Development of Robot-enhanced Therapy for Children with Autism Spectrum Disorders

Project No. 611391

DREAM
Development of Robot-enhanced Therapy for Children with Autism Spectrum Disorders

Agreement Type: Collaborative Project
Agreement Number: 611391

D2.3.3 Tasks for social robots (supervised autonomous version) on developing social skills

Due Date: 31/03/2019
Submission date: **10/05/2019**

Start date of project: **01/04/2014** Duration: **60 months**
Organization name of lead contractor for this deliverable: **Babes Bolyai University**
Responsible Person: **Daniel David**

Revision: **1.0**

Project co-funded by the European Commission within the Seventh Framework Programme		
Dissemination Level		
PU	Public	**PU**
PP	Restricted to other programme participants (including the Commission Service)	
RE	Restricted to a group specified by the consortium (including the Commission Service)	
CO	Confidential, only for members of the consortium (including the Commission Service)	

Table of contents

Executive Summary

The deliverable *D2.3 Tasks for social robots (supervised autonomous version) on developing social skills* provides the primary results of the evaluation of Robot-enhanced Therapy (RET) in clinical settings. This preliminary version of the deliverable will focus on the randomized clinical trial, including preliminary findings concerning the effectiveness of RET using a supervised autonomous version (*T2.3*). The main outcomes for which the effectiveness of RET is being tested are: joint attention, imitation, and turn-taking skills. Thus, in this deliverable (*D2.3*) we will present the theoretical background, objectives, design, procedure, environmental setup, and final results from the clinical trial conclusions and discussions.

Principal Contributors

The main authors of this deliverable are as follows (in alphabetical order):

Daniel David, Babes-Bolyai University
Anca Dobrean, Babes-Bolyai University
Silviu Matu, Babes-Bolyai University
Radu Soflau, Babes-Bolyai University
Aurora Szentagotai, Babes-Bolyai University

Revision History

Version 1.0 (23-03-2018)
First draft, describing the rationale of the task and the methodology
Version 2.0 (25-03-2018)
Second draft, results included
Version 3.0 (31.03.2019)
Final results of the trial and DREAM Lite study

Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder defined by persistent deficits in social communication and social interaction and restricted, repetitive patterns of behavior, interests, or activities (APA, 2013). ASD children commonly experience difficulties with social interactions. Thus, impairments in imitation, joint-attention and turn-taking skills are common in ASD (Dawson et al., 2004). These impairments are particularly problematic given that imitation, joint-attention and turn-taking are important prerequisites for developing social communication skills. For example, it was found that improvements in imitation can facilitate the recognition of peers and caregivers as “social others”, the hand-eye coordination, and the later development of communication skills (Ricks & Colton, 2010). Imitation also enables children to learn new information from his/her social environment (Cabibihan, Javed, Ang, & Aljunied, 2013). Similarly, joint-attention (i.e., the ability to focus simultaneously on the same object/activity with another social partner) is particularly important for perceiving the social others, for a successful learning (Ricks et al., 2010) and for the acquisition of language (Dawson et al., 2004). In what turn-taking skills are concerned, it appears that they play a fundamental role in regulating conversations (Ricks et al., 2010, Cabibihan et al., 2013) and social interactions. Due to these impairments in social skills, ASD children have difficulties sustaining a conversation or playing a game in which the partners’ roles constantly alternate. As a consequence, the three social skills are often targeted by various interventions that have been developed for ASD.

Applied Behavioral Analysis (ABA) / Cognitive Behavioral-Therapy (CBT) has extensive support for its effectiveness (e.g., Peters-Scheffer, 2011; Virués-Ortega, 2010) for ASD. More recently, ABA / CBT techniques have been adapted in order to be delivered by a social robot to address the difficulties displayed in autism. The use of a social robot in the therapy of ASD children has received increasing attention in the last years, given that social robots are considered to be of potential added value in the interventions developed for children with ASD (Cabibihan et al., 2013; Thill et al., 2012). Still, a relative reduced number of studies have focused on testing whether such robots could ameliorate the social interaction and communication deficits that are characteristic to this population (David, Matu & David, 2014). Although the results are promising (see also Pennisi et al., 2016), most of the available studies implemented single case experiments or had reduced sample size. Moreover, RET has been seldom compared with standard therapist-delivered interventions.

The main aim of the work conducted in this work package is to test *the efficacy* of RET in improving performance of children with ASD by supervised autonomous behavior of the robot. We investigate if a social robot can improve the social abilities of children with ASD and whether RET produces similar or better results than a standard therapist-delivered intervention. The social abilities targeted by our intervention program are imitation, joint-attention and turn-taking. In order to achieve this goal, we run the first large scale *randomized clinical trial* comparing RET with standard intervention for children that have an ASD diagnosis (**Study 1**). In order to ensure high control of the experimental design we used stringent inclusion and exclusion criteria and all intervention session were delivered in a controlled environment by the same experienced therapists.

We extend this work by investigating how RET intervention perform in realistic conditions by running an *effectiveness trail* using a simplified version of the DREAM framework, called DREAM Lite. For this purpose (**Study 2**) we employed a randomized trial design which compared RET with a control group (which did not receive any intervention as part of the study). In this study we used wider inclusion criteria and randomly allocated all children that had ASD symptoms and delivered the intervention in an ecological set-up as provided by multiple special education and treatment centers.

Study 1

Methods

Design

A blinded, randomized, equivalence clinical trial was designed to attain these aims. Participants are randomly assigned to receive one of the two interventions: the standard therapist-delivered intervention (SHT) or the robot-enhanced intervention (RET). The randomization procedure was carried out by an independent researcher using a digital random number generator. The information regarding the allocated condition for each child was communicated to the research team, which has subsequently informed the parents. The protocols of the two interventions are identical (i.e., identical tasks for both groups). The only difference between the two intervention groups is at the level of the interaction partner who delivers the intervention: in one case a therapist and in the other case a social robot (Nao).

Participants

Seventy children (11 females) with a mean age of 4.7 years, with an ASD diagnosis confirmed by the ADOS assessment, were included in the final trail. Three children (2 in the RET condition) were excluded after allocation because they showed high level of target skills at the initial assessment. Twenty-six children completed at least half of the SHT protocol and twenty-five completed at least half of the RET intervention. Figure 1 bellow present the CONSORT diagram summarizing inclusion, exclusion and dropout in each group.

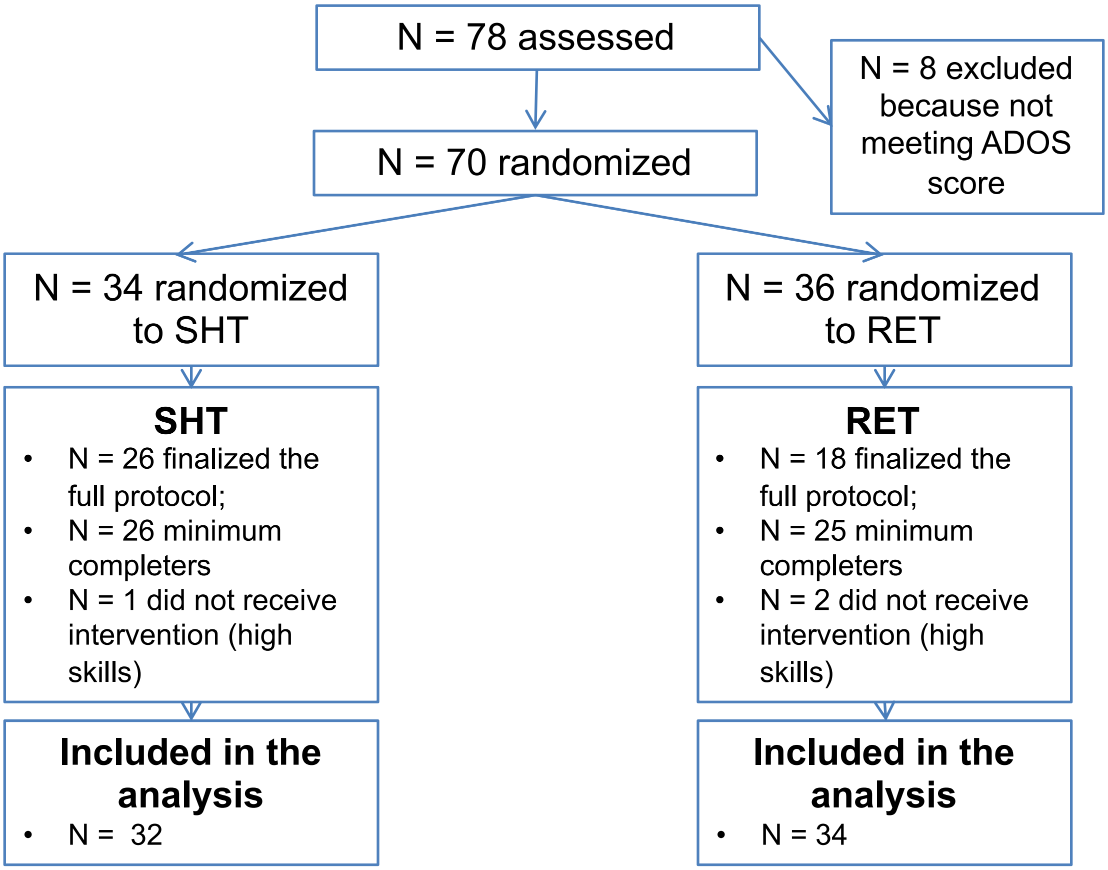

Figure 1. CONSORT diagram presenting randomization, completers, drop-outs and number of subjects included in the analysis for each of the two groups.

Procedure

All recruited participants are required to attend 12 sessions: 2 initial evaluation sessions (i.e., pre-intervention), 8 intervention sessions, and 2 final evaluation sessions (i.e., post-intervention). The 45 minutes sessions are hold bi-weekly. The overall structure of the 12 sessions is described below.

Sessions 1 and 2 are designed to undergo a comprehensive psychological evaluation and to determine the baseline level of the three social abilities (i.e., imitation, joint-attention and turn-taking abilities). The *Autism Diagnostic Observation Schedule (ADOS)* is administrated to each child in order to assess the social and communicative behavior associated with autism. In addition, the ASD screening instrument *Social Communication Questionnaire (SCQ)* is completed by the children's parents/caregivers. The baseline levels of the three targeted social abilities are determined through a series of subtasks that are similar to those designed for intervention (as detailed below).

Sessions 3 to 10 are used for training imitation, joint-attention and turn-taking abilities.

Sessions 11 and 12 firstly aim to determine the level of the three abilities after the interventions. (i.e., through tasks in which the targeted behaviors are preceded by a stimulus but are not followed by any feedback). As it is the case for the initial evaluation sessions, in the post-intervention assessment the standard psychological instruments for ASD symptoms (*ADOS, SCQ*) are also applied, aiming to determine if the benefits of the interventions are generalized.

Intervention protocol

As previously specified, the intervention protocol is identical in the two interventions, concerning the involved tasks. Thus, we will present the shared protocol. The tasks that have been developed for the training of the imitation, joint-attention and turn-taking abilities follow a structured behavioral approach called discrete trial training (DTT). According to Smith (2001), DTT can be particularly useful for the acquisition of new behaviors and for teaching discriminative behavioral responses. Working from the assumptions of this approach, the training tasks take place in a highly structured learning environment and are being directed by the interaction partner (i.e., the Softbank Robotics' Nao robot *OR* a psychotherapist). Thus, all training tasks are organized around a table, which in the case of joint attention and turn taking tasks incorporates a large touch-screen. During training, each targeted behavior/action is preceded by the partner's discriminative stimulus or instruction (e.g., "Do as I do!") and followed by a contingent reinforcement (e.g., "Try again!", "Well done!"). The behaviors are presented over multiple and successive trials and explicit prompting is giving when the child doesn't succeed to accomplish the targeted behavior after several trails. Each action is repeated three times. The intervention protocol is personalized for the needs of each child. Thus, the training of each ability starts from the baseline level (i.e., the level determined in the first two evaluation sessions). As the child's performance improves, the training moves to the next level (i.e., a sub-task of an increased level of difficulty is approached). The goal of the intervention is to reach the highest level possible for each child, on each of the three social abilities. In the following section we will briefly describe the levels of difficulties, as well as the structure of the tasks used to train imitation, joint attention, and turn-taking abilities

A. Imitation

Both the child and the therapist are sitting at a table during the imitation task. If the robot provides the intervention, then the robot is placed on a table in front of the child. Each child is asked to imitate the actions made by the interaction partner (therapist or robot). The imitation ability is trained though the following sub-tasks:

- level 1 of difficulty: imitation with objects (e.g., moving a car, pretending to drink from a cup);
- level 2 of difficulty: imitation of gestures (meaningful movements; e.g., waving one hand and say "bye-bye").

Table 1. The structure of the imitation task.

Instruction
Provided by the interaction partner (robot or human) *"Do as I do!"*
Response
Provided by the child Moving arms/objects in similar ways as the interaction partner.
Consequence
Provided by the interaction partner (robot or human) Depending on the child's answer: *If the child executes the requested movement correctly, he/she receives positive feedback: "Well done!"* *If the child doesn't execute the requested movement, he/she receives encouraging feedback: "Try again!"*

B. Joint-attention

The task is presented to the child in a context (e.g., "Now, we will play another game. In this game I will show you the objects I've seen in an office"; see Table 2). Then, two pictures are displayed simultaneously on a big touch-screen incorporated in the table: one on the left side and the other one on the right side. In this task the child has to look at the picture indicated by the robot/therapist. There are different ways of indicating one of the two displayed images, the number of modalities used to indicate the picture determining the task's level of difficulty:

- level 1 of difficulty: simultaneously looking at one picture, pointing to that picture and saying "Look!";
- level 2 of difficulty: simultaneously looking at one picture and pointing to that picture;
- level 3 of difficulty: looking at one picture.

Table 2. The structure of joint-attention task.

Instruction

Provided by the interaction partner (robot or human) "Please, pay attention to what am I looking at!"
Response
Provided by the child Looking at the picture indicated by the robot/human.
Consequence
Provided by the interaction partner (robot or human) Depending on the child's answer If the child looks at the picture indicated by the robot/human, he/she receives positive feedback: "Well done!" If the child doesn't look at the picture indicated by the robot/human, he/she receives encouraging feedback: "Try again!"

C Turn-taking

The turn-taking ability involves different activities during which the child and his partner have to play by taking turns. The sub-tasks were designed to be implemented on a big touch-screen tablet (Sandtray) and include: sharing information, assigning items to categories and continuing repeating patterns activities. As it can be seen in Table 3, in all subtasks the interaction partner provides an instruction / question before the targeted behaviors and administers a consequence depending on the child's response / behavior.

- sharing information: On the screen of the tablet are displayed simultaneously 5 pictures. In this task the child has to choose a picture from a series of 5 pictures displayed on a touchscreen (when it is his turn) and wait when the robot chooses a picture (when is interaction partner`s turn).
- categories:
    - level 1: 3 pictures are displayed simultaneously on the screen of the tablet (two of them represents categories and the third one is the item that has to be categorized). In this task the child has to categorize the items (when it is his turn) and wait when the robot categorizes (when is interaction partner`s turn). At this level, the categories are familiar to the children between the ages of 3 to 7 years (e.g., fruits vs. vegetables) and the items that have to be categorized appear one by one;
    - level 2: 10 pictures are displayed simultaneously on the screen of the tablet (two of them represents categories and the rest of eight are the items that have to be

categorized). In this task the child has to categorize the items (when it is his turn) and wait when the robot categorizes (when is the interaction partner`s turn). At this level, the categories are more complex and the child has to choose one picture at a time from a larger number of pictures that are simultaneously displayed (ground vehicles vs. water vehicles).

- patterns
    - level 1: 6 pictures are displayed simultaneously on the screen of the tablet (2 of them in the middle of the screen and the rest of them arranged in a string). The child has to continue the pattern illustrated by the string (when it is his turn) and wait when the robot adds a picture to the string (when is robot`s turn). At this level of difficulty the repetitive pattern consists of two or three repetitive items and the only relevant criterion for categorization is the geometrical shape (e.g., rectangle, rectangle, triangle, rectangle, rectangle, ...);
    - level 2: 10 pictures are displayed simultaneously on the screen of the tablet (4 of them in the middle of the screen and the rest of them arranged in a string). The child has to continue the pattern illustrated by the string (when it is his turn) and wait when the robot adds a picture to the string (when is robot`s turn). At this level of difficulty four items repeat and there are two relevant criterions based on which the categorization has to be made: the geometrical shape and its color (e.g., green squire, star, orange squire, circle, green squire, star, ...; see Table 3).

Table 3. The structure of turn-taking task

Sharing information
Instruction
Provided by the interaction partner (robot / human) *a. "It's your turn first! What's your favorite [...]?"* *b. "Now it's my turn!"*
Response
Provided by the child a. The child chooses a picture that represents what he/she likes the most. b. The child waits his turn (doesn't move his/her hands above the touchscreen of the Sandtray when is the partner's turn)

Consequence

Provided by the interaction partner (robot / human)

Depending on the child's answer

a. If the child chooses a picture from those shown on the touch-screen, he/she receives positive feedback: "You showed me very nicely what you like!"

If the child doesn't choose a picture from those shown on the touch-screen, he/she receives no feedback.

b. If the child waits his/her turn (doesn't move his/her hands above Sandtray), he/she receives positive feedback: "You have waited very nicely!"

If the child doesn't wait his/her turn (he/she moves his/her hands above the Sandtray), he/she receives an encouraging feedback: "You have to wait! It's my turn."

Categories

Instruction

Provided by the interaction partner (robot / human)

a. "Let's sort [...]! It's your turn first."

b. "Now it's my turn."

Response

Provided by the child

a. The child categorizes the items.

b. The child waits his/her turn (doesn't move his/her hands above the touchscreen of the Sandtray when is the partner's turn).

Consequence

Provided by the interaction partner (robot / human)

Depending on the child's answer

a. If the child categorizes correctly, he/she receives positive feedback: "You sorted the picture correctly. Well done!"

If the child categorizes incorrectly, he/she receives encouraging feedback: "You sorted incorrectly. Try again!"

b. If the child waits his turn (doesn't move his hands above the Sandtray), he/she receives positive feedback: "You have waited very nicely!"

If the child doesn't wait his/her turn (he/she moves his hands above the Sandtray), he/she receives an encouraging feedback: "You have to wait! It's my turn."

Patterns

Instruction

Provided by the interaction partner (robot / human)

a. "Let's continue the string!"

b. "Now it's my turn."

Response

Provided by the child

a. The child continues the pattern illustrated by the string.

b. The child waits his/her turn (doesn't move his/her hands above the touchscreen of the Sandtray when is the partner's turn).

Consequence

Provided by the interaction partner (robot / human)

Depending on the child's answer

a. If the child continues the pattern correctly, he/she receives positive feedback: "You matched the picture correctly. Well done!"

If the child continues the pattern incorrectly, he/she receives encouraging feedback: "You've matched the picture incorrectly. Try again next time!"

b. If the child waits his/her turn (doesn't move his/her hands above the Sandtray), he/she receives positive feedback: "You have waited very nicely!"

If the child doesn't wait his turn (he/she moves his hands above the Sandtray), he/she receives an encouraging feedback: "You have to wait! It's my turn."

Outcomes and measures

Primary outcomes

The primary outcomes correspond to the three social skills that are being targeted by this randomized clinical trial: imitation, joint-attention, and turn-taking abilities. To determine the level of imitation, children's performance in the imitation task is coded with a score of "1" when the children execute the requested movement correctly and with a score of "0" when children do not execute the requested movement. Similarly, the level of joint-attention is determined based on the performance from the imitation tasks that is coded with a score of "1" – when the children look at the picture indicated by the interaction partner and with a score of "0" - when the children do not look at the indicated picture. In the case of turn-taking, a score of "1" is given when children wait for their turn (don't move their hands above the touchscreen of the tablet when it is the partner's turn) and score of "0" is assigned when the children do not wait for their turn (they move their hands above the tablet).

Secondary outcomes

In addition, we are also investigating some secondary outcomes that are of relevance for ASD interventions. Thus, we also compare the two interventions on engagement in the task, verbal utterances, and knowledge/cognitive abilities. Eye-contact and the reported positive emotions are used as indicators for task engagement, given that this outcome refers to the child's interest and enthusiasm for performing the task. The assessment of verbal utterances targets both initiations and contingent responses (i.e., meaningful verbal productions of child) in this trial. The knowledge/cognitive abilities refer to the extent to which the children are able to share information, sort items, and to continue a repetitive pattern correctly during the turn-taking subtasks. Thus, when the children choose the right picture, share information or sort the items correctly a score of "1" is assigned. Otherwise, a score of "0" is assigned.

Standardized instruments

A number of standardized instruments are being used in order to determine the eligibility of the participants as well as to assess the generalization of results. *Autism Diagnostic Observation Schedule* (*ADOS*; Lord, Risi, Lambrecht et al., 2000) and *The Social Communication Questionnaire* (*SCQ*; Rutter, Bailey & Lord, 2003).

Statistical analysis

Data analysis was conducted using the intent-to-treat principle, meaning that all participants that received at least one intervention session was included in the analysis. In order to be able to include participants that dropped-out, we used a linear mix-model approach with maximum likelihood method which can estimate the relative evolution of the two groups taking into account different number of subjects at different time points. The mixed model included fixed effects for session number, group, level of the difficulty for each task in each session, and baseline scores were necessary. Random intercepts and slopes were added in the model and to see if better fits the data and assessing its parsimony using the AIC indicator. Increases in the AIC lead to the exclusion of the random effect form the model. As a covariance structure, we used an AR(1) auto-regressive model to account for the repeated measure design. We used this type of approach for both primary outcomes (i.e., IM, JA and TT scores) but also for secondary (i.e., measures of engagement).

To test for the equivalence of the two interventions, we computed effect sizes and 95% CI for the difference between the two groups in the final assessment of each of the primary outcomes. For the secondary outcomes, we computed similar effect sizes based on the overall mean in each group. The effect sizes were calculated based on the estimated marginal means and their standard error (from which we derived the standard deviation). The confidence interval for the difference between the two groups was then compared with the equivalence interval established a priori between -0.80 and 0.80, which are the limits of a medium effect size. This analysis allowed us to say with a 95% confidence if the two interventions are equivalent within a medium effect size interval of variation bellow and above 0.

For standardized instruments (i.e., ADOS and SCQ) and other analysis (e.g., baseline comparisons), we used analysis of variance (multivariate or univariate) with between or mixed design, depending on the nature of the data. We also used logistic regression to identify predictors of drop-out.

Results

Pre-test comparison

A Multivariate Analysis of Variance (MANOVA) was computed in order to check if groups were equivalent in baseline across main outcome variables (social skills that are target by the intervention), as well as symptoms severity, as assessed by ADOS.

The results revealed a significant multivariate effect of group, Wilks' Lambda = .021, F (6, 51) = 391.15, p < .000 $\eta^2_{partial}$ = .979. The univariate analysis indicated that this effect was due to a difference between the participants in the two groups in JA skills, F (6, 56) = 11.67, p = .001, $\eta^2_{partial}$ = .172. Descriptive values presented in Figure 2 suggest that the SHT group had better JA skills compared to the RET group at the initial assessment. Baseline JA scores were used as covariates in future analysis for this skill.

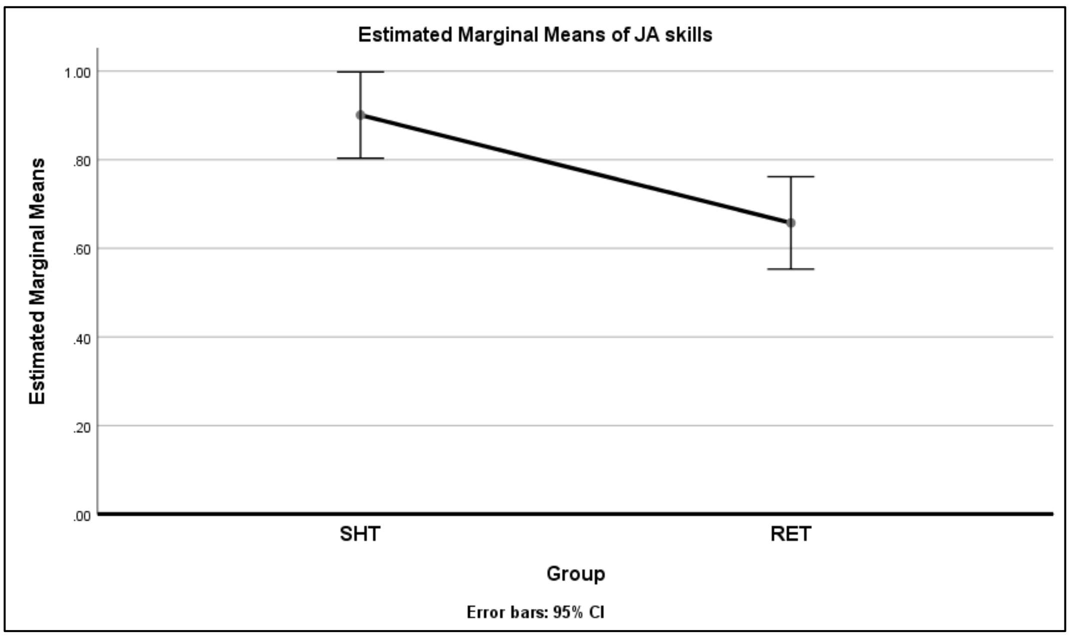

Figure 2. Baseline differences between groups on JA skills at initial assessment.

Dropout analysis

We tested whether the dropout differed between the two groups. For this analysis, we defined a dropout as participant that exited the study before he has completed at least 50% of the intervention sessions (2 initial assessments and 4 interventions sessions). The level of dropout was comparable across the two groups, with a 21.2% dropout in the SHT condition, and a 24.2% dropout in the RET condition. A logistic regression was used to check for additional predictors of dropout. Although the overall model was not significant, χ^2 (8) = 13.36, p = .100, a detailed analysis of the individual coefficients in Table 4, indicates the JA baseline skill are a significant predictor of dropout. The negative coefficient suggest that lower levels of skills are associated with a higher probability of quitting the study.

Table 4. *Logistic regression coefficients for predictors of dropout.*

	B	*Std. error*	*Wald*	*df*	*p*
ADOS total score (pre-test)	.14	.169	.72	1	.395
IM score (pre-test)	-1.06	2.01	.28	1	.598
JA score (pre-test)	-5.04	2.06	6.02	1	.014
TT score for info-sharing task (pre-test)	-.100	1.42	.01	1	.944
TT score for patterns task (pre-test)	1.60	2.08	.60	1	.440
TT score for category task (pre-test)	1.17	2.19	.29	1	.591
Group: SHT vs. RET*	1.48	1.12	1.73	1	.189
Gender: Male vs. Female*	1.12	1.30	.73	1	.393

Note: "*" indicates the category of reference for categorical predictors.

Main outcomes

For IM skills, we found a significant effect of the first level of difficulty indicating that higher difficulty predicted lower performance. We also found an effect of time and more specific comparisons indicated that all session yielded better performance than the initial one, with the exception of the first intervention session which was not statistically different. Figure 3 summarizes the evolution of the two groups across all assessments and interventions for imitation performance.

We computed the effect size and the 95% CI for the difference between SHT and RET on the final assessment measure based on the estimated marginal means and their standard deviations from the above model. The computed effect size was d = 0.05, 95% CI [-0.45, 0.56], which is within the equivalence limit interval of [-0.80, 0.80]; see Figure 8 bellow. This result suggests that the two interventions are equivalent in term of their effectiveness.

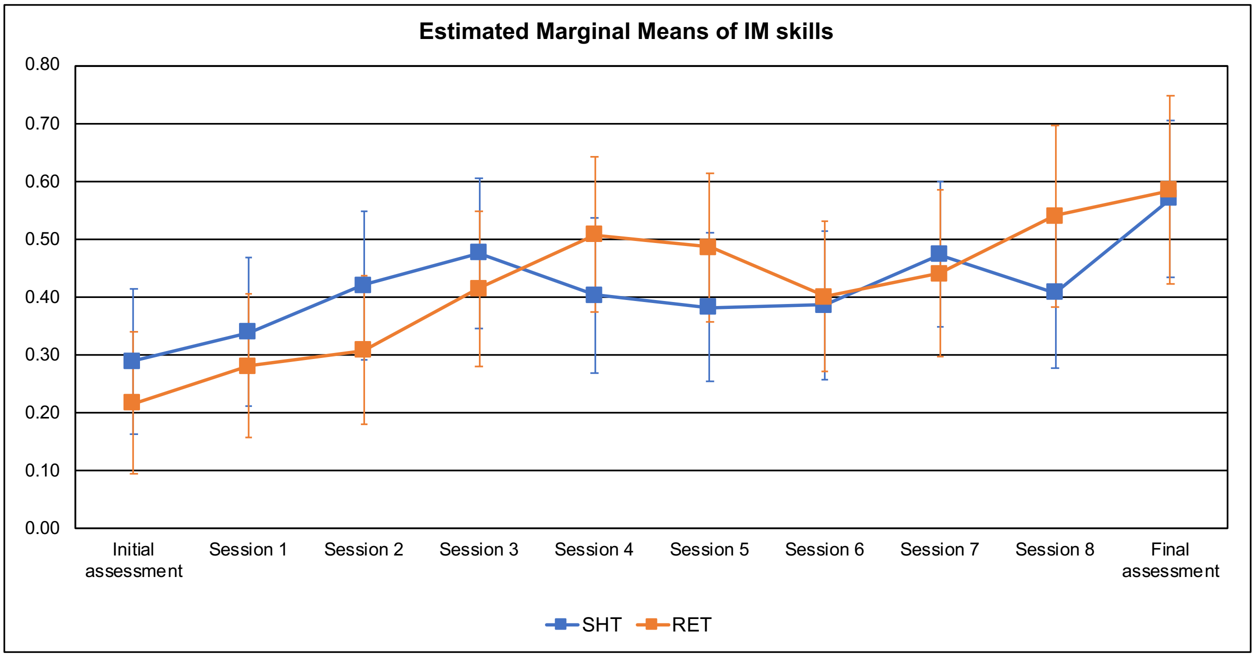

Figure 3. Estimated Marginal Means and 95% CI of IM skills across assessment and interventions sessions.

For JA skills, we found significant effects for the level of difficulty, initial skills levels (used as a control variable) and for the interaction between group and session.

The individual coefficient for the first and second levels of difficulty contrasted to the third one (used as reference) were both positive, indicating that the third level predicted lower performance scores However, the largest coefficient was for Level 2, and this was the only one that significantly differed from the third one, suggesting that look + point led the highest performance. Session number was not a significant predictor for JA skills, suggesting that overall performance has not improved across the full sample. However, the interaction between group and session number was significant. Within and between groups pairwise comparisons indicated that this effect was due to a significant difference favoring SHT on the second intervention session and a difference favoring SHT on the seventh intervention session. Baseline scores were also a significant and positive predictor of JA scores, suggesting that children that had better skills at the beginning of the intervention, performed overall better during the study. Figure 4 summarizes the evolution of the two groups across all sessions for JA.

Similar to IM analysis, we computed the effect size and the 95% CI for the difference between SHT and RET on the final assessment measure based on the estimated marginal means and their standard deviations from the above model. The computed effect size was $d = 0.04$, 95% CI [-0.47, 0.56], which is within the equivalence limit interval of [-0.80, 0.80]; see Figure 8 bellow. Similar to IM, this result suggests that the two interventions are equivalent in term of their effectiveness for JA skill. Yet, it is important to acknowledge that overall there were no significant improvements in this skill across groups.

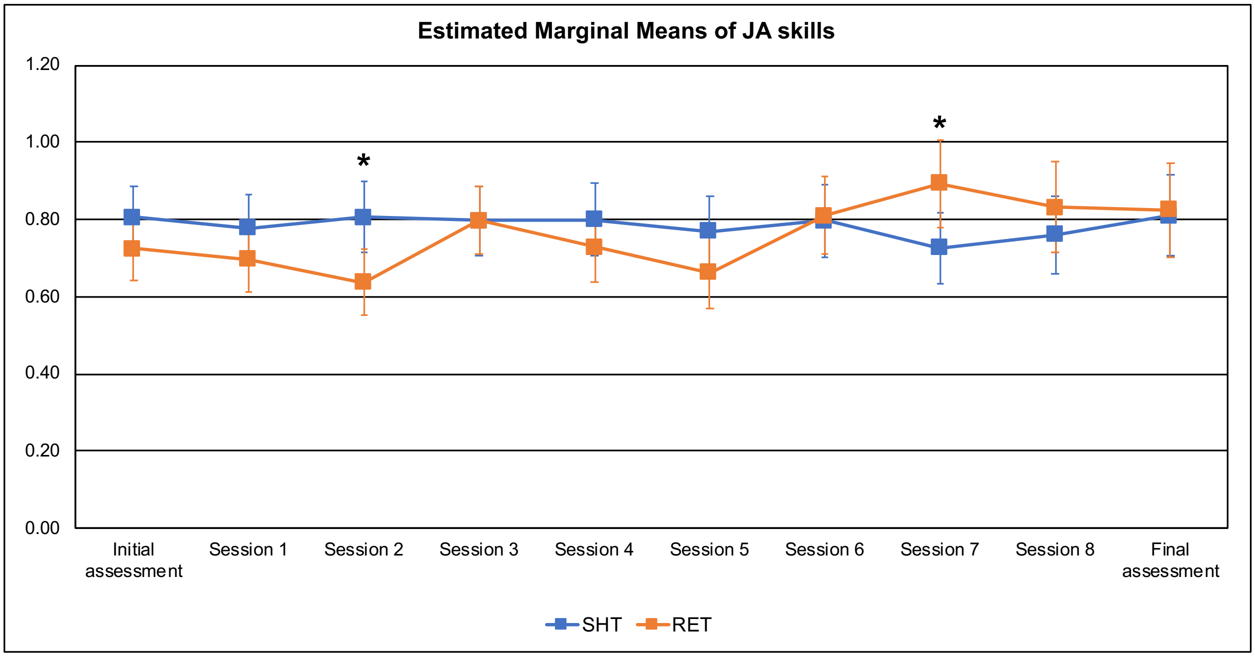

Figure 4. Estimated Marginal Means and 95% CI of JA skills across assessment and interventions sessions. "*" marks statistically significant differences.

For TT, we split the analysis based on the specific type of task, namely info-sharing, patterns completion and categorizing images. The overall results for the comparison between SHT and RET can be found in Figure 8, represented as an average effect size of the difference.

We first analyzed the results for **TT info-sharing** using a similar approach as for the other skills. The fixed effect coefficient for the level of difficulty was only marginally significant. Looking at the detailed parameter estimates, the first level had a negative estimated compared to the second one (used as reference) suggesting that higher difficulty for this task might actually yield better performance. There was a significant fixed effect for session number and pairwise comparisons of estimated marginal means indicated that this difference was due to a significant increase in scores form session one, to the first session intervention across the full sample. No other comparison with the initial assessment indicated significant difference. A similar analysis, using the session number as a continuous covariate indicated similar results as those described above, with a significant and positive the fixed slope for session number, suggesting that overall the children improve their performance on TT info-sharing. Figure 5 summarizes the evolution of the two groups across all assessments and intervention sessions for TT info-sharing.

The computed effect size based on the estimated marginal means and their standard error for the final assessment shows an effect of d = -0.09, with 95% CI [-0.60, 0.42]. These results place the difference between the two interventions in the equivalence limit.

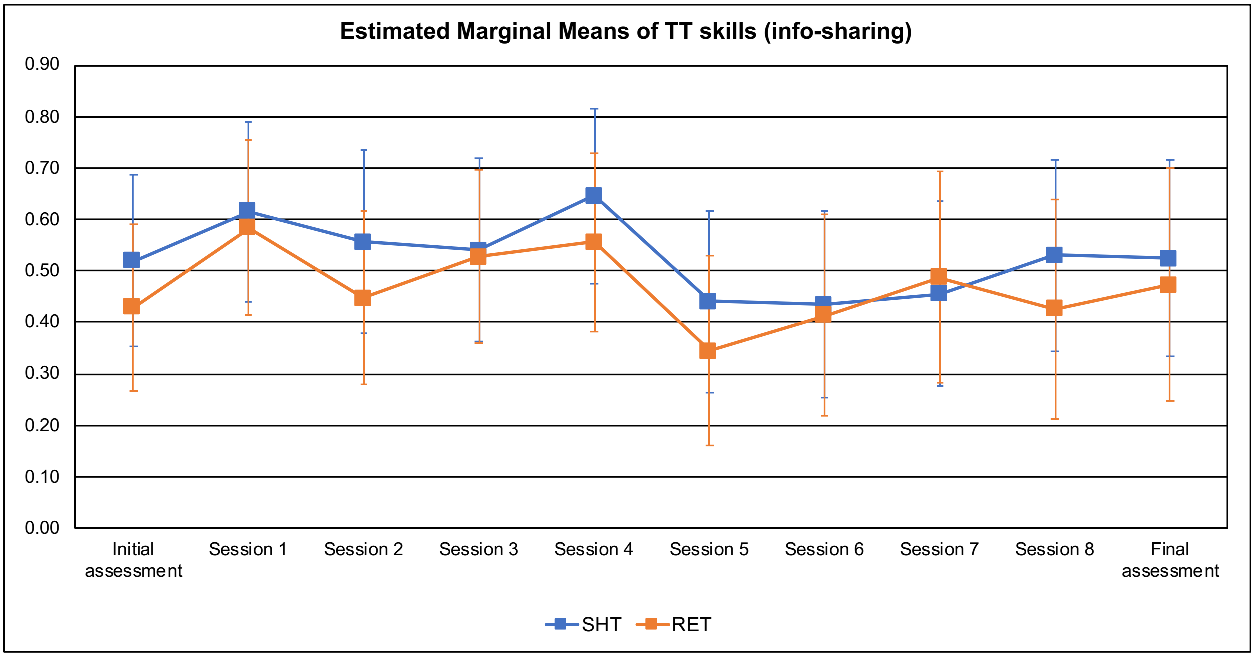

Figure 5. Estimated Marginal Means and 95% CI of TT skills across assessment and interventions sessions for info-sharing.

A similar analysis for **TT patterns** indicated that there was an overall effect for session number which indicated that performance increased across sessions for the full sample. There was also a significant effect for group, which indicated an overall tendency for the SHT group to have higher scores than the RET group, however there was no significant interaction and the groups do not differ significantly at the end of the intervention. Figure 6 sumarizare the evolution of the two groups across these sessions.

The computed effect size at the end of the intervention was d = -0.34 favoring the SHT group, 95% CI [-0.86, 0.17] which is not statistically significant, however it crosses the lower bound of the equivalence limit. This could be interpreted as results being unclear for this specific outcome.

For **TT categories** we found a significant effect of session number and for the interaction between session and group. The specific comparisons indicted that performance increased across sessions and that the interaction was due to between group differences at initial assessment (higher in RET), session 4 and session 8 (for the last two sessions the score was higher in SHT). The were no between-groups differences. Figure 7 summarizes the evolution of the two groups on this outcome.

The standardized difference between the two interventions was d = -0.425 [-0.93, 0.09], which was not significant but outside the equivalence limit, suggesting that the equivalence is unclear for this outcome.

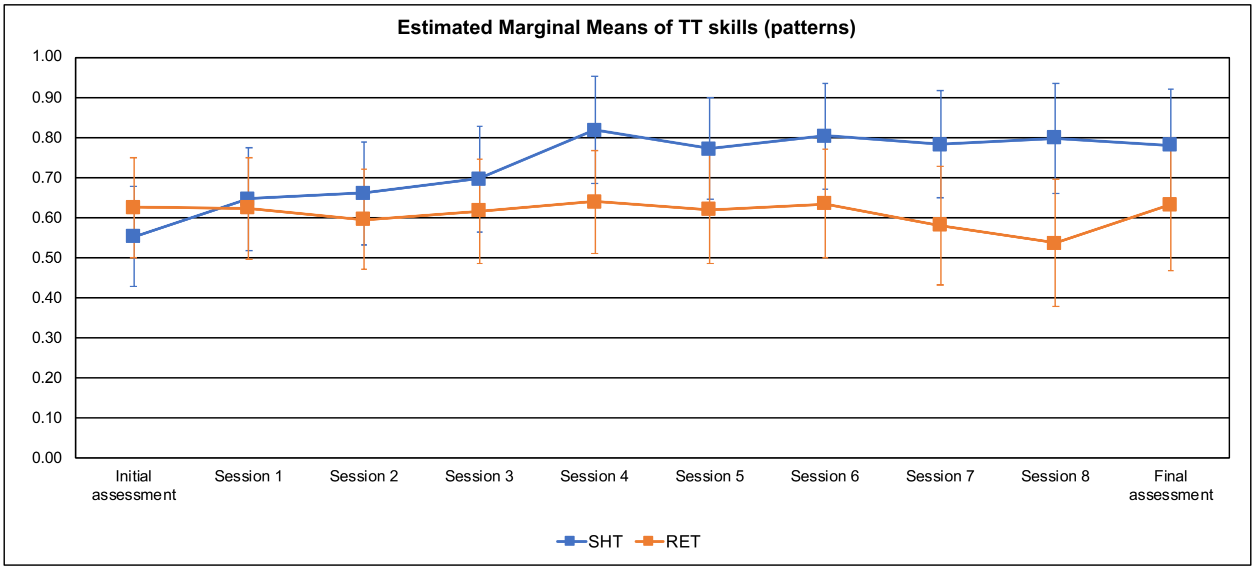

Figure 6. Estimated Marginal Means and 95% CI of TT skills across assessment and interventions sessions for patterns matching task.

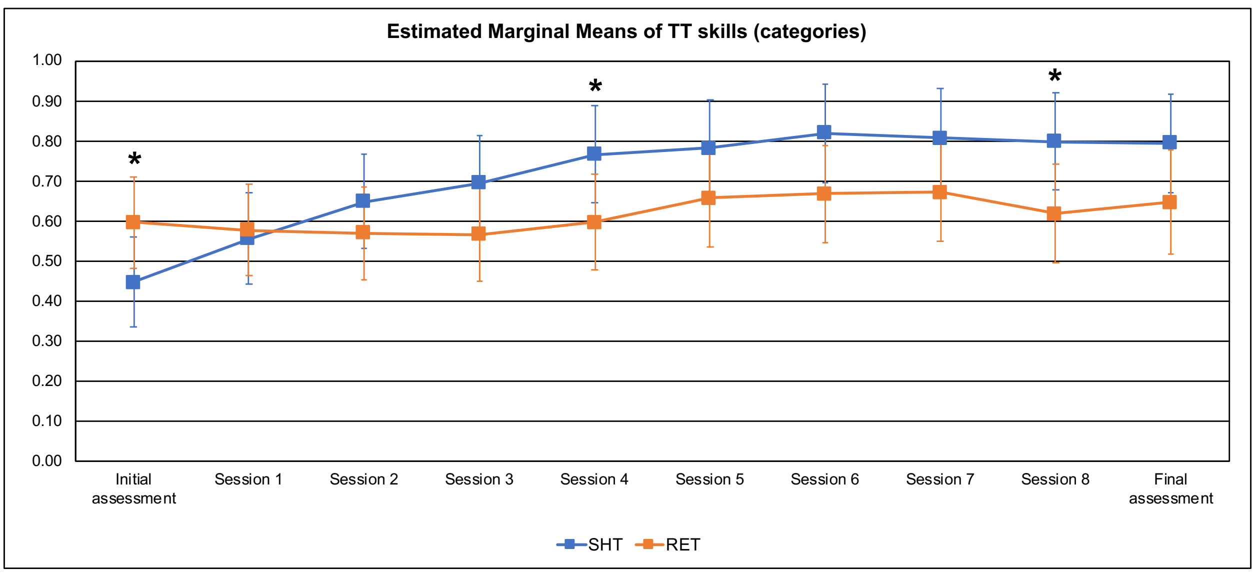

Figure 7. Estimated Marginal Means and 95% CI of TT skills across assessment and interventions sessions for categories sorting task. "*" marks statistically significant differences.

The overall results for the main outcomes suggest that the children in the study experienced significant improvements for almost all trained skills, with the exception of JA, where improvements form baseline were not observed. The two interventions tend to have similar evolutions across time, with few exceptions, the children have similar level of skills at the end of the intervention (there are no statistically significant differences). In terms of equivalence, the difference between the two interventions falls within the equivalence limits, except for TT patterns and categories, where the result is uncertain. However, when averaging the effect size of the difference between the two interventions across TT tasks, the lower bound of the difference falls just within the equivalence interval, indicating that, we can consider them similar. Figure 8 summarizes the results in terms of equivalence for the main outcomes.

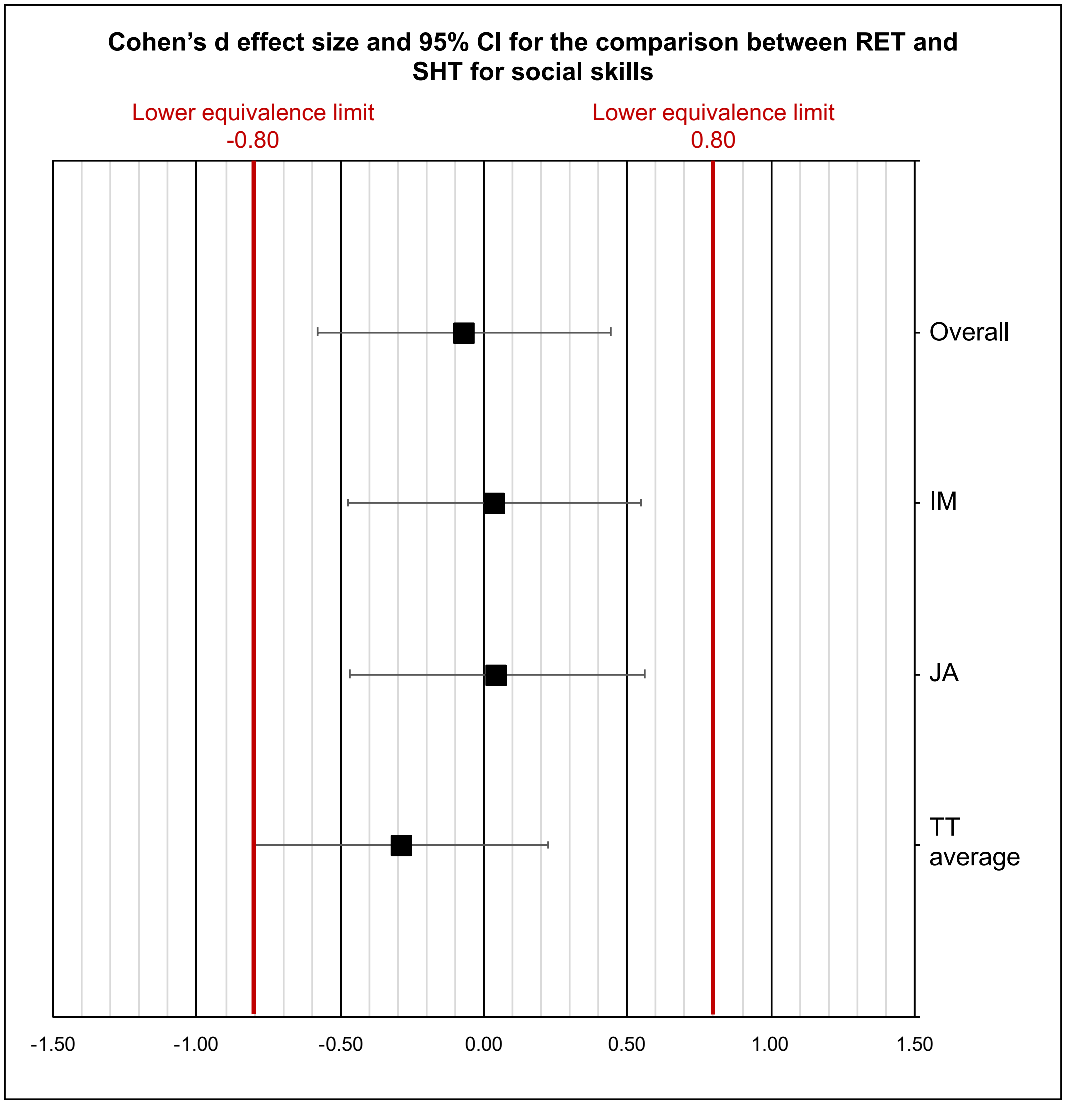

Figure 8. Cohen's *d* effect size for the difference between RET and SHT and 95% CI for targeted skills and equivalence limits at the final assessment. Higher values of the effect size favor RET.

Secondary outcomes

A similar analysis (mixed linear models) was carried for each secondary outcome, namely mutual gaze, smiling and sitting at the table, which are all indicators of engagement in the session. The analysis was conducted across all type of tasks (IM, JA or TT) which was included as a nested factor (within session) in the model. For each of these outcomes the results have largely favored the RET condition, with large effect sizes, pointing that the children have been much more engaged in the intervention when the interaction partner was the robotic agent. The results are similar / stable if we quantify engagement variables as the number of behaviors reflecting the expected behavior or as a ratio between the number of behaviors and the total duration of the session.

For mutual gaze the RET condition had an overall significant higher score compared to SHT, and this was reflected across almost all sessions (see Figure 9), but also across all types of task (see Figure 10).

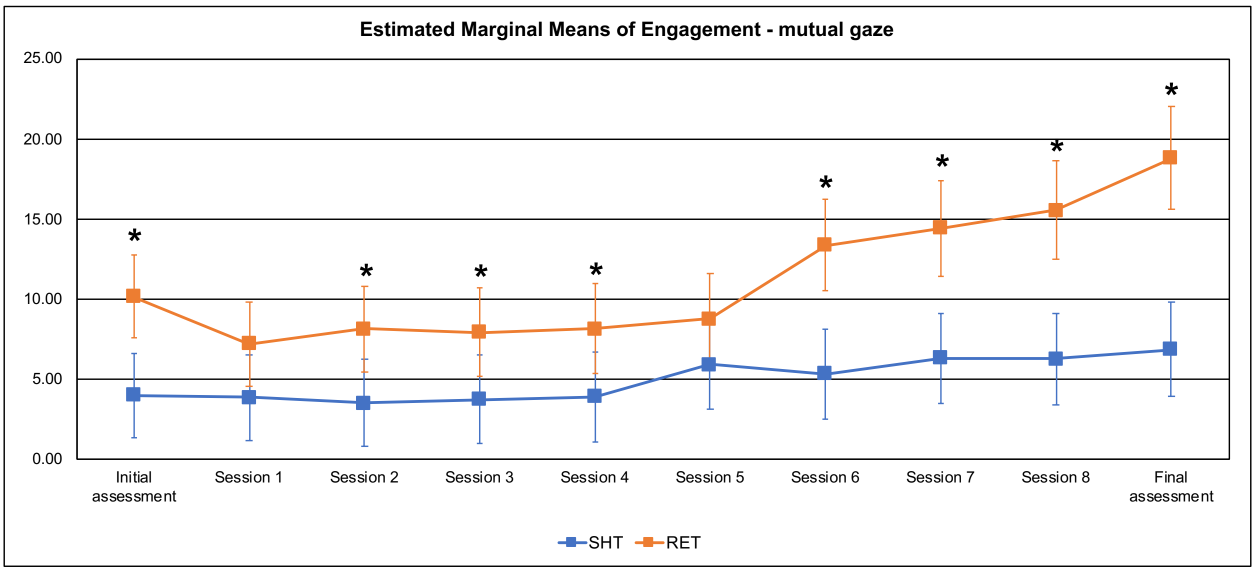

Figure 9. Estimated Marginal Means and 95% CI of TT skills across assessment and intervention sessions for mutual gaze. "*" marks statistically significant differences.

A similar pattern of results emerged also for **smiling**. The was a significant group effect which was reflected across all sessions, with the exception of the initial assessment (see Figure 11). Also, the groups were different (in favor of RET) across all tasks, except for JA (see Figure 12).

Finally, the results were similar for **sitting** at the table during the task. There was a group effect favoring RET, and the specific difference were significant across all sessions, except for baseline assessment (see Figure 13). The groups were also different across task type, but only for IM, TT patterns, and TT categories.

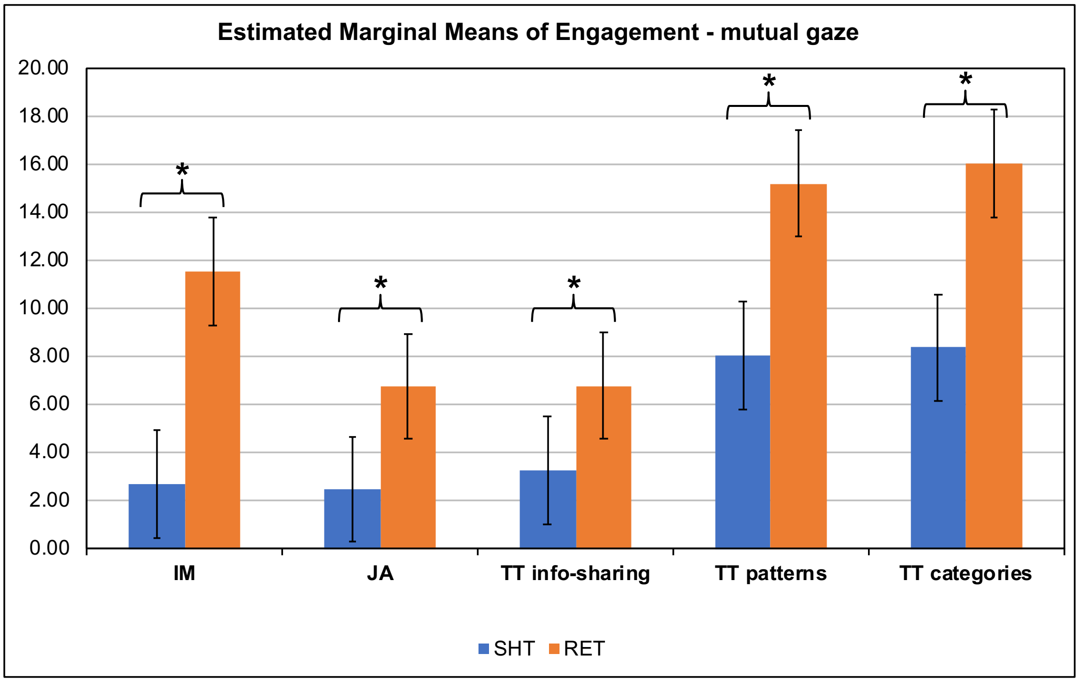

Figure 10. Estimated Marginal Means and 95% CI of TT skills across type of task for mutual gaze. "*" marks statistically significant differences.

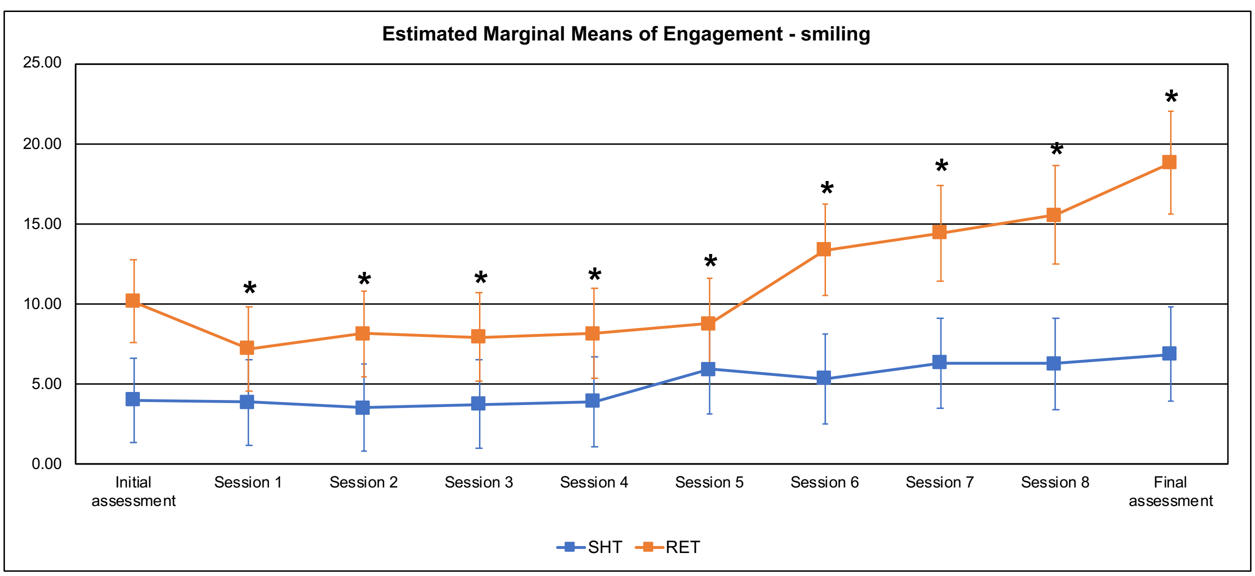

Figure 11. Estimated Marginal Means and 95% CI of TT skills across assessment and intervention sessions for smiling. "*" marks statistically significant differences.

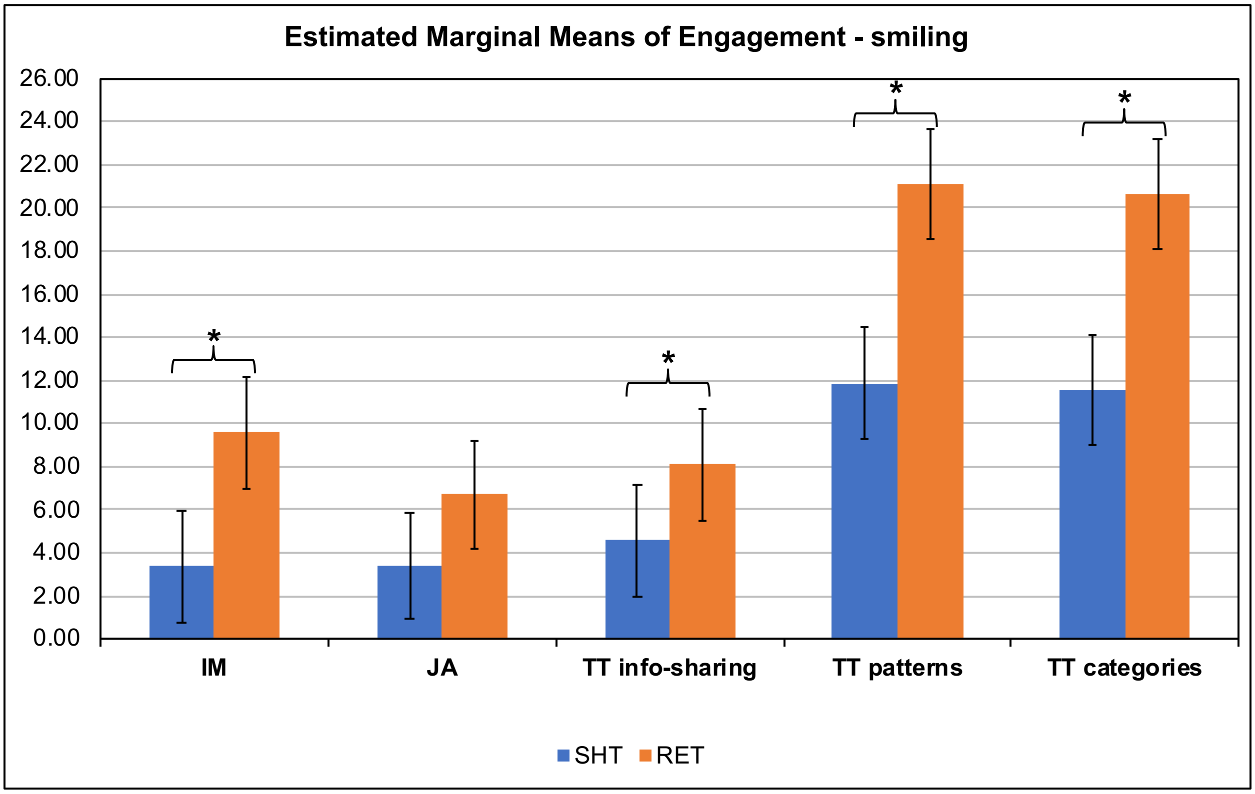

Figure 12. Estimated Marginal Means and 95% CI of TT skills across type of task for smiling. "*" marks statistically significant differences.

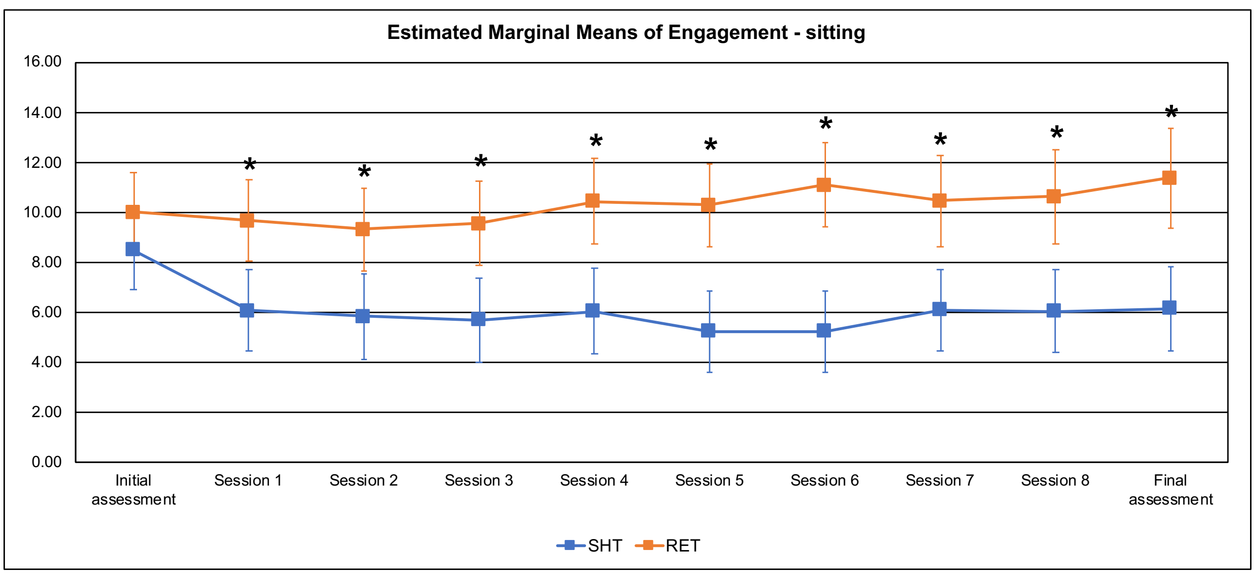

Figure 13. Estimated Marginal Means and 95% CI of TT skills across assessment and intervention sessions for smiling. "*" marks statistically significant differences.

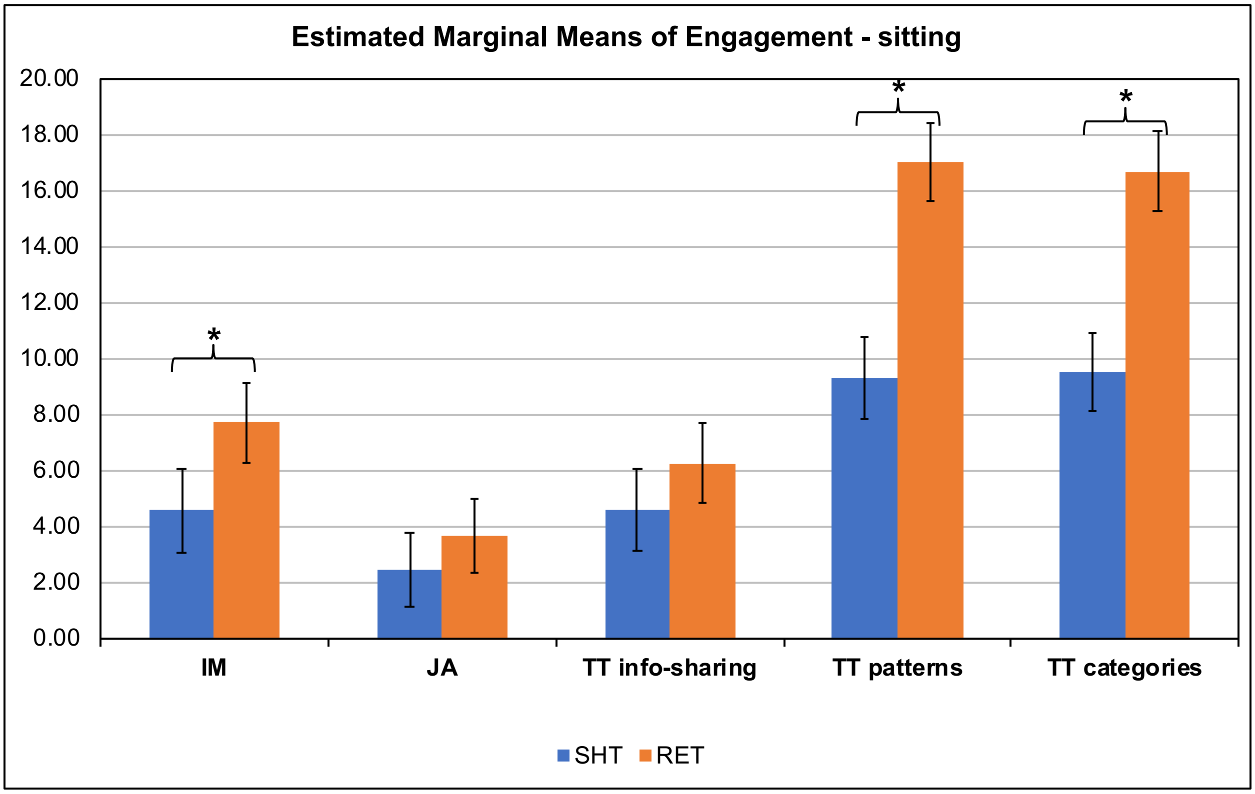

Figure 14. Estimated Marginal Means and 95% CI of TT skills across type of task for smiling. "*" marks statistically significant differences.

Equivalence analysis based on computed effect sizes (across all sessions) and their confidence intervals points that the RET condition is superior to the SHT condition on all secondary outcomes included in the analysis (see Figure 15).

In the previous studies conducted in this project (the single case series) we found that several children experienced higher levels of positive emotions compared to baseline or SHT conditions. Although this would be a relevant secondary outcome for our trial, it was not possible to code and analyze at this point the emotional reactivity express by the children during this study. However, the data is available in the recorded video and will be extracted and analyzed as a continuation of the activities of the project.

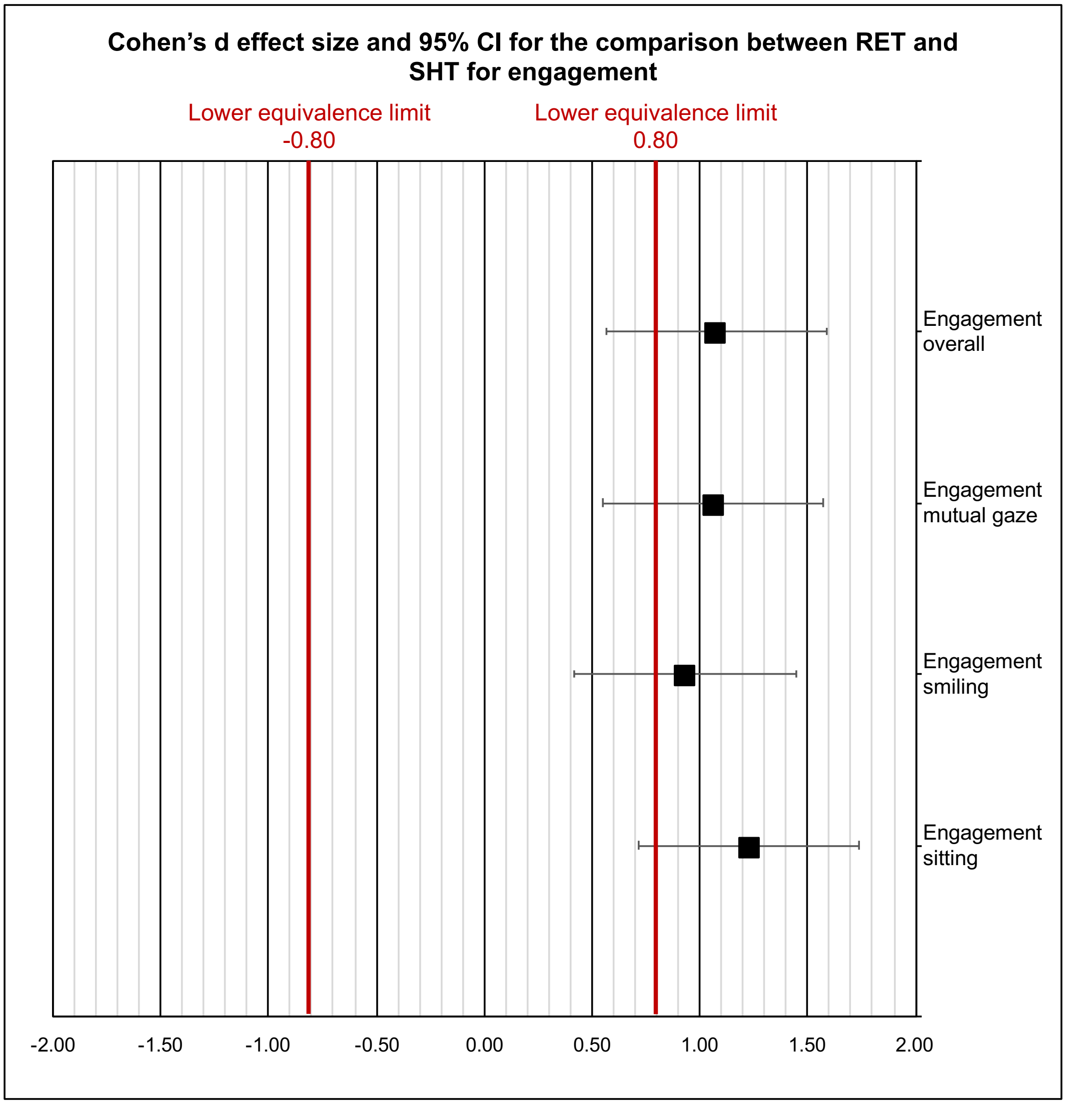

Figure 15. Cohen's *d* effect size for the difference between RET and SHT and 95% CI for engagement and equivalence limits across all sessions and tasks. Higher values of the effect size favor RET.

Standardized instruments

The analysis of the standardized instruments of autism symptoms, namely ADOS and SCQ revealed a significant time effect indicated that the scores are decreasing from pre-test to post-test across the overall sample. The interaction between time and group was not significant which points that the evolution of the two groups was similar on these scales. Figures 15 and 16 present these results.

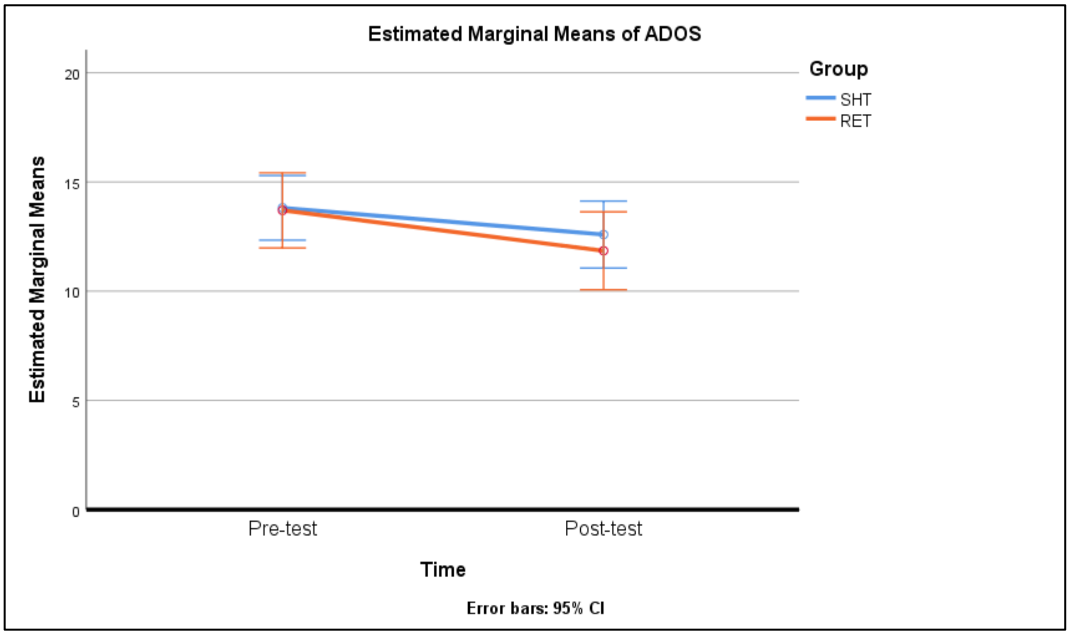

Figure 16. Estimated Marginal Means and 95% CI of ADOS total score at pre-test and post-test for SHT and RET.

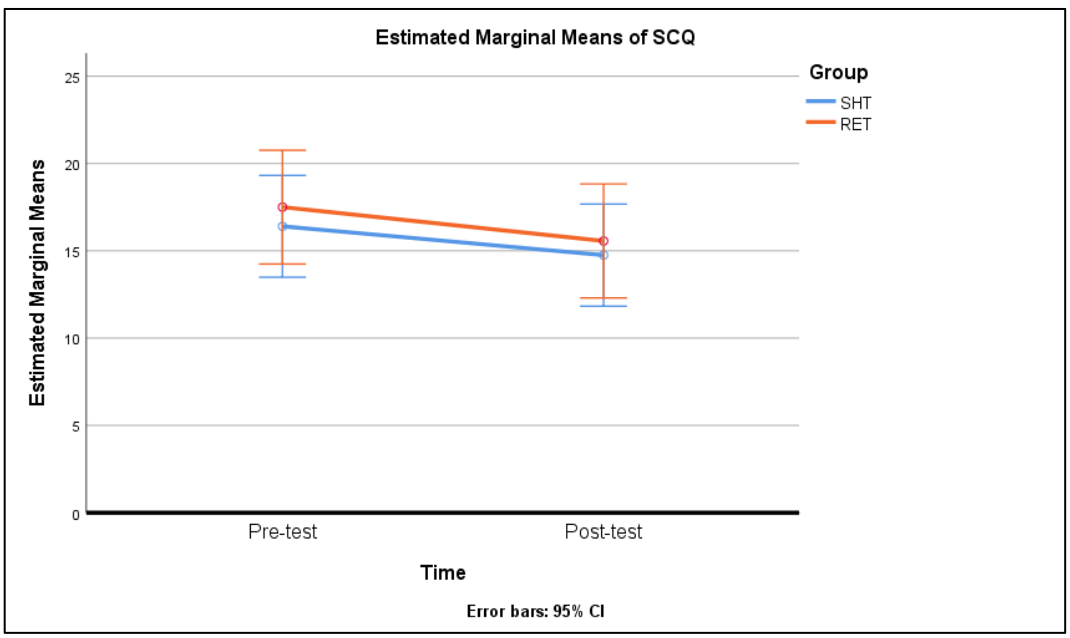

Figure 17. Estimated Marginal Means and 95% CI of SCQ total score at pre-test and post-test for SHT and RET.

Satisfaction about the intervention

As we asked the parents of the children that were included in the study how satisfied they were about several aspects of the intervention (e.g., satisfaction about the degree to which the intervention was personalized for the needs of the child, satisfaction related to the improvements in the target abilities), we compared the frequency of the average scores indicated a low vs. high degree of satisfaction. The two groups were similar in this regard, with a small percentage of parents in the RET condition reporting lower levels of satisfaction. Figure 18 below sumarizes this result.

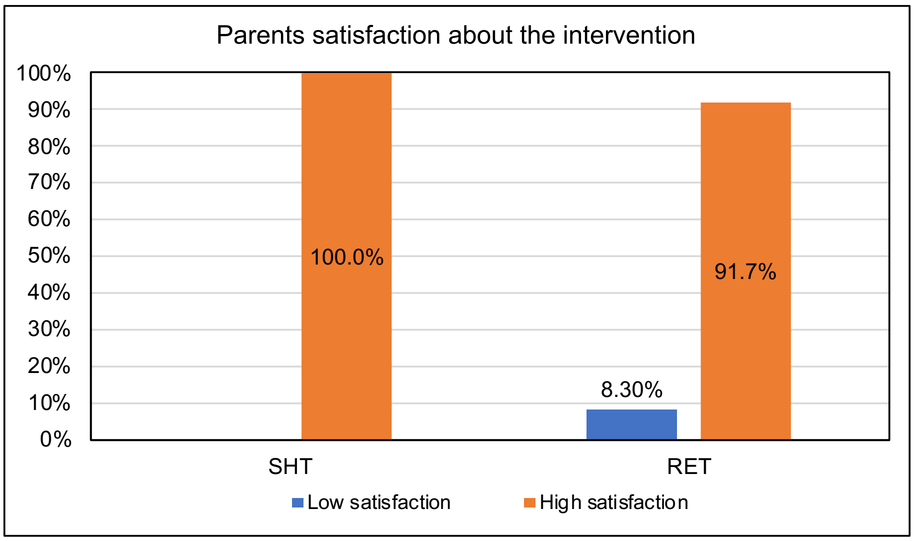

Figure 18. Frequency of parents reporting scores that indicate low and high levels of satisfaction.

Study 2

Methods

Design

In order to test the DREAM Lite solution, we conducted an effectiveness study implemented in 10 special education institutions and therapy centers for ASD children in Cluj-Napoca, Romania. The children were randomly allocated to one of two conditions: a RET intervention delivered via the DREAM Lite solution, or control group in which the children waited for two weeks before they receive the same intervention. All children were recruited from special education institutions and thus all children followed regular therapeutic and educational programs outside of the study protocol. The children were included if they had a history of diagnosed ASD symptoms and were able to understand the instructions required for assessment and for delivering the RET intervention.

Participants

A total of 79 participants (11 females) were included in this study, with an average age of 5.63 years. The number of participants randomly allocated in each group was 39 for control and 40 for RET. Two participants (one in each group did not receive the intervention because of reasons unrelated to the study. Twelve participants (7 in RET) were excluded after allocation because they showed difficulties in following the instructions or presented high levels of distress during the initial assessment or first intervention. Figure 19 presents the CONSORT diagram for the DREAL Lite effectiveness study.

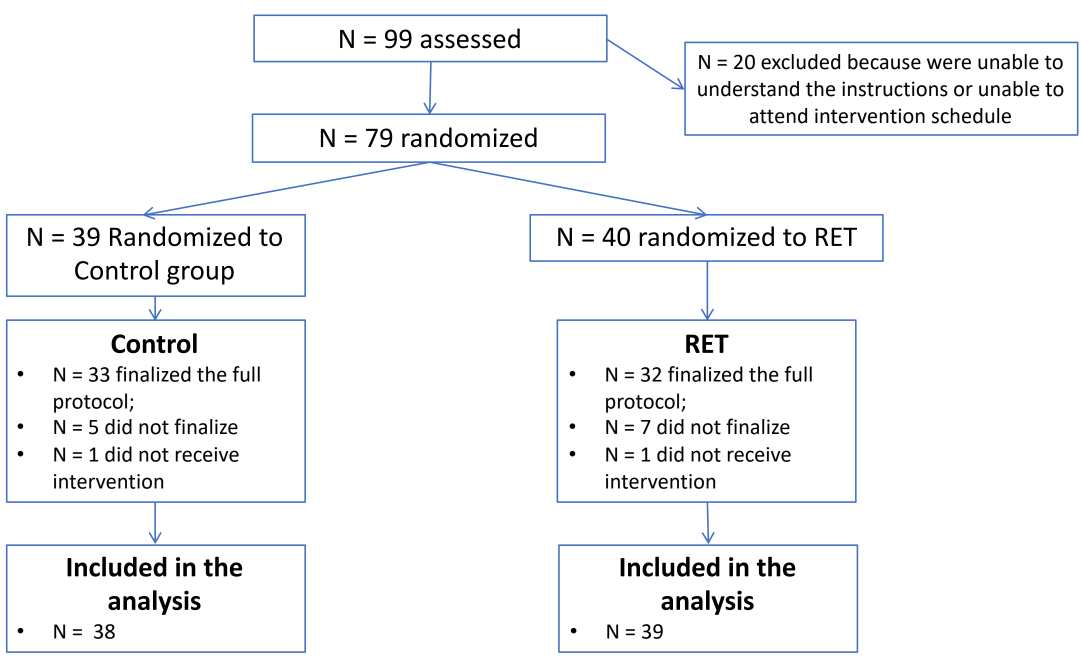

Figure 19. CONSORT diagram for the DREAM Lite study.

Procedure

The intervention consisted of five sessions: a behavioral assessment based on the Autism Diagnostic Observation Schedule-2 (ADOS-2) which was meant to evaluate the baseline of the children' social interaction skills; three intervention sessions of about 30 minutes each, performed daily or every second day, which targeted imitation, joint attention, and turn-taking skills; and a final evaluation session which was identical to the initial one. The intervention sessions followed the discrete trial training and the protocol described above form the clinical trial. The robot partner presented a discriminative stimulus in the form of an instruction (e.g., "do like me", for the imitation intervention), waited for the behavioral response of the child, and offered a contingent feedback (positive social reinforcement or an indication to correct the behavior). The control group was also assessed at the beginning of the wait period and just before starting the RAT intervention.

Measures

We measured the social interaction skills of the children based on the dedicated module form the ADOS-2, as well as the imitation, joint attention and turn-taking skills, based on behavioral tasks performed in interaction with a human partner. We also measured the performance in imitation, joint attention and turn-taking during each intervention session. To increase measurement accuracy, an external observer coded the performance during these sessions using a standardized observation grid. The performance of the child in each trial, across each task and each level of difficulty was coded as good or bad performance. We also video recorded each session and an experienced supervisor reviewed the initial coding. Finally, we used self-report questionnaires to assess parents' satisfaction with the intervention.

Statistical analysis

As for the clinical trial presented above, we looked at the effects of the two groups on targeted skills. However, given that the control condition did not receive any intervention as part of the study (although many of the child received standard treatment as part of their usual treatment) and the fact that the initial and final assessments in the two groups were conducted only by the therapist (and thus, the assessment during the intervention with the robot is not equivalent to the initial and final assessment), we could not analyses these skills from pre-test to post test, by taking to account the evolution across the intervention sessions. Thus, we first looked for changes between baseline and after the wait period (for the control condition) and between baseline and after the intervention (for the RET condition). Changes across intervention sessions were introduced in a separate analysis by combining the control group (after the waiting time; they also received the same intervention after the waiting time) and the RET group.

As an analytical strategy, we used a linear mixed model approach for skills measurements. The average scores for each skill in each session were used in the analysis. For standardized instruments, we used a mixed within-between analysis of variance, with time of assessment (prevs. post-test) as the within-subject factor, and the group of intervention as the between-subject factor.

Results

Between groups comparisons on targeted abilities

For IM, we found a significant time effect and a significant interaction between time and group, indicating the evolution in the two groups is different. However, there was no clear difference between the two groups any of the two points, although the RET condition had lower scores at pre-test and a significant and positive slope for the change in scores. To better understand the interaction effect, we used baseline scores as a covariate and found that the estimated marginal means at post-test were significantly higher for RET than for the control group. These results are depicted in Figure 20 bellow.

For JA, we found a significant time effect, but no interaction effect, indicating that the changes in the two groups were similar. The contrast between the initial and the final assessment indicated that scores have increased across the overall sample. These results are depicted in Figure 21.

The results were similar **for TT**, with a significant time effect, but no interaction effect, suggesting that the two groups had a comparable evolution. The scores also increased from the initial to the final assessment (Figure 22).

The fact that the time effect was significant across all outcomes, suggest that both groups have improved. The fact that the control groups experienced increases in performance scores is not unexpected given that most of the children are in fact doing some form of psychological interventions which likely to targets social skills. Also, the improvement in scores might as well be attributed to a repetition effect, the assessment having also a training effect.

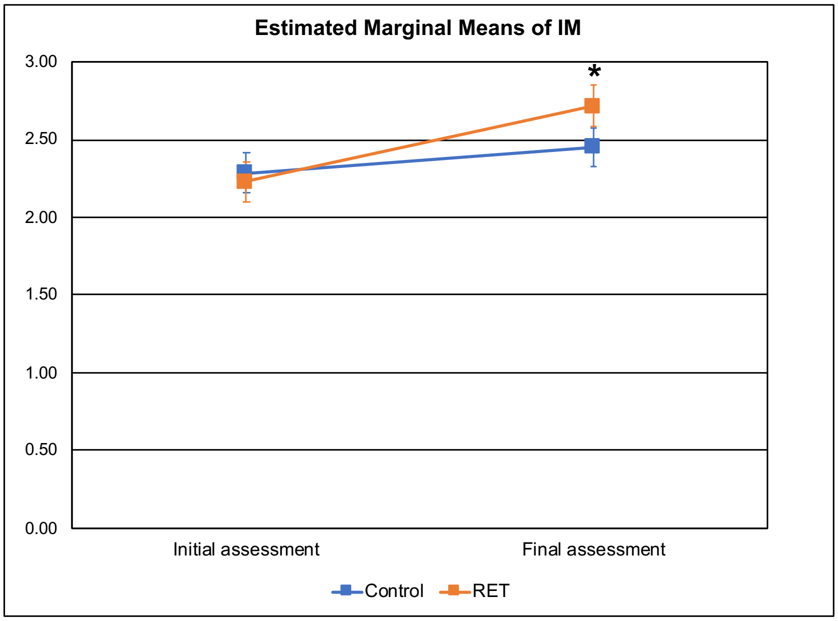

Figure 20. Pre- to post-test changes in IM scores across the two groups. "*" marks statistically significant differences

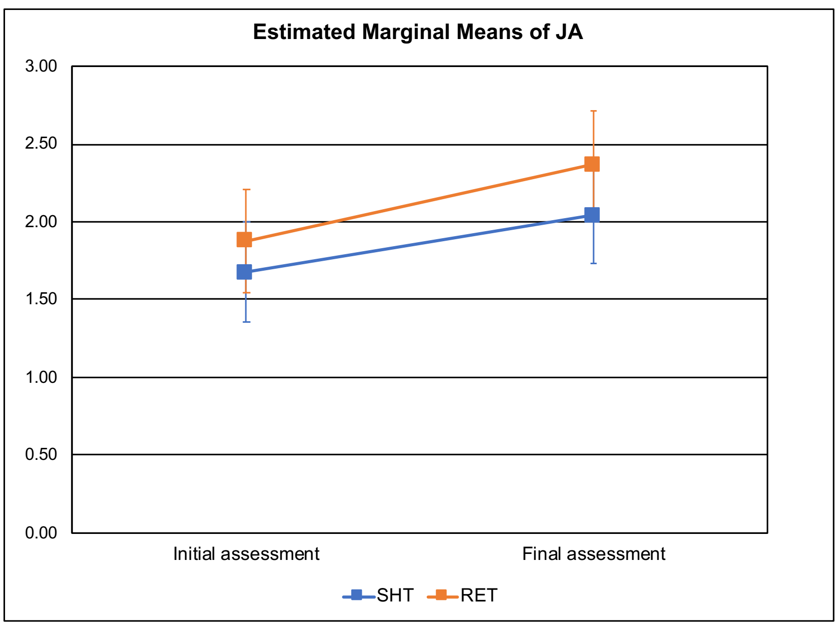

Figure 21. Pre- to post-test changes in JA scores across the two groups.

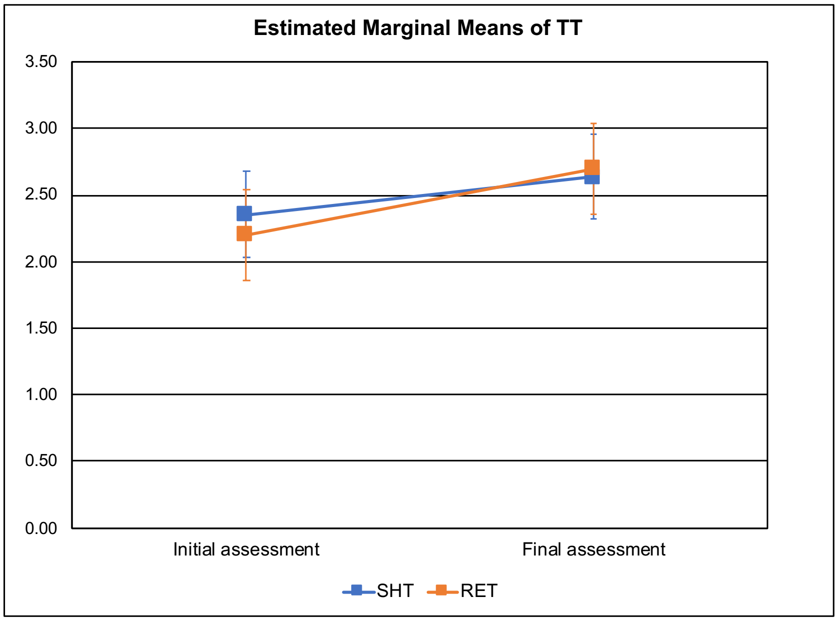

Figure 22. Pre- to post-test changes in TT scores across the two groups.

Within group changes during RET intervention

Changes during the intervention sessions with NAO indicated that the children improved their scores across both IM (see Figure 23) and JA (see Figure 24) where there were statistically significant differences between the initial and the final interventions session. For TT however (see Figure 25), the scores improved form the first session to the second one but decreased afterwards in the thirds session to a level that was no longer different from the first one.

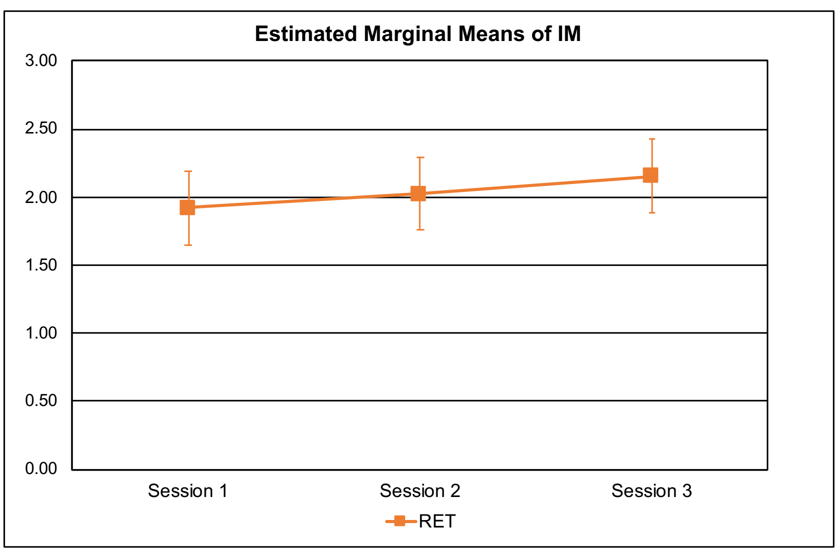

Figure 23. Changes in IM scores during RET intervention.

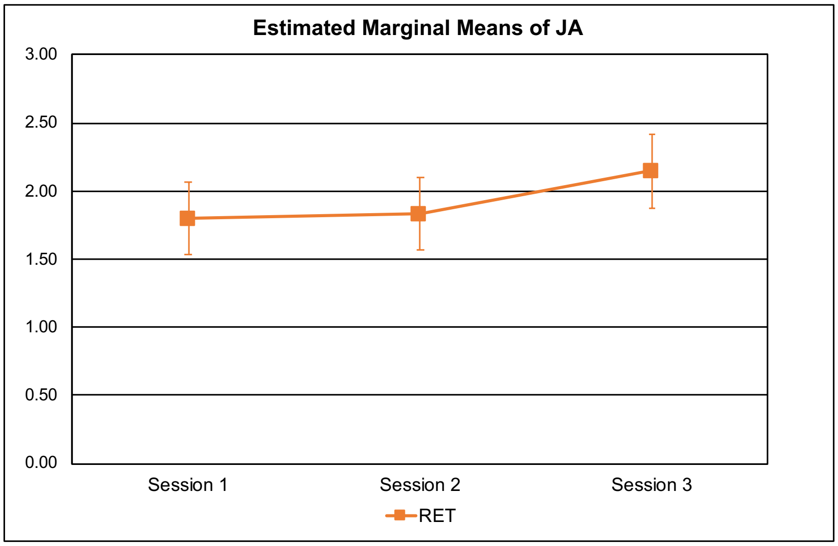

Figure 24. Changes in JA scores during RET intervention.

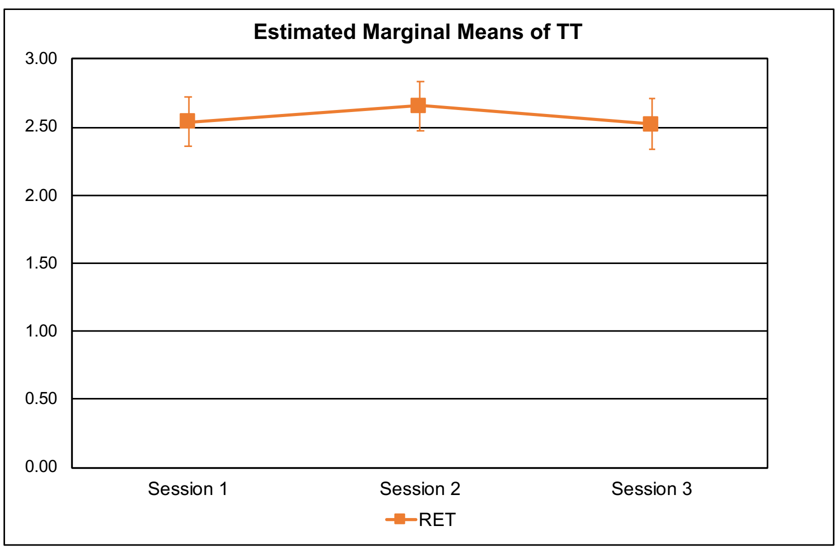

Figure 25. Changes in TT scores during RET intervention.

Changes on parents and teachers reports

We asked both parents and teachers (or in some cases the therapist working on a regular basis with the child) to report on children social skills at the initial and final assessments in both groups. Both parents and teachers were asked to fill the social interaction sub-scale form the TRIAD (see Figures 26 and 27). We compared pre- to post-test changes in the two group and the results showed significant increases on both assessments. However, for the RET condition had significantly higher scores at post-test for parents' assessment, compared to the control condition. The same effect was not present for teachers' assessment.

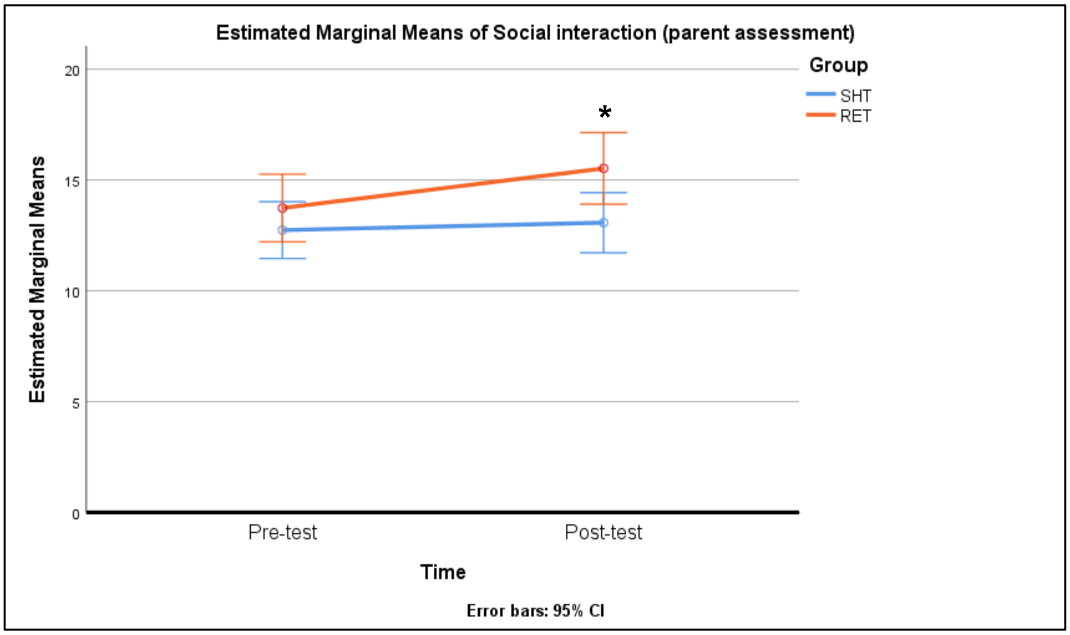

Figure 26. Changes in social interaction as assessed by parents. "*" marks statistically significant differences.

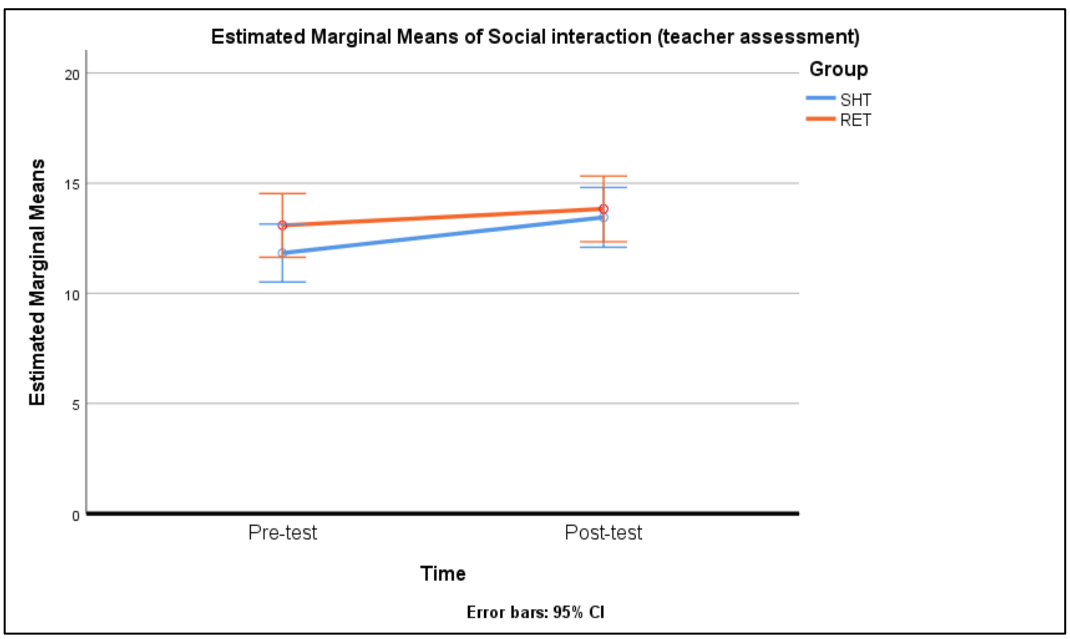

Figure 27. Changes in social interaction as assessed by teachers. "*" marks statistically significant differences.

Parents satisfaction with the RET intervention

We asked also asked the parents to report on their level of satisfaction about the RET intervention, and similar to the trail the great majority of the parents reported scores that indicated a high level of satisfaction (see Figure 28).

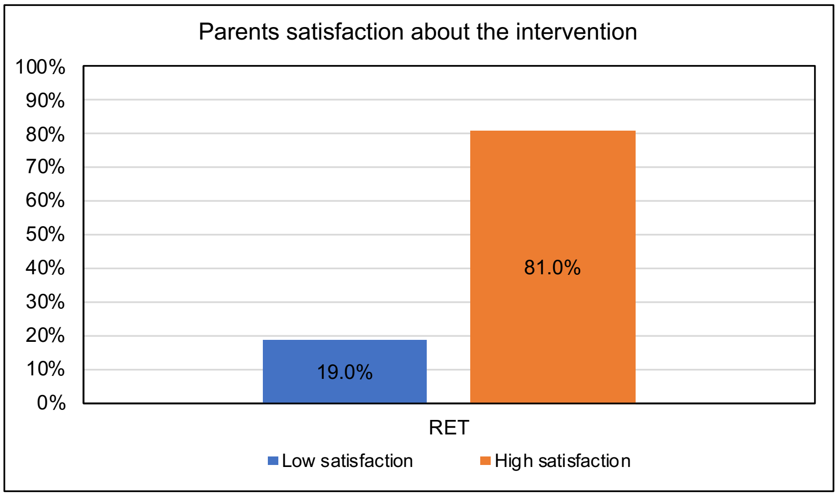

Figure 28. Frequency of parents reporting scores that indicate low and high levels of satisfaction about the RET intervention.

Conclusion

The final results in the clinical trial indicated that both groups had an overall a positive impact on the targeted outcomes, with significant gains on IM and TT, and to some extend on JA, but also on standardized measures (i.e., ADOS and SCQ). The comparisons between the two groups at the end of the interventions showed no significant differences, and the average effect sizes across all social skills were in the equivalence margin that we established *a priori*. For secondary outcomes however, there was a clear advantage for the RET condition, with effect sizes that crossed the equivalence margin pointing significant and large differences between the two groups. These results are a major addition to the research field, this study being the first randomized clinical trial that compared RET with standard intervention, contrasting their impact on several outcomes, from ASD symptoms (measured by standardized instruments), psychological mechanisms involved in ASD (social skills) and process variables related to the intervention (engagement in the task). These results bring RET in the realm of evidence-based interventions for ASD children.

In relation to the DREAM Lite solution, our results were very promising, despite the fact that it was brief, and it was delivered to a heterogonous and severely impaired sample. More research is needed to assess its long-term effects, but our results support the idea that this simplified implementation of the DREAM framework could be used easily to disseminate RET interventions for ASD children in real-life context. DREAM Lite could be used as an additional tool to standard therapy to enhance the learning of social skills in ASD children.

References

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: American Psychiatric Association.

Cabibihan, J. J., Javed, H., Ang, M., & Aljunied, S. M. (2013). Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. *International journal of social robotics*, *5*(4), 593-618. doi: 10.1007/s12369-013-0202-2.

David, D., Matu, S. A., & David, O. A. (2014). Robot-based psychotherapy: Concepts development, state of the art, and new directions. *International Journal of Cognitive Therapy*, *7*(2), 192-210. doi: 10.1521/ijct.2014.7.2.192.

Dawson, G., Toth, K., Abbott, R., Osterling, J., Munson, J., Estes, A., & Liaw, J. (2004). Early social attention impairments in autism: social orienting, joint attention, and attention to distress. *Developmental psychology*, *40*(2), 271. doi: 10.1037/0012-1649.40.2.271.

Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., ... & Rutter, M. (2000). The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of autism and developmental disorders, 30*(3), 205-223. doi: 10.1023/A:1005592401947.

Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S., & Pioggia, G. (2016). Autism and social robotics: A systematic review. *Autism Research*, *9*(2), 165-183. doi: 10.1002/aur.1527.

Peters-Scheffer, N., Didden, R., Korzilius, H., & Sturmey, P. (2011). A meta-analytic study on the effectiveness of comprehensive ABA-based early intervention programs for children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *5*(1), 60-69. doi: 10.1016/j.rasd.2010.03.011.

Ricks, D. J., & Colton, M. B. (2010, May). Trends and considerations in robot-assisted autism therapy. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on* (pp. 4354-4359). IEEE.

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*. Western Psychological Services.

Thill, S., Pop, C. A., Belpaeme, T., Ziemke, T., & Vanderborght, B. (2012). Robot-assisted therapy for autism spectrum disorders with (partially) autonomous control: Challenges and outlook. *Paladyn, 3(4)*, 209-217. doi: 10.2478/s13230-013-0107-7.

Virués-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood: Meta-analysis, meta-regression and dose–response meta-analysis of multiple outcomes. *Clinical psychology review*, *30*(4), 387-399. doi: 10.1016/j.cpr.2010.01.008.