HUMAN-ROBOT INTERACTION

Efficacy and effectiveness of robot-assisted therapy for autism spectrum disorder: From lab to reality

Daniel David[1,2], Paul Baxter[3], Tony Belpaeme[4], Erik Billing[5], Haibin Cai[6], Hoang-Long Cao[7,8], Anamaria Ciocan[1], Cristina Costescu[1], Daniel Hernandez Garcia[9], Pablo Gómez Esteban[7], James Kennedy[10]†, Honghai Liu[11], Silviu Matu[1], Alexandre Mazel[12,13], Mihaela Selescu[1], Emmanuel Senft[14], Serge Thill[15], Bram Vanderborght[7], David Vernon[16], Tom Ziemke[17]*

The use of social robots in therapy for children with autism has been explored for more than 20 years, but there still is limited clinical evidence. The work presented here provides a systematic approach to evaluating both efficacy and effectiveness, bridging the gap between theory and practice by targeting joint attention, imitation, and turn-taking as core developmental mechanisms that can make a difference in autism interventions. We present two randomized clinical trials with different robot-assisted therapy implementations aimed at young children. The first is an efficacy trial (n = 69; mean age = 4.4 years) showing that 12 biweekly sessions of in-clinic robot-assisted therapy achieve equivalent outcomes to conventional treatment but with a significant increase in the patients' engagement. The second trial (n = 63; mean age = 5.9 years) evaluates the effectiveness in real-world settings by substituting the clinical setup with a simpler one for use in schools or homes. Over the course of a modest dosage of five sessions, we show equivalent outcomes to standard treatment. Both efficacy and effectiveness trials lend further credibility to the beneficial role that social robots can play in autism therapy while also highlighting the potential advantages of portable and cost-effective setups.

INTRODUCTION

This paper reports on two large-scale clinical trials comparing robot-assisted therapy (RAT) for children with autism spectrum disorder (ASD) with current standard psychological interventions, in both formal (clinical) and informal (educational) settings. The use of the robots in the treatment of ASD has been investigated by researchers for about 25 years (*1*–*5*), and the general idea is both evolutionary and challenging. Considering autistic children's general interest in technology and the fact that interaction with social robots is considered safe and enjoyable (*6*, *7*), it has been argued by several authors that the involvement of social robots in psychological interventions can lead to better results and cost-effective interventions in achieving therapeutic and educational objectives (*5*, *8*, *9*). However, it is challenging to include social robots in the therapeutical process because good collaboration among end users, psychologists, teachers, and researchers is required to design and build appropriate and beneficial interventions for children with autism. The general idea of RAT has instilled high hopes, not least in parents, for improving the current services for children with autism. However, when we analyzed the effectiveness of social robots in autism therapy reported in randomized clinical trials, the results were mixed, meaning that both positive and negative outcomes have been reported [for recent reviews, see (*5*, *10*, *11*)]. Although there have been some trials with positive results, there is certainly a need for trials that better integrate robot-assisted activities in the therapeutical process, explore the core developmental mechanisms, integrate them in educational contexts, and explore long-term results (*12*). Without sufficient data, there is a risk of instilling false hopes and causing unjustified costs for patients and their families. A recent meta-analysis on the use of social robots for autistic children (*5*) suggested that more trials are needed to identify specific characteristics of RAT (e.g., types of robots, clear and transparent sample selection, and settings) and better integration techniques in the therapeutic process. This paper contributes to filling that gap.

Most previous studies on RAT for children with ASD are proof-of-concept studies, clinical cases/single-case experiments, exploratory studies with small samples, experimental rather than clinical interventions, and often not guided by the evidence-based psychotherapy framework targeting established key mechanisms, such as joint attention, attention, and turn taking (*13*). Moreover, the interventions investigated in these studies target a very broad range of different abilities (e.g., play skills, social interaction abilities, and engagement), which are typically defined differently in different studies, leading to inconsistent results that are not easy to integrate. These inconsistencies and the lack of standardized instruments in measuring the outcomes constitute a substantial challenge for practitioners trying to assess the applicability and effectiveness of RAT in specific cases. Although there are several studies with standardized protocols, large samples of children, and standardized measurements [e.g., (*12*, *14*, *15*)], there is still a shortage of studies rigorously comparing RAT with conventional therapy for children with ASD (in the following, we refer to the latter as "standard human therapy" or SHT hereafter; cf. below). Moreover, as a recent systematic review pointed out, "while significant efforts have been made to develop

[1]Department of Clinical Psychology and Psychotherapy, Babeş-Bolyai University, Cluj-Napoca, Romania. [2]Icahn School of Medicine at Mount Sinai, New York, NY, USA. [3]School of Computer Science, University of Lincoln, Lincoln, UK. [4]IDLab-AIRO, Ghent University & imec, Ghent, Belgium. [5]School of Informatics, University of Skövde, Skövde, Sweden. [6]Department of Computer Science, Loughborough University, Loughborough, UK. [7]Brubotics, Vrije Universiteit Brussel & imec, Brussels, Belgium. [8]College of Engineering, Can Tho University, Can Tho City, Vietnam. [9]School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, UK. [10]Centre for Robotics and Neural Systems, Plymouth University, Plymouth, UK. [11]School of Computing, University of Portsmouth, Portsmouth, UK. [12]Human Machine Interaction, Paris, France. [13]Softbank Robotics, Paris, France. [14]IDIAP Research Institute, Martigny, Switzerland. [15]Donders Institute for Brain, Cognition & Behaviour, Radboud University, Nijmegen, Netherlands. [16]Carnegie Mellon University Africa, Kigali, Rwanda. [17]Department of Computer & Information Science, Linköping University, Linköping, Sweden.
†Present address: Apple Inc., Pittsburgh, PA, USA.
*Corresponding author. Email: tom.ziemke@liu.se

and test the technical aspects, clinical validation and implementation lag behind" (*11*). A recent study by Ghiglino and colleagues (*16*) tested the effects of robot-assisted activities focusing on visuospatial perspective taking integrated into the rehabilitation program offered in a health care institution, and their results indicate that children with ASD gain the most benefit from activities incorporated into interactions with a humanoid robot compared with standard therapy or other interventions.

Despite the differences observed between conventional treatment and RAT in some previous studies, the clinical meaningfulness of these findings is limited in the sense that the core developmental mechanisms of change still remain unknown (*14*). Hence, there is still a lot of uncertainty regarding how RAT really compares with SHT. We addressed some of the limitations in this domain in a large European project called DREAM (Development of Robot-Enhanced Therapy for Children with Autism Spectrum Disorders) through both clinical and technological innovations [see also (*9*, *17*)], which have been integrated and evaluated in the two clinical studies reported in this paper. One of the strengths of our work presented here, we believe, lies in the conceptualization of RAT in accordance with the evidence-based psychotherapy paradigm in clinical psychology, meaning that we used both clinical and psychometric methods for a rigorous clinical assessment of ASD. Previous studies focused either on clinical symptoms of ASD or only basic skills. Furthermore, some recent studies tested the effectiveness of social robots in home-based therapy, an environment that is slightly different from our approach (*18*, *19*). We fully support the general idea of using social robots in home environments in the future, but we also strongly believe that, at this point, more research is needed to investigate the underlying mechanisms of RAT to be used in home settings where other types of challenges can appear (*20*). Central to our work is a distinction between the clinical symptoms of ASD (as documented in DSM-5, the *Diagnostic and Statistical Manual of Mental Disorders*, fifth edition) and the underlying mechanisms. Our hypothesis, based on previous work [e.g., (*21*)], is that imitation, joint attention, and turn-taking are crucial, fundamental abilities in children with ASD. However, no previous study investigating the use of social robots has treated these abilities as the main targets. These fundamental abilities, we believe, also constitute a base for developing more complex skills, such as play and social interactions, and they contribute to the cognitive and social development of children and predict their later communication and intellectual functioning.

Some previous studies have investigated the robot-mediated response to joint attention in a small sample (*21*) of young children with ASD (*22*), with small but not significant group differences being observed when using a Nao robot to elicit joint attention responses. The study took place in a clinical setting, and all participants underwent four sessions of 10 min each. Two recent studies implemented by the same research group investigated the effectiveness of pivotal response treatment with or without robot assistance for improving social skills in children with ASD (*12*, *23*). One of the studies investigated the effectiveness of robot-assisted pivotal response treatment compared with pivotal response treatment in improving general social communication skills and self-initiations measured with both standardized questionnaires and observational grids (*23*). Fifty-two children were assigned to one of the groups and participated in 20 intervention sessions in which self-initiations were targeted using motivational techniques from pivotal response treatment. Their results showed no significant differences between the two groups, although both interventions were effective in improving children's general socio-communication skills. No significant associations were found between the self-initiations and overall communication skills. Therefore, future studies are needed to further investigate the prerequisites and underlying mechanics of the social deficits in ASD. Similar results were identified when comparing robot-assisted pivotal response treatment with treatment as usual (*12*) in a study involving 73 children with ASD. Both studies were conducted in clinical settings and limited their intervention protocols to self-initiations (e.g., asking for objects and asking for help by using utterances).

Another important issue that we try to address in our studies is the generalizability of the results, given that our final goal is to enhance social skills of autistic children in real-world human-human social interactions (*3*, *24*). Considering that the focus of our intervention is on hypothesized underlying mechanisms (imitation, joint attention, and turn-taking), which support social skills acquisition (i.e., learning directly various social/communication behaviors), rather than on clinical symptoms as such, we are conceptualizing the intervention as targeting core developmental mechanisms rather than symptoms. We believe that this can support flexibility and generalizability in learning new social communication behaviors. Consequently, the intervention here is not conceptualized as a series of experimental sessions but as a minimum clinical sequence, which could then be generalized to a full long-term or intense treatment. On that account, we expect to identify differences in the clinical symptoms, although our approach is not a direct and long-term one. Another strength of our approach is that we compare RAT and SHT, both in an efficacy study (i.e., in well-controlled lab conditions) and in an effectiveness study (in a school setting, i.e., in more ecological, real-world conditions), in line with the most rigorous conditions of evidence-based therapy. The work presented here—documenting two complementary clinical studies and addressing efficacy and effectiveness, respectively—contributes to the integration of both these needs: a highly rigorous methodology and effectiveness in real-world settings.

Furthermore, considering the approach to target core developmental mechanisms, we aim to clarify the active elements from the interventions that could help children with autism improve their abilities and reduce their social difficulties. We propose an approach that allows us to identify and confirm the hypothesized underlying mechanisms (imitation, joint attention, and turn-taking) of change in a controlled efficacy study and subsequently to maximize those active elements in an effectiveness study. This approach allows us to clarify what abilities are most important to train in RAT.

In this process, the robot is used as a mediator. As previously mentioned, we strongly believe that if social robots are to be part of the therapeutic process, they need to be adopted by practitioners in their daily activities, and practitioners should be enabled to actively participate in the design of the robot-assisted activities (*25*). The mechanisms and skills learned in interaction with the robot should be generalizable to human-human interactions with the help of the therapist and should not remain limited to human-robot interactions only.

In the development of our studies, we followed a participatory design with the following steps: First, the initial concepts were developed by the psychologists and engineers, after which feedback from teachers, parents, and children was gathered. The tasks in pilot studies [see (*26*, *27*)] were then tested, refining the prototypes on the basis of the feedback and results from these pilot studies. Last, the intervention program was adapted and scaled to include more children.

RESULTS

Our approach moved from a Wizard-of-Oz paradigm, in which the robot was controlled by a human operator, to a semi-autonomous one, in which the robot's autonomous decision-making was supervised by a human therapist. Our platform for delivering semi-autonomous RAT for children with ASD has been made available as an open-source and reusable platform (*17*).

More specifically, to carry out the studies reported here, we developed two technical platforms (see Fig. 1): a stationary system used in controlled clinical trials to evaluate the efficacy of RAT (hereafter referred to as the clinical trials platform) and a smaller, more portable system that was used to evaluate the effectiveness of RAT in real-life educational settings (the educational platform). The clinical trials platform (*9*, *17*, *26*, *27*) consisted of a sensorized table (Fig. 1A) on which the robot was placed, equipped with several cameras that recorded aspects such as the pose, movement, and facial expressions of the child. This sensory information was used by the system to assess whether the behavior of the child was in accordance with the expectations of the clinical intervention scripts. That information, in turn, was used to automatically decide the next actions of the robot, thereby enabling the robot's semi-autonomous behavior. However, therapists always retained the ability to override the system's decisions. The educational platform (*28*), on the other hand, omitted the sensorized table, and the robot was instead controlled via a tablet (Fig. 1B). This made the educational platform more portable, but all clinically relevant attributes were retained, as discussed in more detail below.

Study 1: Efficacy of RAT

Intent-to-treat analysis on all outcomes indicated that the two interventions (RAT versus SHT) were equivalent on all primary outcomes and on clinical outcomes. However, the RAT intervention was superior on several secondary outcomes, reflecting engagement with the intervention (see Fig. 2). Statistical analysis and results are described in more detail in the Supplementary Materials.

SHT was structured as a behavioral teaching program rooted in applied behavior analysis (ABA) discrete trial training methods, an approach well established in autism interventions in Romania. The treatment directly targeted core social cognitive domains—imitation, joint attention, and turn-taking—ensuring alignment with the RAT condition. Sessions consisted of structured play tasks facilitated by a human therapist, where learning followed the discriminative stimulus → child response → contingent reinforcement format. Correct responses were reinforced through positive social feedback (e.g., praise, smiles, and gestures), whereas incorrect or no responses were supported through shaping, modeling, and prompting. These skills were practiced through repetitive, reinforcement-based exercises embedded in natural play contexts to promote consolidation, motivation, and engagement. The structure and dosage of sessions

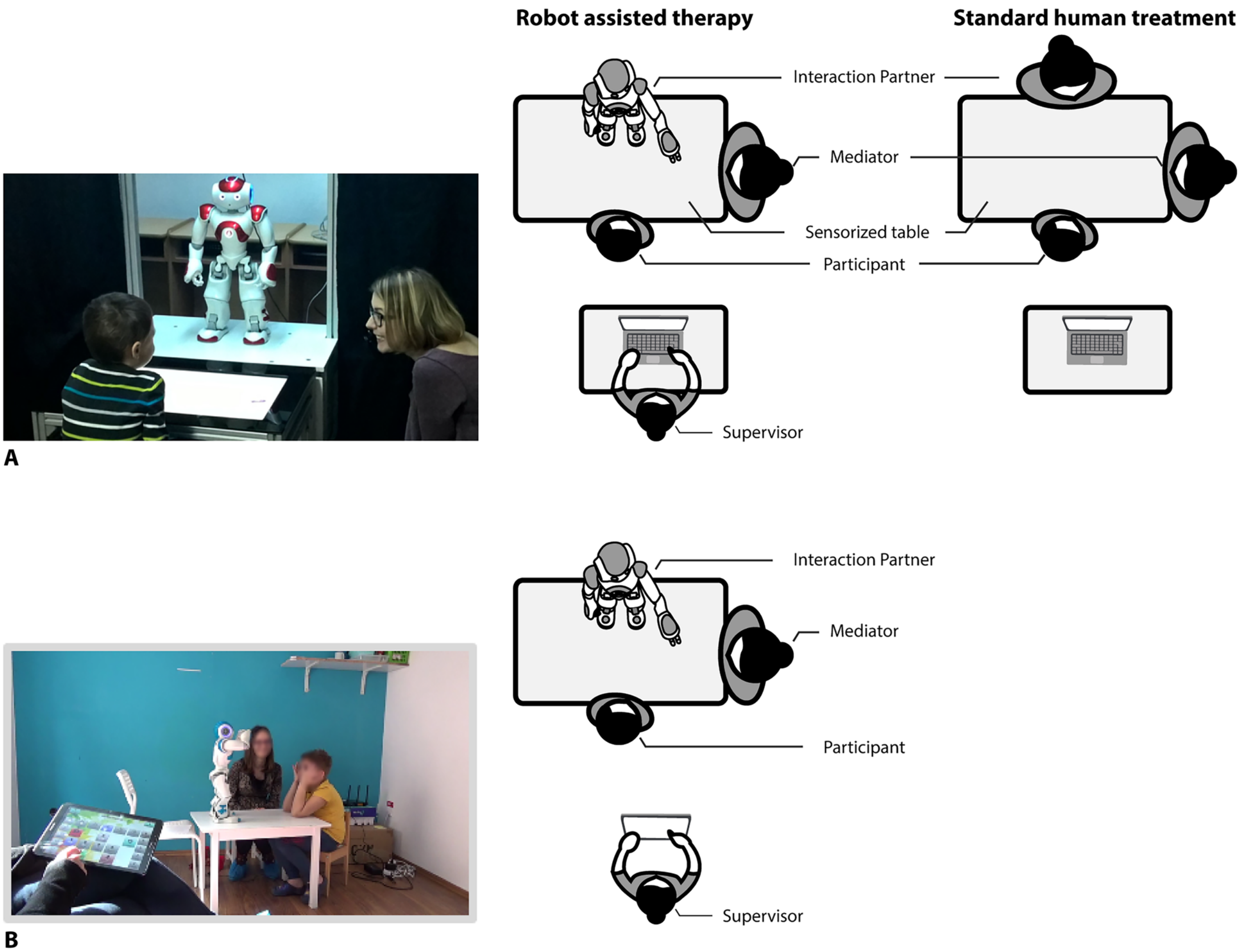

Fig. 1. RAT research platforms. (**A**) Clinical trials platform used in study 1. (**B**) Educational platform used in study 2.

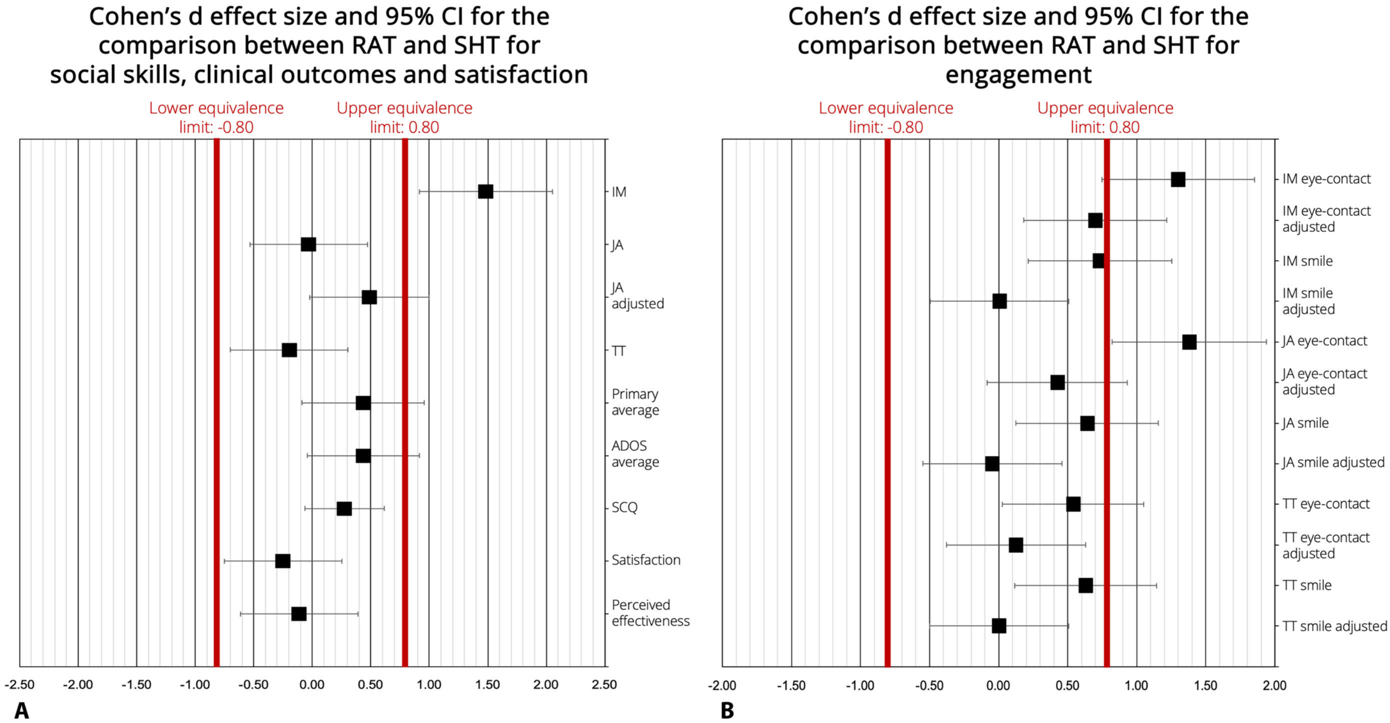

Fig. 2. Study 1 results: Efficacy of RAT using the clinical trials platform—primary outcomes (social skills, clinical outcomes, satisfaction, and perceived effectiveness) and secondary outcomes (eye contact and smiling, i.e., engagement). The comparison of RAT with SHT on social skills, clinical outcomes, and satisfaction (**A**) using Cohen's *d* effect sizes and their 95% confidence intervals indicated a significant difference between RAT and SHT for imitation, $d = 1.48$, 95% confidence interval (0.91; 2.05). All other assessed criteria fell within equivalence limits. The comparison of RAT with SHT on engagement (**B**) using Cohen's *d* effect sizes and their 95% confidence intervals indicated a significant difference between RAT and SHT on eye contact in the joint-attention task when the scores were not adjusted for baseline differences, $d = 1.38$, 95% confidence interval (0.83; 1.93). All other assessed criteria fell within equivalence limits. Cohen's *d* effect sizes for all comparisons were computed using the estimated marginal means and their standard errors from a hierarchical linear model.

(12 in total, with evaluation, training, and postevaluation phases; each lasting 45 min) were intentionally matched to the RAT condition, ensuring comparability in treatment intensity and exposure.

Regarding imitation, our results revealed that children from the RAT group significantly improved their imitation ability compared with those from the SHT group [$F_{1,84.57} = 23.44$, $P < 0.001$]. Also, the results showed that they significantly improved their performance from initial assessment to final assessment [$F_{9,110.95} = 5.38$, $P < 0.001$], and they improved their performance with the increase in difficulty from level 1 to level 2 [$F_{2,84.81} = 4.03$, $P = 0.021$]. Comparisons for the group and time interaction indicate that the scores for the participants in the SHT group did not change from baseline assessment during the intervention, whereas participants in the RAT group reported significant improvements starting with session 3 ($P = 0.004$), and performance remained significantly higher until the final assessment ($P < 0.001$). Comparisons conducted between groups suggested that children in the RAT condition had significantly better performance starting with intervention session 2 ($P = 0.032$) and maintained this advantage until the end of the study ($P < 0.001$). In the case of joint attention, there were no significant differences between the SHT and RAT groups. Comparisons within the SHT group showed decreases in scores when comparing initial assessment with intervention session 5 ($P < 0.001$), session 7 ($P < 0.001$), session 8 ($P < 0.001$), and final assessment ($P = 0.003$). No within-group differences emerged for the RAT group when comparing intervention sessions and the final assessment with baseline scores. The RAT intervention had better scores for the two final intervention sessions compared with SHT; however, the difference was no longer apparent in the final assessment.

For turn-taking, both groups improved their performance during the interventions starting with the first half of the sessions. The overall performance was lower in the information sharing task [mean = 0.48, standard error (SE) = 0.03] compared with categorization (mean = 0.66, SE = 0.3) and pattern completion (mean = 0.65, SE = 0.30). The SHT group showed some advantages, especially in the pattern completion task, but there were no significant overall or final assessment differences between the two groups.

Regarding the results for the secondary outcomes (eye contact and smiling) in the imitation tasks, children maintained eye contact better in RAT sessions (mean = 8.01, SE = 0.53) compared with SHT (mean = 5.54, SE = 0.53). Similar results were obtained for smiling: Children in the RAT group smiled more (mean = 7.72, SE = 0.58) compared with children in the SHT group (mean = 3.80, SE = 0.54). In the joint-attention tasks, children also seemed to enjoy the RAT sessions more, showing increased levels of eye contact (mean = 4.21, SE = 0.33) compared with SHT sessions (mean = 2.94, SE = 0.32), and they smiled more (RAT sessions mean = 4.25, SE = 0.34; SHT sessions mean = 2.60, SE = 0.33). The same pattern emerged in the turn-taking tasks: Although there were some differences between the types of turn-taking tasks, overall, children were more engaged in the RAT tasks, showing more eye-contact episodes (mean = 12.12, SE = 0.98) compared with SHT sessions (mean =

6.93, SE = 0.96), and smiled in RAT sessions (RAT mean = 15.33, SE = 1.27; SHT mean = 8.44, SE = 1.24).

When investigating the differences in clinical symptoms, as illustrated in Fig. 3, we did not find any significant differences between the SHT and RAT groups [$F_{1,60.75} = 3.15$, $P = 0.381$]; however, our results revealed that the scores decreased from pretest (mean = 14.23, SE = 0.48) to posttest (mean = 12.60, SE = 0.52) for both groups. A similar pattern was observed when analyzing the Social Communication Questionnaire (SCQ) results: Scores decreased from pretest (mean = 16.20, SE = 0.70) to posttest (mean = 14.56, SE = 0.81). Both the scores from the Autism Disorder Observation Schedule (ADOS) and from the SCQ were very close to the cutoff scores (for autism diagnoses) reported in some studies.

In summary, the results indicate that RAT is as effective as SHT in changing some of the symptoms and underlying mechanisms (joint attention, imitation, and turn-taking) of ASD. This is important for both RAT and SHT, taking into account the low psychological treatment (two initial evaluation sessions, eight intervention sessions, and two final evaluation sessions). RAT was more effective than SHT in improving various engagement mechanisms. Considering that treatment protocols for ASD are intense or long, our results suggest using RAT independently or as an additional alternative to SHT (e.g., when there is a decrease in motivation during SHT).

Study 2: Effectiveness of RAT

Having established the efficacy for some mechanisms in RAT in the laboratory in study 1, we turned to investigating stability and generalizability in real-life settings in study 2. We conducted an effectiveness study implemented in 10 special education institutions and therapy centers for ASD children in Cluj-Napoca, Romania. The

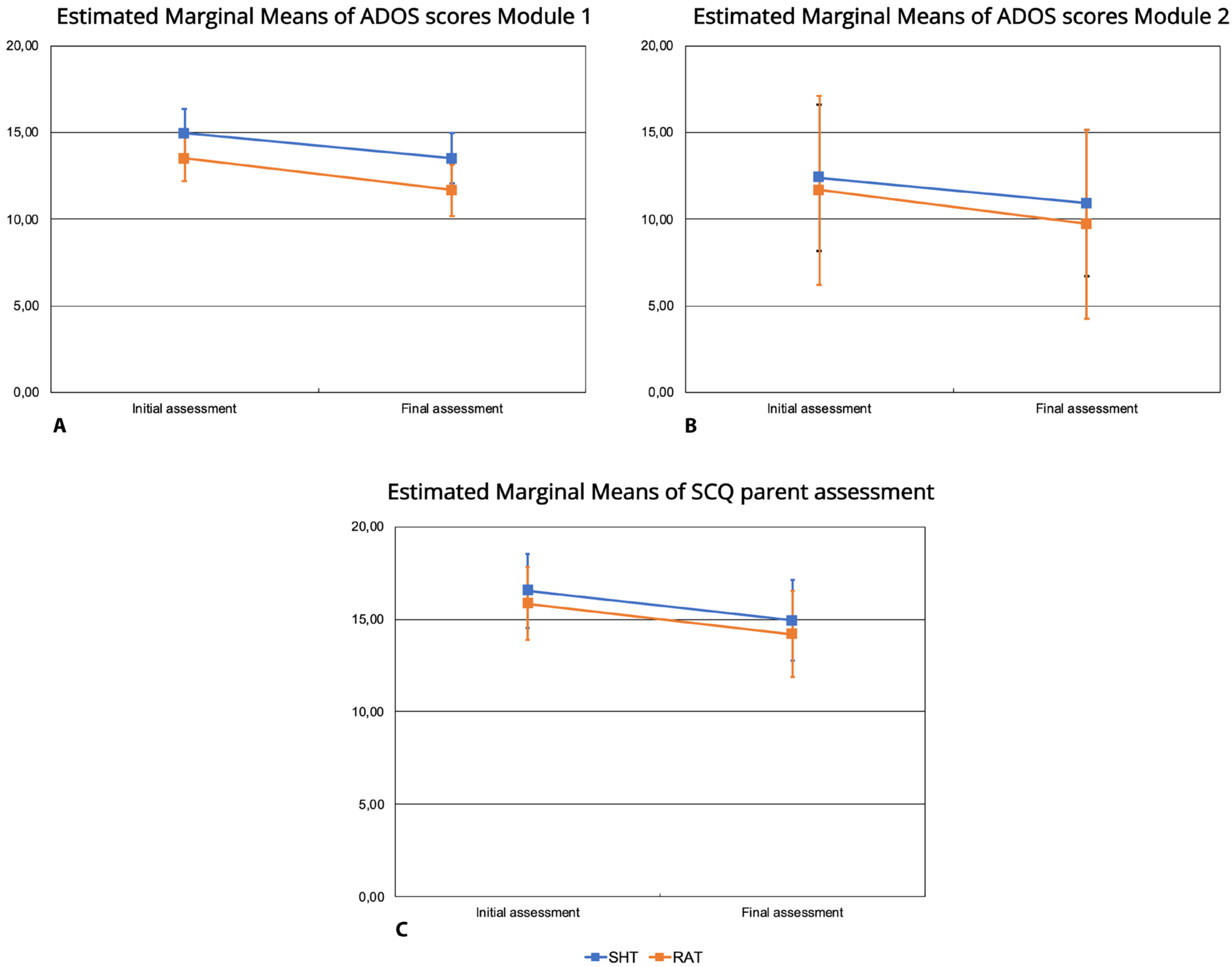

Fig. 3. Study 1 results: Efficacy of RAT using the clinical trials platform—clinical outcomes (ADOS and SCQ). ADOS scores for module 1 (**A**) were analyzed using a hierarchical linear model that included fixed effects for time (pretest versus posttest), group, group × time interaction, and a fixed intercept. Results indicated a significant effect for time [$F_{1,39.75} = 21.46$, $P < 0.001$]. No significant effects for group [$F_{1,60.75} = 3.15$, $P = 0.381$] or interaction [$F_{1,39.75} = 0.30$, $P = 0.586$] were found. The estimated marginal means for the time effect indicated that scores decreased from pretest (mean = 14.23, SE = 0.48) to posttest (mean = 12.60, SE = 0.52). SCQ parent assessments (**C**) were analyzed using a similar hierarchical linear model including fixed effects: intercept, time (pretest versus posttest), group, and group × time interaction. The results indicated a significant effect for time [$F_{1,37.22} = 7.80$, $P = 0.008$] but no significant effects of group [$F_{1,69.81} = 0.27$, $P = 0.607$] or interaction [$F_{1,37.22} = 0.01$, $P = 0.966$]. The estimated marginal means for the time effect indicated that scores decreased from pretest (mean = 16.20, SE = 0.70) to posttest (mean = 14.56, SE = 0.81). ADOS scores for module 2 (**B**) were not analyzed statistically because of low statistical power; see Supplementary Materials for details.

children were randomly allocated to one of two conditions: RAT intervention delivered via the educational platform (including treatment as usual) and a control group receiving only treatment as usual. All children were recruited from special education institutions, meaning that they all followed regular therapeutic and educational programs outside of the study protocol. Participating children had a history of diagnosed ASD symptoms and were able to understand the instructions required for assessment and for delivering the RAT intervention. Therefore, in contrast with study 1, here, we tested RAT in real-life conditions, for a larger age range, and in a transdiagnostic framework (i.e., symptoms of ASD are often comorbid with other clinical conditions). Moreover, this study was conducted as a superiority study, and we expected that the addition of RAT to the standard treatment-as-usual intervention would improve outcomes.

Treatment as usual in the special schools from Romania, where data were collected, entails a structured approach that combines individualized educational programming with therapeutic and social support services. Typically, children benefit from daily participation in small-group and one-on-one instructional activities led by specialized teachers and therapists. The curriculum is adapted to each child's developmental level, focusing on core educational domains (such as communication, social skills, self-care, and academic abilities) and integrating established intervention models such as ABA, speech and language therapy, occupational therapy, and social skills training. Intervention strategies may include structured teaching, visual supports, repetition, reinforcement-based learning, and the use of clear routines and predictable schedules to enhance adaptation and learning. The intervention environment typically emphasizes low student-teacher ratios; individualized supports; and collaboration among educators, therapists, and families to maximize developmental gains and social inclusion.

The results indicated that the RAT group outperformed the control group on imitation abilities. No significant differences between the two groups appeared for joint attention and turn-taking. Results regarding the hypothesized underlying mechanisms are presented in Fig. 4. Results regarding social interaction difficulties, as rated by parents and teachers, are provided in Fig. 5.

Children improved their imitating ability from initial evaluation to final evaluation in the RAT group (pretest mean = 2.18, SE = 0.10 to posttest mean = 2.45, SE = 0.10). However, there was no significant difference between the RAT group and the treatment-as-usual group [$F_{1,61.44} = 0.64$, $P = 0.427$]. Similar results were obtained for turn-taking skills, meaning that there were significant improvements from initial to final assessment (pretest mean = 2.20, SE = 0.12 to posttest mean = 2.53, SE = 0.12) but no significant differences between the treatment-as-usual group and RAT group [$F_{1,60.77} = 0.07$, $P = 0.789$]. For joint attention, there were no significant differences from initial to final assessment [$F_{1,59.64} = 3.20$, $P = 0.079$] individually or between groups [$F_{1,60.91} = 0.01$, $P = 0.933$]. Regarding emotion recognition, the results indicated no significant effects for time [$F_{1,60.15} = 0.83$, $P = 0.367$], group [$F_{1,62.79} = 0.15$, $P = 0.703$], or their interaction [$F_{1,60.15} = 0.449$, $P = 0.506$].

The results revealed a significant effect for time and interaction skill perceived both by parents [$F_{1,51.52} = 4.88$, $P = 0.032$] and by teachers [$F_{1,53.86} = 12.14$, $P < 0.001$] but no significant effect between the treatment-as-usual control group and RAT, either for parents [$F_{1,57.17} = 3.60$, $P = 0.063$] or for teachers [$F_{1,57.67} = 0.001$, $P = 0.971$].

We found in study 2 that RAT was well accepted among children, parents, teachers, and therapists. Both RAT and treatment as usual led to improvements from pre- to posttest on some of the mechanisms (imitation and turn-taking) and in the evaluation of social skills by both parents and teachers. This is an important finding considering the low psychological treatment dosage in study 2 (one initial assessment, three interventions, and one final evaluation session). RAT had a greater effect than the control group (treatment as usual) on imitation.

DISCUSSION

The two studies aimed to investigate the efficacy and effectiveness of RAT for children with autism compared with conventional SHT and treatment as usual. Here, we have proposed and evaluated an approach that bridges the gap between theory and practice, with a focus on training the underlying mechanisms (imitation, joint attention, and turn-taking) that can make a difference in autism interventions. This work validated a standardized protocol for RAT for children with autism that can improve their imitation skills. Although some previous studies demonstrated the effectiveness of social robots in developing joint attention (*29*, *30*), we did not find any significant difference compared to SHT in our work. One important finding from investigating joint attention performance during the RAT sessions in comparison with SHT sessions is that, in the latter, children tended to lose interest and could even perform worse after several sessions than in the initial evaluation because of the repetitive tasks, whereas with RAT they maintained their interest. Their increased interest can be observed in their performance in the joint-attention tasks, their increased positive emotions (e.g., smiling), and their prolonged eye contact. One possible explanation is that our intervention protocol used standardized tasks, and recent studies have shown the need to personalize the intervention according to children's needs (*15*, *31*) to obtain a better influence on children's abilities. A recent study (*12*) reported similar results as ours: no significant differences between SHT and RAT in a large sample after treatment measures. One possible explanation is that we included intensive interventions in our treatment-as-usual conditions, which is different from previous studies, where control groups (*32*) or community-based or ABA-focused interventions (*33*, *34*) were used.

Another important finding pertains to the measured decrease in clinical symptoms of ASD after RAT in study 1. The results highlight significant improvements in core developmental mechanism-level skills—imitation, joint attention, and turn-taking—that served as our primary outcomes. However, generalization to broader, functional domains was not directly assessed. Although reductions in clinical symptom scores on ADOS and SCQ approached relevant cutoffs, promising potential improvements observed in trained tasks do not necessarily equate to generalized social or clinical gains. The lack of group-level differences on validated distal clinical scales further suggests that RAT's effects in this study primarily reflect task-specific learning rather than broad clinical change. We acknowledge this limitation and concur with the need for future studies to incorporate independent generalization measures and assessments of real-world social functioning in diverse, naturalistic environments. Nonetheless, our approach—focused on repeated engagement and systematic training of core social cognitive skills—may offer an innovative pathway for targeted skill acquisition and technology-based intervention development for children with autism. These findings invite further research into how the presence of robotized supports, even in school settings and for short periods, can facilitate foundational social learning, which will require additional validation through

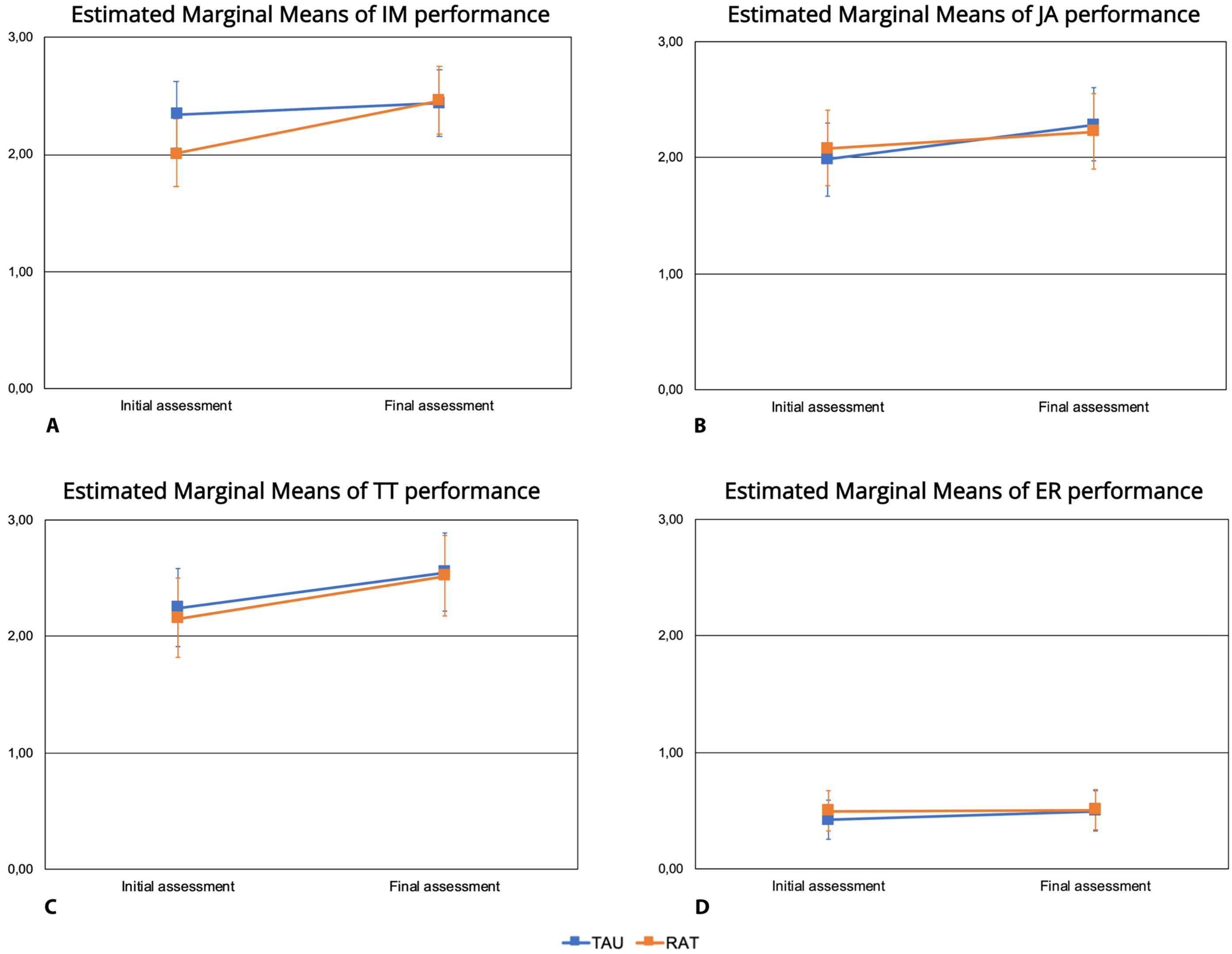

Fig. 4. Study 2 results: Effectiveness of RAT using the educational platform—performance in imitation, joint attention, turn-taking, and emotion recognition. All four measures (imitation, joint attention, turn-taking, and emotion recognition) were analyzed using hierarchical linear models including a fixed intercept and fixed effects for time (pretest versus posttest), group, and group × time interaction. For imitation (**A**), significant effects for time [$F_{1,56.34} = 21.25$, $P < 0.001$] and interaction [$F_{1,56.34} = 8.97$, $P = 0.004$] were found, but the group effect was not significant [$F_{1,61.44} = 0.64$, $P = 0.427$]. The estimated marginal means for the time effect indicated that scores increased from pretest (mean = 2.18, SE = 0.10) to posttest (mean = 2.45, SE = 0.10). The estimated marginal means for the interaction effect indicated that the scores increased significantly in the RAT group from pretest (mean = 2.01, SE = 0.14) to posttest (mean = 2.46, SE = 0.15), $P < 0.001$, but no significant change was observed in the control group, $P = 0.247$. For turn-taking (**C**), results indicated a significant effect for time [$F_{1,58.11} = 11.63$, $P = 0.001$]. The effects for group [$F_{1,60.77} = 0.07$, $P = 0.789$] and interaction [$F_{1,58.11} = 0.08$, $P = 0.775$] were not significant. The estimated marginal means for the time effect indicated that scores increased from pretest (mean = 2.20, SE = 0.12) to posttest (mean = 2.53, SE = 0.12). No significant differences for joint attention (**B**) and emotion recognition (**D**) were found.

generalization-focused, ecologically meaningful outcome measures. Overall, our results indicate the effectiveness of our approach to RAT both in controlled and real-life conditions.

As previously seen in other studies (*35*), RAT outperforms SHT in controlled conditions (study 1) for various engagement measures, which is important from a long-term perspective, to keep children engaged. Also, the engagement and the novelty effect were thoroughly investigated in previous studies in the domain of RAT, and our results reveal that, even if the tasks are repetitive, children were still motivated and interested to continue in the RAT sessions. In real-world conditions (study 2), RAT was found to outperform SHT (in the form of treatment as usual) in the case of imitation. This result can contribute to future development of RAT interventions, suggesting that some of the underlying mechanisms are more appropriate to be trained using social robots.

One important limitation of our work is the absence of a no-treatment control group. Therefore, our results should be interpreted with caution because the outcomes could be also influenced by other factors. Future studies should consider including a treatment-as-usual group (which we did not include in our study 1) and a no-treatment control group. Despite this limitation, the findings provide valuable insights and serve as a foundational basis for further exploration into the use of robot-assisted intervention for children with neurodevelopmental disorders. Moreover, children from our treatment-as-usual

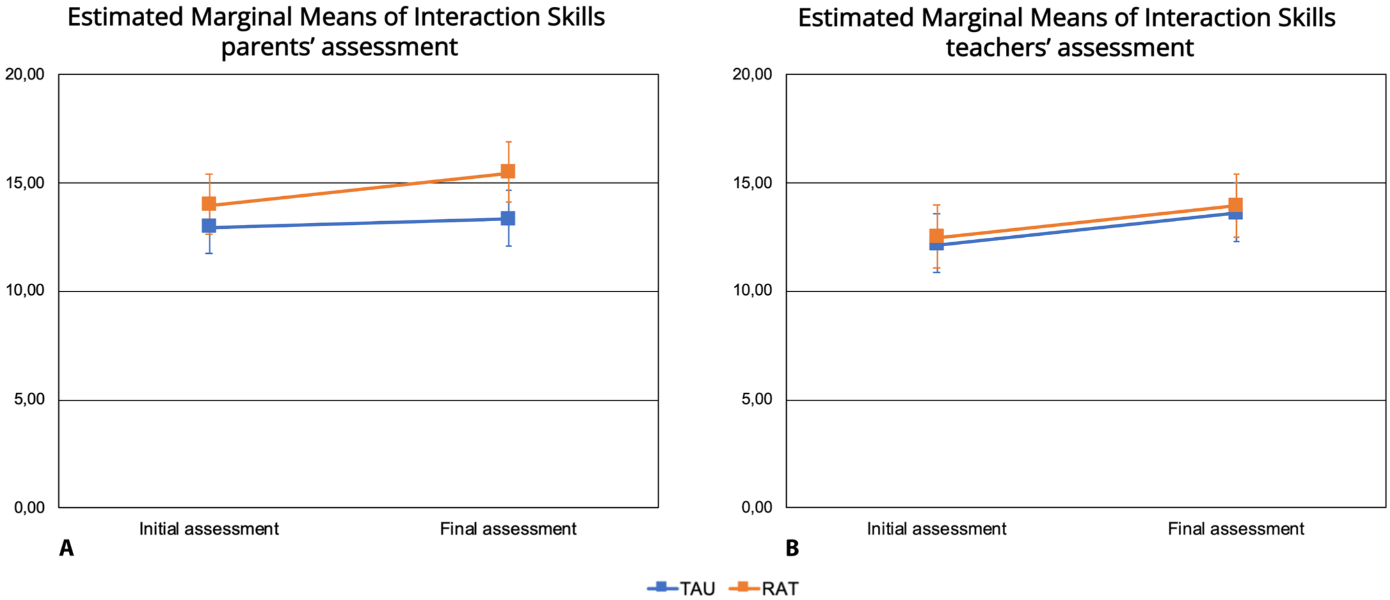

Fig. 5. Study 2 results: Effectiveness of RAT using the educational platform—performance in social interaction assessed by parents and teachers. Interaction skill assessments (by both parents and teachers) were analyzed using hierarchical linear models including fixed effects for the intercept, time (pretest versus posttest), group, and group × time interaction. Results for parents' assessments (**A**) indicated a significant effect for time [$F_{1,51.52} = 4.88$, $P = 0.032$] but not for group [$F_{1,57.17} = 3.60$, $P = 0.063$] or interaction [$F_{1,51.52} = 1.742$, $P = 0.193$]. The estimated marginal means for the time effect indicated that scores increased from pretest (mean = 13.45, SE = 0.45) to posttest (mean = 14.38, SE = 0.48). Results for teachers' assessments (**B**) indicated a significant effect for time [$F_{1,53.86} = 12.14$, $P < 0.001$] but not for group [$F_{1,57.67} = 0.001$, $P = 0.971$] or interaction [$F_{1,53.86} = 0.52$, $P = 0.472$]. The estimated marginal means for the time effect indicated that scores increased from pretest (mean = 12.27, SE = 0.49) to posttest (mean = 13.46, SE = 0.49).

group received evidence-based interventions, so when compared with evidence-based interventions, even a similar performance or a small effect on the RAT group represents an important and relevant gain.

The fact that the training program used here specifically targeted mechanisms such as imitation, joint attention, and turn-taking, subsequently used as metrics to evaluate improvement, could also be considered a drawback. This overlap between the intervention focus and the assessment metrics may introduce bias, potentially influencing the perceived effectiveness of the training. This limitation underscores the need for future research to incorporate a broader range of outcome measures, including those that assess generalized and functional improvements beyond the directly targeted skills. The measurements used in this study for clinical symptoms were ADOS and SCQ, instruments that were not so sensitive to the improvements made by children in RAT sessions to show the difference between the two groups.

Another limitation of our studies, as for most other studies, is related to the reduced dose of the treatment (intensity or duration). In future work, these protocols should be tested in more intense clinical interventions for longer durations. On the other hand, this can also be considered a strength because it has allowed us to test both our RAT approach's internal validity (in the efficacy study) and the external validity (in the effectiveness study).

An important limitation of study 2 is the small number of intervention sessions provided. With only three intervention sessions, it is challenging to determine the long-term effectiveness and sustainability of the observed improvements. The short duration of the intervention may not have been sufficient to evaluate lasting changes or fully capture the potential benefits of the training. Consequently, the results might reflect initial gains rather than enduring improvements in the targeted outcomes.

A final limitation of the work presented here lies in the standardized approach used for all participants, as mentioned above. Recent studies highlight the importance of tailored intervention, and future studies should consider the codesign of the intervention session together with end users (*36*, *37*). Effective treatment for ASD often requires personalized interventions tailored to individual needs, especially because ASD encompasses a wide range of characteristics and varies substantially from individual to individual. Rigid protocols may fail to provide the personalized attention necessary for effective intervention, potentially leading to suboptimal outcomes.

A critical consideration in evaluating the effectiveness of RAT is determining whether the observed increase in children's engagement is due to the unique features of the robot itself or simply the novelty of the activity. In our study, it is unlikely to be a novelty effect because we saw an increase in engagement over several sessions. Rather, the robot's design and interactive capabilities, and the way the tasks are built, appear to intrinsically motivate and sustain engagement. This suggests a more durable and substantive therapeutic benefit, demonstrated by children showing increased engagement over several intervention sessions.

Overall, the results contribute to an understanding of the relevance of technological solutions in treating children with ASD and reducing the human workload that traditional therapies impose. Evidence-based psychological treatments for ASD (e.g., derived from applied behavioral analysis) are intense and of long duration. Many patients with ASD and their families do not have access to necessary treatments, cannot afford the costs, do not respond well, or relapse. In the absence of available therapy, many children diagnosed with ASD are treated solely with medication despite the lack of persistent effects. In summary, the work presented here constitutes one of the largest studies on RAT in autism treatment, further extending the existing clinical evidence base, and it provides a systematic approach to evaluating both efficacy and effectiveness in an integrated manner through the combination of complementary clinical studies using a unified approach. Future studies could consider more than the three skills addressed here (or others) in their protocols, and there is still scope to personalize the interventions according to children's needs.

MATERIALS AND METHODS

Study 1: Efficacy of RAT using the clinical trials platform

In study 1, using the clinical trials platform, we investigated the efficacy of RAT compared with SHT in an equivalence trial design. We established an equivalence margin of Cohen's $d = 0.79$, meaning that the maximum standardized difference between the two groups should not be larger than a medium effect size. The clinical trial protocol was registered on ClinicalTrials.gov with the number NCT03323931. The study received prior ethical approval from the Scientific Council of Babes-Bolyai University in Cluj-Napoca, Romania, where the trial was conducted (record no. 30664/10 February 2017). A letter of consent was signed by at least one parent before the study, expressing their consent to record assessments and interventions and to use the data and recordings for scientific purposes in an anonymous fashion.

Sample

Participants were between 2.5 and 7 years and had a formal diagnosis of autism in their medical record, confirmed by an assessment with the ADOS (*38*) carried out by clinical psychologists with an ADOS certification (and blind to the study design). Most of the children had a score of 12 or above (51 children), meaning that they have a moderate to substantial need for support (*39*). Eighteen children had scores from 7 to 11, meaning that they only need minimal support from the caregivers. Children were excluded if they had other comorbid neurodevelopmental or learning disorders; if they had high levels of social skills at the initial assessment, as indicated by perfect performance on all of the assessment tasks; or if they lacked the ability to understand the task requirements and follow the instructions in a session of about 10 to 15 min. Figure 6 presents a CONSORT (Consolidated Standards of Reporting Trials) diagram providing details on the study participants. The final sample included 69 participants, of which 12 were female (17.4%), with a mean age of 52.58 months (SD = 11.66).

Measures

Primary outcomes were measured in terms of imitation performance, joint attention performance, and turn-taking performance. Imitation performance was measured using a dichotomic scale (correct or incorrect answer). A good performance rating was achieved when the child had correctly imitated the behavior of the play partner (human or robot), including both the movement and the sound made by the partner. For each session, an average score was calculated for each child and used in data analysis by dividing the number of correct behavioral imitations by the number of attempts.

Joint attention performance was measured on the same dichotomic scale. Children received a positive rating if they followed the cues shown by the interaction partner. A correct answer counted if the child looked at an object displayed on the left or the right side of a large touchscreen placed in front of them. The eye-tracking algorithms of the clinical trials platform were used to detect the eye gaze of the child (*40*), whereas the observing clinician used visual input coming from the left and right cameras (positioned with respect to the child) to decide the rating of the behavior. Similar to imitation, an average score was calculated for each child in each session.

Turn-taking performance was also measured as correct or incorrect for each trial with the help of the sensorized table and supervised by the clinician. Children received a positive performance rating if they waited for the partner's turn in the game without touching the screen while the partner performed the action. The semi-autonomous supervised system determined whether the child waited their turn by means of skeleton and posture analysis based on the video and depth cameras in the setup (*40*–*42*). An average score for each child for each session was also used for turn-taking in data analysis.

Secondary outcomes included the engagement of the child, measured in terms of eye gaze and smiling at the play partner during the sessions. Engagement measures were chosen so that they offer an estimate of the degree to which the child was focusing on the treatment activities and enjoying them, rather than being distracted, getting involved in other activities, or feeling distressed during the intervention. Eye contact with the play partner was measured in real time (*40*). An eye-contact event was defined by the eye gaze of the child directed toward the head of the play partner for at least 200 ms. The presence or absence of an eye-contact event was determined using eye-tracking algorithms. Smiling during sessions was also considered an indicator of engagement and was measured in real time by the sensorized table detecting smiling expressions on the basis of the facial landmarks collected from the cameras during the intervention (*43*).

ADOS (*38*) is a set of standardized activities designed to cue behaviors that are relevant for the diagnosis of ASD and related to social behavior and communication (e.g., responding when one's name is called, asking for help to get access to food, responding to social signals such as smiles, etc.). ADOS comprises four modules that have different activities adapted to the age and the level of verbal abilities of the child. In the current study, only module 1 (for children who do not use phrase speech) and module 2 (for children who make spontaneous use of meaningful short phrases of up to three words) were used. ADOS offers a coding and rating system tied to clinical criteria for the diagnosis of ASD, allowing the clinician to rate the level of severity and establish a likely diagnosis by comparing the score of the child with the cutoff scores established by research.

The SCQ (*44*) is a questionnaire consisting of 40 items on which a parent of the child rates the presence and the absence of normal and abnormal behaviors that are relevant for ASD. Two forms are available: "current" and "lifetime." The former looks at the child's past 3 months, whereas the latter looks at the child's entire developmental history. The current symptoms version of the questionnaire, which we used to assess parents' perceptions over the changes in ASD symptoms, has shown poorer diagnostic utility in children under 4 years (*45*), and thus we did not use its cutoff scores. Instead, following other studies (*46*), we used this scale as a continuous measure of symptoms as reported by parents to gather information about the generalizability of the intervention effects beyond the interaction with the therapist.

To determine parents' perceptions of child improvements and satisfaction, at the end of the intervention, we asked one parent of each child to rate the perceived improvements of the child overall (one item) and on the three specific skills (three items) using a self-report scale from 1 ("I didn't see any progress") to 5 ("The progress was very good"). We also asked parents to rate their satisfaction with the treatment that the child received using 10 items, each asking about satisfaction with a specific component (e.g., how assessment was conducted, the assessment reports they received, and the degree to which the intervention was

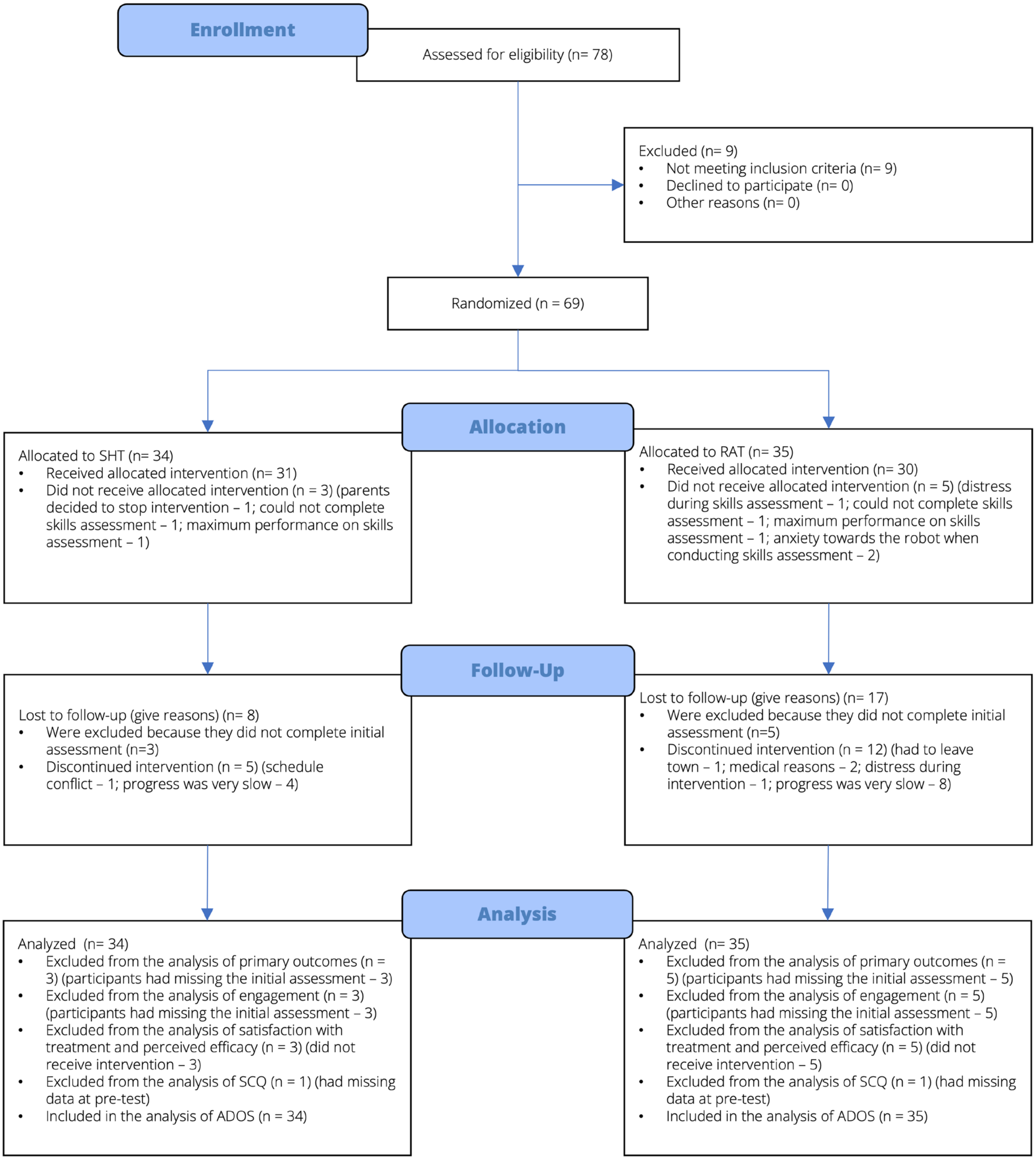

Fig. 6. Study 1: Efficacy of RAT using the clinical trials platform. CONSORT flow diagram showing how participants were enrolled, allocated to the SHT and RAT groups, excluded or discontinued, and analyzed.

tailored for the needs and skills of the child). The satisfaction for each item was rated on a scale from 1 (not at all satisfied) to 5 (very satisfied). Both scales were adapted for the current study and showed excellent reliability, with Cronbach's alphas of 0.88 for perceived improvements and 0.93 for satisfaction.

Procedure

All recruited participants were required to attend 12 sessions: 2 initial evaluation sessions (i.e., preintervention), 8 intervention sessions, and 2 final evaluation sessions (i.e., postintervention). The 45-min sessions were held biweekly. The overall structure of the 12 sessions was similar for both groups. Sessions 1 and 2 were designed to carry out a comprehensive psychological evaluation and to determine the baseline level of the three fundamental social abilities (imitation, joint attention, and turn-taking). ADOS was administrated to each child to assess the social and communicative behavior associated with ASD. In addition, the SCQ was completed by the children's parents or caregivers. Sessions 3 to 10 were used for training imitation, joint attention, and turn-taking. Sessions 11 and 12 first determined the level of the three abilities after the interventions through tasks in which the targeted behaviors were preceded by a stimulus but not followed by any feedback. In the postintervention assessment, ADOS and SCQ were used to determine whether the benefits of the interventions were generalized.

Intervention sessions with the robot assistant had a similar structure and followed the principles of discrete trial training (*47*). Intervention tasks were introduced as gameplay activities. The interaction partner presented a discriminative stimulus and waited for the child's response. If the child response matched the expected one, then the robot provided positive social feedback (e.g., praise). If the child's performance was below the expected one, then the robot provided encouragement to try again. The child was given three opportunities to respond to each exercise. After the third failure, the therapist guiding the intervention prompted the child's behavior.

Study 2: Effectiveness of RAT using the educational platform

Sample

The study sample included 63 participants, of which 7 were females (11.11%), with a mean age of 5.89 years (SD = 2.63). The sample size was calculated using *G**Power with the following specifications: $\alpha = 0.05$, $1 - \beta = 0.80$, and an estimated effect size (Cohen's *d*) for the population of at least 0.80. Children were included if their ages were between 3 and 12 years, if they had a history of diagnosed ASD symptoms, and if they were able to understand the instructions required for assessment and for delivering the RAT intervention. Figure 7 presents the CONSORT flow diagram providing details on the participants.

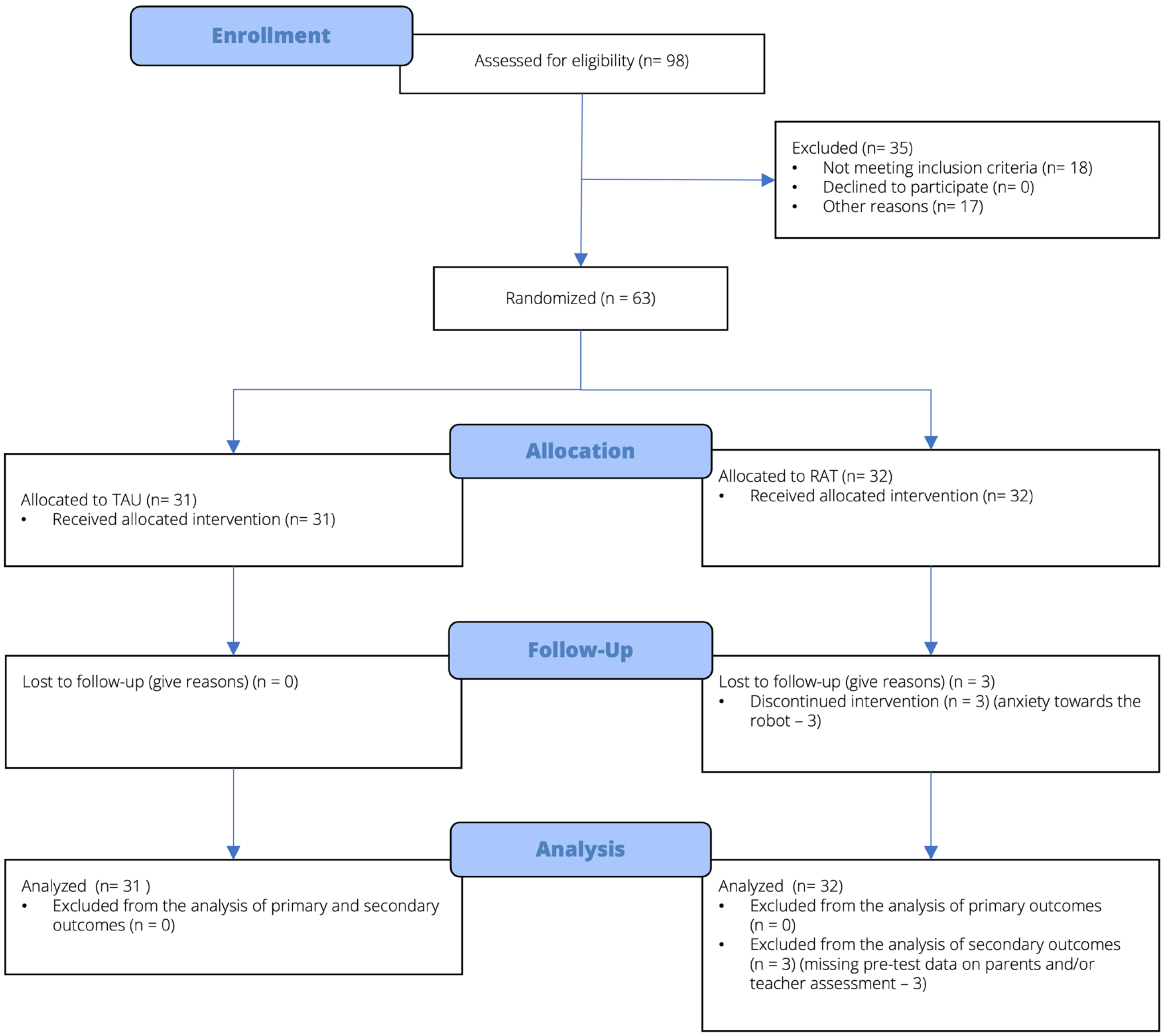

Fig. 7. Study 2: Effectiveness of RAT using the educational platform. CONSORT flow diagram showing how participants were enrolled, allocated to the treatment-as-usual group and RAT group, excluded or discontinued, and analyzed.

Measures

The primary outcomes were constituted by the same basic underlying mechanisms as in study 1 (imitation, joint attention, and turn-taking), measured with specific tasks derived from ADOS. Moreover, we added an additional emotion recognition measure related to the recognition of emotional expressions. As secondary outcomes, we measured the difference between initial assessment and final assessment of the clinical symptoms of social interaction, as rated by parents and teachers.

Emotion recognition was measured during the interaction sessions when children had to indicate the emotions expressed by the robotic play partner: happy, sad, scared, or angry. Children who were verbal could do so by naming the emotion, whereas nonverbal children were given a set of photo cards to express each of the four possible emotions.

To assess the interaction skills as perceived by parents and teachers, for each child in each group, we asked one parent and one teacher to assess the child's ability to interact with others using a set of items extracted from the TRIAD Social Skills Assessment that we used for parents and teachers [(*48*); specifically, we used items 44 to 49 from the parent's module and items 45 to 50 from the teacher's module]. The items were identical for both raters, and we asked them to indicate how competent the child was, on a scale from 1 (not very competent) to 4 (very competent), in expressing feeling, understanding the perspective of others, initiating social interaction, among others. Both raters filled out the items before and after the intervention. The scale was translated and adapted into Romanian language for the present study. The internal consistency, a measure of reliability that evaluates how closely related the items in a questionnaire are, was very good, with Cronbach's alphas of greater than 0.80 for both parents and teachers at both points in time.

Procedure

The intervention was a shortened and simplified version of the one implemented in study 1. We started with a behavioral assessment of imitation, joint attention, turn-taking, and emotion recognition skills. After the initial assessment, we offered children three intervention sessions of ~30 min each, performed daily or every second day, that targeted these skills. This was followed by a final evaluation session identical to the initial one.

In study 2, the robot was controlled by a research assistant present in the room using a DREAM Lite tablet application. The coding of the child's behavior (the degree to which the child imitated, initiated joint attention, and took turns) was made by a graduate-level clinical psychologist specifically trained for this purpose. All sessions were recorded with a video. Each team delivering the intervention, consisting of the therapist, the observer coding child's responses, and the research assistant controlling the robot, had weekly meetings with a supervisor who was a PhD-level experienced clinical psychologist or a special education specialist. During supervision meetings, each individual case was discussed, and intervention sessions were reviewed on the basis of the video.

Supplementary Materials

The PDF file includes:
Materials and Methods
References (*49*–*56*)

Other Supplementary Material for this manuscript includes the following:
MDAR Reproducibility Checklist

REFERENCES AND NOTES

1. K. Dautenhahn, "Robots as social actors: Aurora and the case of autism" in *Proceedings of the Third International Cognitive Technology Conference (CT99)* (Michigan State Univ., 1999), pp. 359–374.
2. K. Dautenhahn, I. Werry, Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmat. Cogn.* **12**, 1–35 (2004).
3. J. J. Diehl, L. M. Schmitt, M. Villano, C. R. Crowell, The clinical use of robots for individuals with autism spectrum disorders: A critical review. *Res. Autism Spectr. Disord.* **6**, 249–262 (2012).
4. B. Scassellati, H. Admoni, M. Matarić, Robots for use in autism research. *Annu. Rev. Biomed. Eng.* **14**, 275–294 (2012).
5. A. Kouroupa, K. R. Laws, K. Irvine, S. E. Mengoni, A. Baird, S. Sharma, The use of social robots with children and young people on the autism spectrum: A systematic review and meta-analysis. *PLOS ONE* **17**, e0269800 (2022).
6. M. Clark, D. Adams, The self-identified positive attributes and favourite activities of children on the autism spectrum. *Res. Autism Spectr. Disord.* **72**, 101512 (2020).
7. M. Jones, M. Falkmer, B. Milbourn, T. Tan, S. Bölte, S. Girdler, The core elements of strength-based technology programs for youth on the autism spectrum: A systematic review of qualitative evidence. *Rev. J. Autism Dev. Disord.* **10**, 441–457 (2023).
8. A. Alabdulkareem, N. Alhakbani, A. A. Al-Nafjan, Systematic review of research on robot-assisted therapy for children with autism. *Sensors* **22**, 944 (2022).
9. H. L. Cao, P. G. Esteban, M. Bartlett, P. Baxter, T. Belpaeme, E. Billing, H. Cai, M. Coeckelbergh, C. Costescu, D. David, A. De Beir, D. Hernandez, J. Kennedy, H. Liu, S. Matu, A. Mazel, A. Pandey, K. Richardson, E. Senft, S. Thill, G. van de Perre, B. Vanderborght, D. Vernon, K. Wakanuma, H. Yu, X. Zhou, T. Ziemke, Robot-enhanced therapy: Development and validation of supervised autonomous robotic system for autism spectrum disorders therapy. *IEEE Robot. Autom. Mag.* **26**, 49–58 (2019).
10. Z. Salimi, E. Jenabi, S. Bashirian, Are social robots ready yet to be used in care and therapy of autism spectrum disorder: A systematic review of randomized controlled trials. *Neurosci. Biobehav. Rev.* **129**, 1–6 (2021).
11. R. Vagnetti, A. D. Nuovo, M. Mazza, M. Valenti, Social robots: A promising tool to support people with autism. A systematic review of recent research and critical analysis from the clinical perspective. *Rev. J. Autism Dev. Disord.* 10.1007/s40489-024-00434-5 (2024).
12. I. van den Berk-Smeekens, M. W. P. de Korte, M. van Dongen-Boomsma, I. J. Oosterling, J. C. den Boer, E. I. Barakova, T. Lourens, J. C. Glennon, W. G. Staal, J. K. Buitelaar, Pivotal response treatment with and without robot-assistance for children with autism: A randomized controlled trial. *Eur. Child Adolesc. Psychiatry* **31**, 1871–1883 (2022).
13. N. L. Robinson, T. V. Cottier, D. J. Kavanagh, Psychosocial health interventions by social robots: Systematic review of randomized controlled trials. *J. Med. Internet Res.* **21**, e13203 (2019).
14. V. Holeva, V. A. Nikopoulou, C. Lytridis, C. Bazinas, P. Kechayas, G. Sidiropoulos, M. Papadopoulou, M. D. Kerasidou, C. Karatsioras, N. Geronikola, G. A. Papakostas, V. G. Kaburlasos, A. Evangeliou, Effectiveness of a robot-assisted psychological intervention for children with autism spectrum disorder. *J. Autism Dev. Disord.* **54**, 577–593 (2024).
15. B. Scassellati, L. Boccanfuso, C. M. Huang, M. Mademtzi, M. Qin, N. Salomons, P. Ventola, F. Shic, Improving social skills in children with ASD using a long-term, in-home social robot. *Sci. Robot.* **3**, eaat7544 (2018).
16. D. Ghiglino, F. Floris, D. De Tommaso, K. Kompatsiari, P. Chevalier, T. Priolo, A. Wykowska, Artificial scaffolding: Augmenting social cognition by means of robot technology. *Autism Res.* **16**, 997–1008 (2023).
17. E. Billing, T. Belpaeme, H. Cai, H.-L. Cao, A. Ciocan, C. Costescu, D. David, R. Homewood, D. Hernandez, P. G. Esteban, H. Liu, V. Nair, S. Matu, A. Mazel, M. Selescu, E. Senft, S. Thill, B. Vanderborght, D. Vernon, T. Ziemke, The DREAM dataset: Supporting a data-driven study of autism spectrum disorder and robot enhanced therapy. *PLOS ONE* **15**, e0236939 (2020).
18. C. Clabaugh, K. Mahajan, S. Jain, R. Pakkar, D. Becerra, Z. Shi, E. Deng, R. Lee, G. Ragusa, M. Matarić, Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders. *Front. Robot. AI* **6**, 110 (2019).
19. R. Pakkar, C. Clabaugh, R. Lee, E. Deng, M. Matarić, "Designing a socially assistive robot for long-term in-home use for children with autism spectrum disorders" in *2019 28th IEEE International Conference on Robot and Human Interactive Communication* (*RO-MAN*) (IEEE, 2019), pp. 1–7; 10.1109/RO-MAN46459.2019.8956468.
20. S. Jain, B. Thiagarajan, Z. Shi, C. Clabaugh, M. J. Matarić, Modeling engagement in long-term, in-home socially assistive robot interventions for children with autism spectrum disorders. *Sci. Robot.* **5**, eaaz3791 (2020).
21. K. K. Poon, L. R. Watson, G. T. Baranek, M. Poe, To what extent do joint attention, imitation, and object play behaviors in infancy predict later communication and intellectual functioning in ASD? *J. Autism Dev. Disord.* **42**, 1064–1074 (2012).
22. Z. Zheng, G. Nie, A. Swanson, A. Weitlauf, Z. Warren, N. Sarkar, A randomized controlled trial of an intelligent robotic response to joint attention intervention system. *J. Autism Dev. Disord.* **50**, 2819–2831 (2020).

23. M. W. De Korte, I. van den Berk-Smeekens, M. van Dongen-Boomsma, I. J. Oosterling, J. C. Den Boer, E. I. Barakova, T. Lourens, J. K. Buitelaar, J. C. Glennon, W. G. Staal, Self-initiations in young children with autism during pivotal response treatment with and without robot assistance. *Autism* **24**, 2117–2128 (2020).
24. H. Kumazaki, Y. Yoshikawa, Y. Yoshimura, T. Ikeda, C. Hasegawa, D. N. Saito, S. Tomiyama, K. M. An, J. Shimaya, H. Ishiguro, Y. Matsumoto, Y. Minabe, M. Kikuchi, The impact of robotic intervention on joint attention in children with autism spectrum disorders. *Mol. Autism* **9**, 46 (2018).
25. C. A. G. J. Huijnen, H. A. M. D. Verreussel-Willen, M. A. S. Lexis, L. P. de Witte, Robot KASPAR as mediator in making contact with children with autism: A pilot study. *Int. J. Soc. Robot.* **13**, 237–249 (2021).
26. D. O. David, C. A. Costescu, S. Matu, A. Szentagotai, A. Dobrean, Developing joint attention for children with autism in robot-enhanced therapy. *Int. J. Soc. Robot.* **10**, 595–605 (2018).
27. D. O. David, C. A. Costescu, S. Matu, A. Szentagotai, A. Dobrean, Effects of a robot-enhanced intervention for children with ASD on teaching turn-taking skills. *J. Educ. Comput. Res.* **58**, 29–62 (2020).
28. A. Mazel, S. Matu, DREAM Lite: Simplifying robot assisted therapy for ASD. arXiv:2104.08034v1 [cs.RO] (2021); https://doi.org/10.48550/arXiv.2104.08034.
29. J. Kajopoulos, A. H. Y. Wong, A. W. C. Yuen, T. A. Dung, T. Y. Kee, A. Wykowska. "Robot-assisted training of joint attention skills in children diagnosed with autism" in *Social Robotics* (*ICSR 2015*), A. E. Tapus, J. C. André, F. Martin, M. Ammi Ferland, Eds. (Springer, 2015), pp. 296–305; https://doi.org/10.1007/978-3-319-25554-5_30.
30. F. Mehmood, Y. Ayaz, S. Ali, A. R. De Cassia, H. Sadia, Dominance in visual space of ASD children using multi-robot joint attention integrated distributed imitation system. *IEEE Access* **7**, 168815–168827 (2019).
31. Z. Shi, T. R. Groechel, S. Jain, K. Chima, O. Rudovic, M. J. Matarić, Toward personalized affect-aware socially assistive robot tutors for long-term interventions with children with autism. *ACM Trans. Hum. Robot Interact.* **11**, 1–28 (2022).
32. T. W. Vernon, A. N. Holden, A. C. Barrett, J. Bradshaw, J. A. Ko, E. S. McGarry, E. J. Horowitz, D. M. Tagavi, T. C. German, A pilot randomized clinical trial of an enhanced pivotal response treatment approach for young children with autism: The PRISM model. *J. Autism Dev. Disord.* **49**, 2358–2373 (2019).
33. A. Y. Hardan, G. W. Gengoux, K. L. Berquis, R. A. Libove, C. M. Ardel, J. M. Phillips, T. W. Frazier, M. B. Minjarez, A randomized controlled trial of pivotal response treatment group for parents of children with autism. *J. Child Psychol. Psychiatry* **56**, 884–892 (2015).
34. F. Mohammadzaheri, L. K. Koegel, M. Rezaee, S. M. Rafiee, A randomized clinical trial comparison between pivotal response treatment (PRT) and structured applied behavior analysis (ABA) intervention for children with autism. *J. Autism Dev. Disord.* **44**, 2769–2777 (2014).
35. A. Tapus, A. Peca, A. Aly, C. Pop, L. Jisa, S. Pintea, A. S. Rusu, D. O. David, Children with autism social engagement in interaction with Nao, an imitative robot: A series of single case experiments. *Interact. Stud.* **13**, 315–347 (2012).
36. J. Borghouts, E. Eikey, G. Mark, C. De Leon, S. M. Schueller, M. Schneider, N. Stadnick, K. Zheng, D. Mukamel, D. H. Sorkin, Barriers and facilitators to user engagement with digital mental health interventions: A systematic review. *J. Med. Internet Res.* **23**, e24387 (2020).
37. C. Costescu, I. R. Podina, A. Voinescu, Digital mental health: Interventions and assessment. *Front. Psychol.* **13**, 1014527 (2022).
38. C. Lord, S. Risi, L. Lambrecht, E. H. Cook Jr., B. L. Leventhal, P. C. DiLavore, A. Pickles, M. Rutter, The Autism Diagnostic Observation Schedule—Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *J. Autism Dev. Disord.* **30**, 205–223 (2000).
39. R. Luyster, K. Gotham, W. Guthrie, M. Coffing, R. Petrak, K. Pierce, S. Bishop, A. Esler, V. Hus, R. Oti, J. Richler, S. Risi, C. Lord, The Autism Diagnostic Observation Schedule—Toddler Module: A new module of a standardized diagnostic measure for autism spectrum disorders. *J. Autism Dev. Disord.* **39**, 1305–1320 (2009).
40. H. Cai, B. Liu, J. Zhang, S. Chen, H. Liu, Visual focus of attention estimation using eye center localization. *IEEE Syst. J.* **11**, 1320–1325 (2015).
41. B. Liu, Z. Ju, N. Kubota, H. Liu, "Online action recognition based on skeleton motion distribution" in *29th British Machine Vision Conference* (British Machine Vision Association, 2018).
42. S. Zhang, H. Yu, J. Dong, T. Wang, L. Qi, H. Liu, "Combining kinect and pnp for camera pose estimation" in *2015 8th International Conference on Human System Interaction* (IEEE, 2015), pp. 357–361.
43. Y. Wang, H. Yu, B. Stevens, H. Liu, "Dynamic facial expression recognition using local patch and LBP-TOP" in *2015 8th International Conference on Human System Interactions* (IEEE, 2015), pp. 362–367.
44. M. Rutter, A. Bailey, C. Lord, *The Social Communication Questionnaire* (Western Psychological Services, 2003).
45. A. R. Marvin, D. J. Marvin, P. H. Lipkin, J. K. Law, Analysis of Social Communication Questionnaire (SCQ) screening for children less than age 4. *Curr. Dev. Disord. Rep.* **4**, 137–144 (2017).
46. A. Pickles, A. L. Couteur, K. Leadbitter, E. Salomone, R. Cole-Fletcher, H. Tobin, I. Gammer, J. Lowry, G. Vamvakas, S. Byford, C. Aldred, V. Slonims, H. M. Conachie, P. Howlin, J. R. Parr, T. Charman, J. Green, Parent-mediated social communication therapy for young children with autism (PACT): Long-term follow-up of a randomised controlled trial. *Lancet* **388**, 2501–2509 (2016).
47. T. Smith, Discrete trial training in the treatment of autism. *Focus Autism Other Dev. Disabil.* **16**, 86–92 (2001).
48. W. L. Stone, L. Ruble, E, Coonrod, S, Hepburn, M. Pennington, C. Burnette, N. B. Brigham, *TRIAD Social Skills Assessment* (Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), ed. 2., 1998).
49. P. G. Esteban, P. Baxter, T. Belpaeme, E. Billing, H. Cai, H. L. Cao, M. Coeckelbergh, C. Costescu, D. David, A. De Beir, J. Kennedy, H. Liu, A. Mazel, A. Pandey, K. Richardson, E. Senft, S. Thill, G. Van de Perre, B. Vanderborght, D. Vernon, H. Yu, T. Ziemke, How to build a supervised autonomous system for robot-enhanced therapy for children with autism spectrum disorder. *Paladyn J. Behav. Robot.* **8**, 18–38 (2017).
50. H. Cai, Y. Fang, Z. Ju, C. Costescu, D. David, E. Billing, T. Ziemke, S. Thill, T. Belpaeme, B. Vanderborght, D. Vernon, K. Richardson, H. Liu, Sensing-enhanced therapy system for assessing children with autism spectrum disorders: A feasibility study. *IEEE Sens. J.* **19**, 1508–1518 (2018).
51. L. Flight, S. A. Julious, Practical guide to sample size calculations: An introduction. *Pharm. Stat.* **15**, 68–74 (2016).
52. A. F. Zuur, E. N. Ieno, N. J. Walker, A. A. Saveliev, G. M. Smith, *Mixed Effects Models and Extensions in Ecology with R* (Springer, 2009).
53. A. Feingold, Confidence interval estimation for standardized effect sizes in multilevel and latent growth modeling. *J. Consult. Clin. Psychol.* **83**, 157–168 (2015).
54. E. J. Mascha, D. I. Sessler, Equivalence and noninferiority testing in regression models and repeated-measures designs. *Anesth. Analg.* **112**, 678–687 (2011).
55. C. K. Enders, *Applied Missing Data Analysis* (Guilford Publications, 2022).
56. J. Siddique, C. H. Brown, D. Hedeker, N. Duan, R. D. Gibbons, J. Miranda, P. W. Lavori, Missing data in longitudinal trials–part B, analytic issues. *Psychiatric. Ann.* **38**, 793–801 (2008).

Acknowledgments
Funding: This work was funded by the European Commission, FP7 grant agreement 611391 (DREAM, "Development of Robot-Enhanced therapy for children with AutisM spectrum disorders," 2014-2019). T.Z. was also supported by ELLIIT, the Excellence Center at Linköping-Lund in Information Technology. **Author contributions:** Conceptualization: B.V., C.C., D.D., E.B., S.M., S.T., T.B., and T.Z. Methodology: C.C., D.D., D.V., E.B., H.L., S.M., S.T., and T.B. Investigation: A.C., A.M., C.C., D.D., D.H.G., E.B., E.S., H.C., H.-L.C., J.K., M.S., P.B., P.G.E., S.M., and S.T. Visualization: C.C., E.B., E.S., and S.M. Funding acquisition: B.V., C.C., D.D., D.V., S.T., T.B., and T.Z. Project administration: A.M., B.V., D.D., D.V., H.L., T.B., and T.Z. Supervision: B.V., D.D., H.L., T.B., and T.Z. Writing—original draft: C.C., D.D., and S.M. Writing—review and editing: C.C., D.D., D.V., E.B., S.M., S.T., T.B., and T.Z. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The data and code are available via the links below. The complete clinical trial platform—comprising the physical design, the data analysis algorithms, intervention scripts and robot control functions—is available as open source at https://github.com/dream2020/DREAM, licensed under GNU GPL v.3. In addition, the dataset collected for the present study, comprising close to 4000 sessions, is released publicly at https://github.com/dream2020/data. This dataset comprises secondary data extracted by the sensorized table during therapy, in which footage of children is excluded to preserve anonymity.

Submitted 30 April 2025
Resubmitted 30 April 2025
Accepted 2 December 2025
Published 24 December 2025
10.1126/scirobotics.adl2266