

Table Supplementary Material 3: RCT studies and respective main results

	Groups	Pre Test Evaluation	Post Test Evaluation	p value: Post Test > Pre Test	Observations and Main conclusions
Costecu					Robotic therapy can reduce significantly the negative emotion intensity associated with negative social events in ASD children
Frequency of correct strategies to be used in a social situation	Control group: 15	-	14.77 ± 3.39	p = 0.586	
	Robotic group: 12	11.64 ± 7.63	13.18 ± 5.01	p = 0.491	
Frequency of used rational beliefs	Control group: 15	-	4.33 ± 5.82	p < 0.031 ***	
	Robotic group: 12	2.38 ± 4.73	13.46 ± 8.88	p = 0.000 ***	
Frequency of used irrational beliefs	Control group: 15	-	6.06 ± 5.98	p = 0.292	
	Robotic group: 12	5.80± 5.37	4.00 ± 6.09	p = 0.251	
Emotional intensity rating scale for angry and sadness	Control group: 15	-	101.77 ± 39.31	p < 0.008 ***	
	Robotic group: 12	120.08 ± 32.68	78.17 ± 38.40	p = 0.003 ***	
Frequency of using adaptive or dezadaptive behaviours	Control group: 15	-	9.37 ± 4.80	p = 0.08 ***	
	Robotic group: 12	7.08 ± 5.23	11.00 ± 4.74	p = 0.121	
De Korte					In functional self initiations, a larger growth was verified for children in the robotic group compared to the control group. Therefore, robots within an established intervention, namely PRT for ASD, may be a valuable treatment component.
Number of self-initiations	Control group: 20	17.58 ± 6.50	16.59 ± 7.71	No linear growth over time Estimate = -0.10 (0.10), 95% CI: -0.31 to 0.11	
	Robotic group: 24	14.69 ± 5.63	19.74 ± 4.80	Linear growth over time Estimate = 0.17(0.07), 95% CI: 0.03-0.31	
Social Responsiveness Scale teachers	Control group: 20	84.74 ± 26.65	80.25 ± 25.15	not significant	
	Robotic group: 24	78.25 ± 23.66	65.05 ± 20.83	not significant	
Social Responsiveness Scale parents	Control group: 20	90.80 ± 18.60	81.91 ± 20.63	The change on the parent-rated in follow up was significantly larger (p = 0.011) in the robotic group compared with the control group.	
	Robotic group: 24	84.03 ± 23.03	69.21 ± 20.04		
Marino					A significant positive effect of the social robot was found in respect to children’s learning of emotional nature, causes and regulation in relation to specific social contexts.
Test of Emotional Comprehension	Control group: 7	4.14 ± 1.46	5.00 ± 1.31	p = 0.318	
	Robotic group: 7	4.57 ± 1.18	7.29 ± 1.16	p = 0.001***	
Emotional Lexicon Test	Control group: 7	4.71 ± 0.45	5.29 ± 0.452	p = 0.128	
	Robotic group: 7	4.71 ± 0.70	7.00 ± 0.00	p = 0.001***	

Table Supplementary Material 3: RCT studies and respective main results

	Groups	Pre Test Evaluation	Post Test Evaluation	p value: Post Test > Pre Test	Observations and Main conclusions
Srinivasan					The rhythm and comparison groups significant effects on the JTAT from the pretest to the posttest, which was not verified in the robotic group. The lack of adaptability of the context lead to a lack of engagement and boredom. Advances on the autonomy and contingent responding of the robot are required in the future.
Joint Attention Test	Control group: 12	SMD = 0.71 CI(SMD) = -0.01 to 1.43		p = 0.004 ***	
	Rythm group: 12	SMD = 0.55, CI (SMD) = -0.13 to 1.24		p = 0.005 ***	
	Robotic group: 12	SMD = 0.25, CI (SMD) = -0.38 to 0.89		p = 0.27	
Yun					Robotic and standard therapy had similar positive results in terms of eye contact and facial emotion recognition.
Behavioral measurements - Autism diagnostic observation schedule - Korean version of the vineland adaptive behavior scale - Social communication questionnaire - Korean version of the child behavior checklist	Control group: 7			time x group p > 0.05	
	Robotic group: 8				
Frequency of eye contact	Control group: 7	17.14 ± 12.65	73.81 ± 12.79	p = 0.018*	
	Robotic group: 8	20.00 ± 12.18	77.92 ± 18.47	p = 0.012*	
Accuracy of facial emotion expression (session 1 to 4)	Control group: 7	14.29 ± 37.80	88.10 ± 20.89	p = 0.024	
	Robotic group: 8	16.25 ± 25.04	58.21 ± 48.75	p = 0.046	
Accuracy of facial emotion expression (session 4 to 5)	Control group: 7	60.00 ± 44.72	90.00 ± 19.15	p = 0.078	
	Robotic group: 8	55.42 ± 41.13	82.50 ± 34.54	p = 0.104	
Zheng					Significant effects in terms of the joint-attention tasks mediated by the robot and of the interaction in STAT with human were not found, when examining children with ASD compared to the control children. However significant effects were found when constructing subgroups of children: improvement in specifically constructed “improved” subgroup was significant as was decline in constructed “non-improved” subgroups.
Screening Tool for Autism in Toddlers and Young children	Control group: 11	3.18 ± 0.97	3.05 ± 0.80	time x group p = 0.77	
	Robotic group: 12	3.13 ± 1.09	2.80 ± 1.30		
Average prompt level that a participant needed to hit a target in a trial	Control group: 11	2.17 ± 0.71	2.29 ± 0.90	time x group p = 0.97	
	Robotic group: 12	2.39 ± 1.11	2.53 ± 1.12		
Target hit rate	Control group: 11	0.94 ± 0.09	0.86 ± 0.17	time x group p = 0.66	
	Robotic group: 12	0.91 ± 0.13	0.88 ± 0.13		

Table Supplementary Material 3: RCT studies and respective main results

	Groups	Pre Test Evaluation	Post Test Evaluation	p value: Post Test > Pre Test	Observations and Main conclusions
So 2017					Social robots can effectively teach gestural recognition and gestural production to children with low functioning ASD
Number of gestures recognized Training Scenarios	Control group: 6	-	-	p = 0.71	
	Robotic group: 7	-	-	p < 0.001 ***	
Number of gestures recognized Non training scenario	Control group: 6	-	-	p = 0.61	
	Robotic group: 7	-	-	p < 0.001 ***	
Number of gestures recognized Non training scenario with human	Control group: 6	-	-	p = 0.40	
	Robotic group: 7	-	-	p < 0.001 ***	
Number of gestures produced Training Scenarios	Control group: 6	-	-	p < 0.25	
	Robotic group: 7	-	-	p < 0.001 ***	
Number of gestures produced Non training scenario	Control group: 6	-	-	p < 0.76	
	Robotic group: 7	-	-	p < 0.005 ***	
Number of gestures produced Non training scenario with human	Control group: 6	-	-	-	
	Robotic group: 7	-	-	p < 0.16	
So 2018					A robotic therapy based on gestures can reduce the delay in the production of intransitive gestures in young children with ASD when compared with TD children
Number of correctly produced gestures training scenario	Control group: 15	-	-	time x group p < 0.001 ***	
	TD group: 15	-	-		
	Robotic group: 15	-	-		
	TD group vs Control group	p < 0.05 (TD > Control)	NA		
	TD group vs Robotic group	p < 0.04 (TD > Robotic)	p < 0.001 (Robotic > TD)		
	Robotic group vs Control group	NA	p < 0.001 (Robotic > Control)		
Number of correctly produced gestures non training scenario	Control group: 15	-	-	time x group p < 0.003 ***	
	TD group: 15	-	-		
	Robotic group: 15	-	-		
	TD group vs Control group	p < 0.03 (TD > Control)	NA		
	TD group vs Robotic group	p < 0.05 (TD > Robotic)	p < 0.27 (Robotic > TD)		
	Robotic group vs Control group	NA	p < 0.001 (Robotic > Control)		
Number of appropriate gestures - refer to delayed post test	Control group: 15	-	-	group effect p < 0.005 ***	
	TD group: 15	-	-		
	Robotic group: 15	-	-		
	TD group vs Control group	-	p < 0.005 (TD> Control)		
	TD group vs Robotic group	-	p< 0.72 (TD> Robotic)		
Number of gestures with verbal imitation - refer to delayed post test	Control group: 15	-	-	not significant	
	Robotic group: 15	-	-	p < 0.05 ***	

Table Supplementary Material 3: RCT studies and respective main results

	Groups	Pre Test Evaluation	Post Test Evaluation	p value: Post Test > Pre Test	Observations and Main conclusions
So 2019					The group effect in terms of eye gaze was significant, $\beta=.96$, $SE=.24$, being higher in the robotic group. Both robot- and human-based intervention programs are effective in promoting gestural recognition and production skills in children with low-functioning autism. The characteristics of each individual and the relation with the respective effect should be investigated.
Number of correct recognized gestures	Control group: 11 Robotic group: 12	-	-	the time point x group interaction was not significant, $p < .23$.	
Number of produced gestures	Control group: 11 Robotic group: 12	-	-	the time point x group interaction was not significant, $p < .23$.	
So 2020					The robotic intervention enhanced both initiation of joint attention and the play skills of children with ASD.
Initiation of Joint Attention	Control group: 11	-	-	$p < 0.81$	
	Robotic group: 12	-	-	$p < .001$ ***	
Response to Joint Attention	Control group: 11	-	-	Non significant main and interaction effects verified	
	Robotic group: 12	-	-	$p < 0.95$	
Other directed functional play	Control group: 11	-	-	$p < 0.89$	
	Robotic group: 12	-	-	$p < .001$ ***	
Self directed functional play	Control group: 11	-	-	$p < 0.89$	
	Robotic group: 12	-	-	$p < 0.85$	
Symbolic play	Control group: 11 Robotic group: 12	-	-	Non significant effects verified	
Social Responsiveness Scale	Control group: 11	72.72 $\pm$ 26.29	76.27 $\pm$ 29.38	$p < 0.33$	
	Robotic group: 12	70.42 $\pm$ 23.45	61.08 $\pm$ 21.11	$p < 0.11$ ***	