstly, as would be expected, the number of maths and stone-age image libraries completed is strongly positively correlated. Secondly, there is a strong positive correlation between the overall performance and maths image library success rate, but not for the stone-age image library success; this is despite there being an overall higher success rate for the stone-age ($mean_{SA} = 0.934$, $n = 30$, 95% CI=[0.887,0.982]) than maths image libraries ($mean_{maths} = 0.862$, $n = 30$, 95% CI=[0.816,0.908]). Thirdly, there is a moderate negative (significant) correlation between change in stone-age performance after repeats and both number of math repeats and change in math score. Furthermore there is a strong negative correlation between overall performance and number of math image library repeats (the more repeats needed, the lower the overall score, and vice versa). The presence of only a few significant results here make patterns and trends difficult to extract, but in general the results seem to suggest that performance and change in performance is inversely associated for the two image library subjects.

One final aspect to note regarding the separate image library subject correlations is that the gender of the child does not appear to be strongly (or significantly) associated with any of the other variables. For this reason, the effect of gender is not considered further for the present paper (although there may be related phenomena worth further investigation).

IV. DISCUSSION

A central facet of the experimental setup and task context used is that it is a fundamentally collaborative task between a child and a robot (figure 1). Note that despite collaboration not being enforced (i.e. rather than having an explicit turn-taking structure, it is possible for the child to complete the task on their own if he/she ignores the robot), collaborative behaviours are indeed typically observed [14]. It is in this interactive context that the results obtained should be considered. With the robot taking on the role of a peer (see section II-C), the extent of feedback provided to the child is relatively sparse (owing to the desire for the robot to have the same level of apparent knowledge as the child). Nevertheless, this feedback serves to highlight to the child where image libraries are to be repeated: any subsequent change in performance (such as the mean increase observed in the present study) may thus be mediated by this interactive context. The collaborative nature of the task may also provide additional motivation for performance improvement (beyond the desire to move on to another image library) [15], although this effect requires further empirical investigation.

The results have shown that at the group level, there is some apparent benefit for repeating a categorisation task if initial performance is low (figure 4), and that this benefit is greater for the familiar subject than the unfamiliar one (figure 6). Familiarity may in this case not be the only distinguishing characteristic between the two types of image library, with other aspects such as the level of abstraction or topic-related enjoyment that may be important: this requires further investigation, although we note that levels of self-reported enjoyment remain high [11]. However, it is also clear that (as may perhaps be expected) there is a high degree of variability between subjects. Examining these more closely indicates that repeats for maths is positively correlated with score change, but inversely correlated with overall performance – a relationship not present for the stone-age subject.

One feature of the results is that both the overall repeat rate and the overall score change as a result of repeats is higher for the familiar subject (maths) than for the unfamiliar subject (stone-age). We suggest that this may be related to the sparsity of the feedback: recall that correct/incorrect feedback is only provided on the touchscreen itself in response to a categorisation. In a familiar task, the children would already know the features of the problem (what is involved in multiplication for example), and so even sparse feedback is confirmatory. Conversely, this may not be true for a novel problem, in which case only sparse feedback may not be as helpful. This seems to be supported by the correlation results, where there was a strong negative correlation between repeat rates and overall performance for the familiar subject, but not for the unfamiliar subject. This leads to a hypothesis for future study that for unfamiliar tasks (to the children), richer feedback is required than for familiar tasks.

Note however that the relationship between the group data and the correlations remains ambiguous in some respects. For example, the negative correlation between change in math performance and change in stone-age performance (table IV) requires further investigation. It is likely that the wider context for the individual needs to be taken into account, as in the discussion of sparse feedback and role of interactivity above. One possibility not explored in the present study is the role of attention: repetition could be a means of re-orienting attention back to the task after a lapse of concentration or misunderstanding (cf. the outlier mean score change in the first image library, figure 4). More generally, the question is – what characteristics of the individual (or the circumstances) predispose them (or not) to gain more from repetition?

This widening of scope seems necessary when performing a more fine-grained analysis. Returning to the notion of repeating collaborative tasks based on initial performance, we have seen that while in general there could be some benefit, it is necessary to consider this from the perspective of the children's individual differences and of the task engaged in (familiar and unfamiliar in this case). While the present study has only provided a descriptive analysis of the data obtained, it provides a number of pointers to phenomena that should be further researched.

ACKNOWLEDGMENT

This work was supported by the EU FP7 project DREAM (grant number 611391, http://dream2020.eu), and the H2020 project L2TOR (grant number 688014, http://www.l2tor.eu). The authors would like to thank Salisbury Road Primary school (Plymouth, U.K.) for their participation in the study.

REFERENCES

[1] F. Tanaka and S. Matsuzoe, "Children Teach a Care-Receiving Robot to Promote Their Learning: Field Experiments in a Classroom for Vocabulary Learning," *Journal of Human-Robot Interaction*, vol. 1, no. 1, pp. 78–95, 2012.

[2] T. Belpaeme, P. Baxter, R. Read, R. Wood, H. Cuayahuitl, B. Kiefer, S. Racioppa, I. Kruiff-Korbayova, G. Athanasopoulos, V. Enescu, R. Looije, M. Neerincx, Y. Demiris, R. Ros-Espinoza, A. Beck, L. Canamero, A. Hiolle, M. Lewis, I. Baroni, M. Nalin, P. Cosi, G. Paci, F. Tesser, G. Sommavilla, and R. Humbert, "Multimodal Child-Robot Interaction : Building Social Bonds," *Journal of Human-Robot Interaction*, vol. 1, no. 2, pp. 33–53, 2012.

[3] I. Leite, G. Castellano, A. Pereira, C. Martinho, and A. Paiva, "Modelling Empathic Behaviour in a Robotic Game Companion for Children : an Ethnographic Study in Real-World Settings," in *HRI'12*. Boston, MA, U.S.A.: ACM Press, 2012, pp. 367–374.

[4] R. Ros, I. Baroni, and Y. Demiris, "Adaptive humanrobot interaction in sensorimotor task instruction: From human to robot dance tutors," *Robotics and Autonomous Systems*, vol. 62, no. 6, pp. 707 – 720, 2014.

[5] S. Lemaignan, A. Jacq, F. Garcia, D. Hood, A. Paiva, and P. Dillenbourg, "Learning by teaching a robot: The case of handwriting," *IEEE Robotics Automation Magazine*, vol. PP, no. 99, pp. 1–1, 2016.

[6] B. Skinner, "The science of learning and the art of teaching," *Harvard Educational Review*, vol. 24, pp. 86–97, 1954.

[7] A. S. Palincsar, "Social constructivist perspectives on teaching and learning," *Annual Reviews of Psychology*, vol. 49, pp. 345–375, 1998.

[8] P. Dillenbourg, "What do you mean by collaborative learning?" in *Collaborative Learning: Cognitive and Computational Approaches*, P. Dillenbourg, Ed. Elsevier, 1999, pp. 1–15.

[9] T. Kanda, T. Hirano, D. Eaton, and H. Ishiguro, "Interactive Robots as Social Partners and Peer Tutors for Children: A Field Trial," *Human-Computer Interaction*, vol. 19, no. 1, pp. 61–84, jun 2004.

[10] J. Kennedy, P. Baxter, and T. Belpaeme, "Comparing Robot Embodiments in a Guided Discovery Learning Interaction with Children," *International Journal of Social Robotics*, vol. 7, no. 2, pp. 293–308, 2015.

[11] P. Baxter, E. Ashurst, R. Read, J. Kennedy, and T. Belpaeme, "Robot education peers in a situated primary school study: Personalisation promotes child learning," *PLOSONE*, in review.

[12] P. Baxter, E. Ashurst, J. Kennedy, E. Senft, S. Lemaignan, and T. Belpaeme, "The Wider Supportive Role of Social Robots in the Classroom for Teachers," in *1st Int. Workshop on Educational Robotics at the Int. Conf. Social Robotics*, Paris, France, 2015.

[13] P. Baxter, R. Wood, and T. Belpaeme, "A Touchscreen-Based Sandtray' to Facilitate, Mediate and Contextualise Human-Robot Social Interaction," in *7th ACM/IEEE International Conference on Human-Robot Interaction*. Boston, MA, U.S.A.: IEEE Press, 2012, pp. 105–106.

[14] P. Baxter, R. Wood, I. Baroni, J. Kennedy, M. Nalin, and T. Belpaeme, "Emergence of Turn-taking in Unstructured Child-Robot Social Interactions," in *HRI'13*, no. 1. Tokyo, Japan: ACM Press, 2013, pp. 77–78.

[15] A. Coninx, P. Baxter, E. Oleari, S. Bellini, B. Bierman, O. B. Henkemans, L. Canamero, P. Cosi, V. Enescu, R. R. Espinoza, A. Hiolle, R. Humbert, B. Kiefer, I. Kruijff-korbayova, R. Looije, M. Mosconi, M. Neerincx, G. Paci, G. Patsis, C. Pozzi, F. Sacchitelli, H. Sahli, A. Sanna, G. Sommavilla, F. Tesser, Y. Demiris, and T. Belpaeme, "Towards Long-Term Social Child-Robot Interaction: Using Multi-Activity Switching to Engage Young Users," *Journal of Human-Robot Interaction*, vol. 5, no. 1, pp. 32–67, 2016.