# Analyzing the Effectiveness of Standard Human Treatment (SHT) in Comparison to Robotic Enhanced Therapy (RET)

Hari Menon

## Abstract

This project examines whether humans or robots are better therapists for children with varying degrees of Autism by comparing the Standard Human Treatment (hereafter, SHT) to Robotic Enhanced Therapy (hereafter, RET). An analysis of the standard deviation of eye gaze data from the DREAM Dataset (i.e., a pre-existing experiment that recorded changes in ADOS scores in children based on scripts used in both Human and Robotic intervention methods) revealed that Robotic Enhanced Therapy had a higher standard deviation in eye gaze, showing that the child was more engaged with (and therefore paid more attention to) the robotic therapist in comparison to the human counterpart. This conclusion ultimately serves as an important step forward in the process of finding the most effective intervention method to help children with Autism Spectrum Disorder (hereafter, ASD) manage the associated symptoms.

## Introduction

ASD is a neurodevelopmental condition which as of 2021, impacts 1 in 127 people across all ages, gender, and ethnic backgrounds (World Health Organization [WHO], 2025). This condition is characterized by a spectrum of symptoms including challenges in social communication, restricted interests, and repetitive behaviors (Hamad et al., 2025). ASD can further impose medical consequences. For example, up to 60% of children with ASD have abnormal electroencephalogram (EEG), corresponding to abnormal electrical activity in the brain; 80% of individuals with ASD have sleep disorders, most prevalent in children; 46-84% of children with ASD have gastrointestinal (GI) problems, including chronic constipation and chronic diarrhea; 25% of children with ASD have immune deficiency and dysfunction (Al-Beltagi, 2021). In addition, ASD is associated with genetic and neurological disorders such as atypical connectivity of different brain regions that encompass functions such as social processing, sensory integration, and emotional function (Lord et al., 2018).

It has been shown by previous works that interventions "significantly improves the social, communicative, and daily life skills" of children with ASD (Du et al., 2024). Therefore, it is necessary that effective interventions are developed to help these children. Eye gaze data is a key indicator of whether or not an intervention is effective. In the context of physical activities that require a lot of eye movement, this data can show the ability for an intervention method to engage the child and help them develop these social, communicative, and other such skills through critical thinking.

The present study tests whether robotic therapy is less attention-grabbing than human therapy, and as a result would show a smaller standard deviation of the eye gaze data. This connection can be made since the prompts used, as described below, require the child to engage in certain activities, and hence move their eyes a lot to follow the interaction partner successfully. As a result, a higher standard deviation of the eye gaze data can be correlated with an intervention method being more attention-grabbing. This hypothesis was tested using existing data in the DREAM Dataset (Billing, 2020), where the standard deviation of the eye gaze data was compared across participants that experienced RET (vs. SHT).

The DREAM Dataset is a database that consists of behavioral data collected from children diagnosed with ASD across over 3000 sessions ranging from 3 to 87 minutes, comprising 300 hours of therapy. Half of the children interacted with the social robot NAO, supervised by a therapist. The other half, who constituted the control group, interacted directly with an interaction partner who was also supervised by a therapist. A representation of this setup is depicted below:

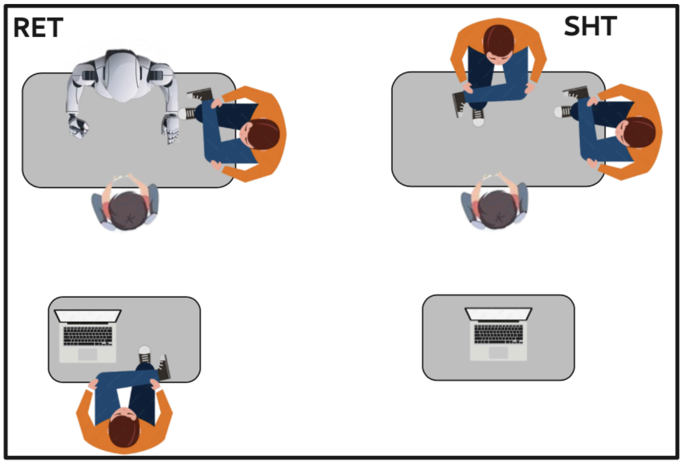


*Figure 1. Illustrates the setup of the SHT Condition and RET Condition. In the RET Condition, a supervisor manages a robot, which serves as the focal interaction partner. In the SHT Condition, the child is provided task scripts by a human interaction partner, with their interactions overlooked by a therapist. Figure adapted from Billing, Erik, et al. "The Dream Dataset: Supporting a Data-Driven Study of Autism Spectrum Disorder and Robot Enhanced Therapy." PloS One, U.S. National Library of Medicine, 21 Aug. 2020, pmc.ncbi.nlm.nih.gov/articles/PMC7444515/.*

Both groups followed the same standard protocol for cognitive behavioral therapy, Applied Behavior Analysis (ABA). This is an intervention method that uses positive reinforcement, specifically a reward, to encourage certain behaviors with the belief that it will increase the individuals' use of that behavior (Billing, 2020). Each session was recorded with three Red-Green-Blue (RGB) cameras and two Red, Green, Blue-Depth (RGBD) cameras (Kinect) which were analyzed with image processing techniques to identify the child's behavior during therapy. The orientation of the dimensions in which the cameras recorded data is shown below:

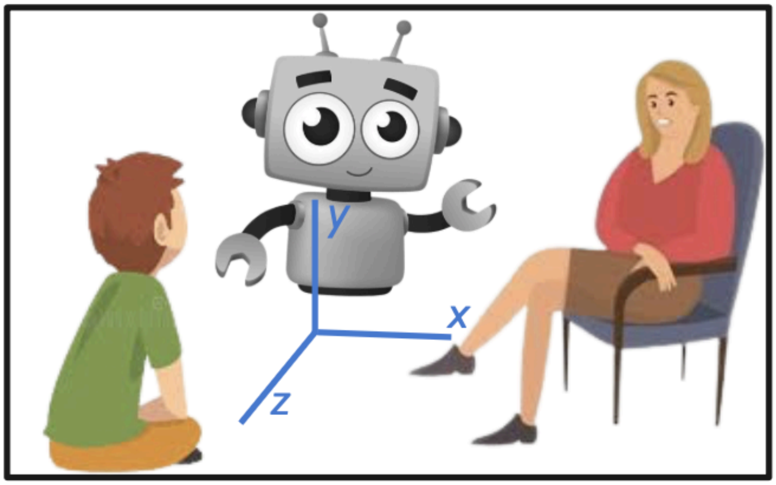


*Figure 2: Demonstrates how eye gaze data was recorded and reported. Specifically, eye tracking data was recorded and reported in 3D space, utilizing the x, y, and z axes to visualize the data. Figure adapted from "DREAM2020/Data: A Public Release of the Dream Dataset, Comprising Anonymous Behavioural Data from Autism Therapy." GitHub, github.com/dream2020/data.*

.

This public version of the database does not contain any recorded video material or other personal data. However, it includes anonymized data describing the child's movements, head position and orientation, and eye movements, all specified in a common coordinate system. Furthermore, descriptive data about the participant in the form of the child's age, gender, autism diagnosis (ADOS), and more details is included.

Each participant began with a diagnostic test to evaluate their ADOS score, a quantitative measure of the severity of behaviors characteristic of ASD. Higher ADOS scores reflect increased severity. Then, eight intervention sessions across the participants were carried out in their assigned condition. In each intervention session, participants followed task scripts that target three social skills, which have been shown to commonly affect individuals on the spectrum, with different levels of difficulty: imitation (Ingersoll, 2008), joint attention (Dawson et al., 2004), and turn-taking in collaborative play (Wimpory et al., 2000). In the end, there is a final intervention session.

Each intervention/therapy session proceeds as follows: the interaction partner (either a robot or human) provides a discriminative stimulus (e.g. an instruction to perform a behavior that pertains to a relevant social skill). Then, the partner waits for a response from the child. If the child's behavior is aligned with expectation, the partner provides positive feedback. If the child's behavior is not aligned with expectation, the partner provides an indication to try again. If the child fails three attempts, a behavioral prompt is used by the interaction partner instead. A behavioral prompt is a form of assistance given to the child to help them perform a behavior they were not able to independently learn.

For example, the interaction partner might ask the child, as a part of the imitation task script, to imitate eating. If the child is not able to imitate that action after three attempts, the partner instead, as a behavioral prompt, opens its mouth widely as a way of giving assistance to the child with the motion of eating.

The following goes into the specifics about the task scripts:

In the imitation task, there are three task scripts. In the first task script, the child is asked to imitate objects (e.g. how to play with a toy car). In the second task script, common gestures (e.g. waving goodbye) are tested. Finally, in the last task script gestures without a significant meaning (e.g. moving the hand in an uncommon gesture) are tested.

There were also three joint-attention task scripts, which vary in difficulty. For the first level of difficulty, the interaction partner points and looks at an object while giving a verbal cue (e.g. “look”). In the second level of difficulty, the partner points and looks at an object without giving a verbal cue. Finally, in the third level of difficulty, the interaction partner simply looks at an object.

Finally, for turn-taking tasks scripts, there were two levels of difficulty per task. The first task was sharing information that one likes based on five simultaneously displayed pictograms, and the levels depended upon the complexity of the pictograms. The second task was categorizing objects. In the first level of difficulty,one object was displayed at a time. In the second level of difficulty, eight objects were displayed at the same time, and the child had to categorize one of them correctly. The third task was completing a series of pictures in a pattern, in which the first level of difficulty was choosing one of two pictures that continued said pattern and the second level of difficulty was where one of four pictures had to have been chosen by the child to continue the provided pattern.

## Methodology

The eye gaze data was downloaded from the GitHub page for the DREAM Dataset in .json files (Billing, 2020). From here, the available data from the final intervention sessions was isolated from participants and put on a table:

| | participant | condition | task | test_phase_clean | rx | ry | rz |
|---|---|---|---|---|---|---|---|
| 0 | 3 | SHT | TT | final | 0.404704 | 0.247085 | 0.206432 |
| 1 | 4 | RET | IM | final | 0.424303 | 0.344187 | 0.297862 |
| 2 | 4 | RET | JA | final | 0.382371 | 0.315656 | 0.256236 |
| 3 | 4 | RET | TT | final | 0.432072 | 0.234277 | 0.222070 |
| 4 | 5 | RET | IM | final | 0.347938 | 0.206577 | 0.219314 |
| 5 | 5 | RET | JA | final | 0.433943 | 0.266127 | 0.261098 |
| 6 | 5 | RET | TT | final | 0.384158 | 0.264066 | 0.207673 |
| 7 | 8 | SHT | IM | final | 0.321460 | 0.210945 | 0.134044 |
| 8 | 8 | SHT | JA | final | 0.367391 | 0.253160 | 0.179187 |
| 9 | 8 | SHT | TT | final | 0.359239 | 0.257901 | 0.197960 |
| 10 | 10 | RET | IM | final | 0.303737 | 0.275165 | 0.176881 |
| 11 | 10 | RET | JA | final | 0.294434 | 0.287932 | 0.213947 |
| 12 | 10 | RET | TT | final | 0.339127 | 0.230829 | 0.192712 |

*Figure 3: Small sample of data isolated from participants for which it was available. The "participant" refers to the specific ID assigned to each participant in the DREAM study. TT stands for the "Turn Taking" task script, IM stands for the "Imitation" task script, and "JA" stands for the "Joint-Attention" task script. The "test_phase_clean" is used to ensure that data from no other intervention but the final is included. The "rx", "ry", and "rz" refers to the standard deviations across the x, y, and z axes, respectively.*

Of the provided variables, only the participant number, condition, task, test phase ("test_phase_clean"), and the standard deviations across all axes was included. The rest were ignored since they were either deemed irrelevant to the results of determining whether SHT or RET is more engaging (e.g., age) or could not be effectively controlled/accounted for (e.g. ADOS score).

## Results

To compare whether the standard deviation of eye gaze was larger for SHT or RET, a linear regression was conducted using the standard deviation of eye gaze data as the dependent variable and condition as the independent variable. This analysis was conducted for each axis of the provided eye gaze data in the DREAM Dataset.

Contrary to the predicted effect, the bar chart among the x-axis revealed that the standard deviation of eye gaze was significantly greater for RET in comparison to SHT for the Turn-Taking and Imitation task script. Figure 4 depicts these results. This was verified quantitatively using a linear regression that modeled the standard deviation of the eye gaze in the x dimension as a function of the SHT or RET condition and task used. I found that the standard deviation of the eye gaze data in the x dimension was lower for SHT relative to the RET condition, $t = -3.1$, $p = .004$, $B = -0.05$, 95% $CI = [-0.086, -0.017]$. However, there was no significant difference between the standard deviation of the eye gaze data between the two conditions for the Join-Attention task script.

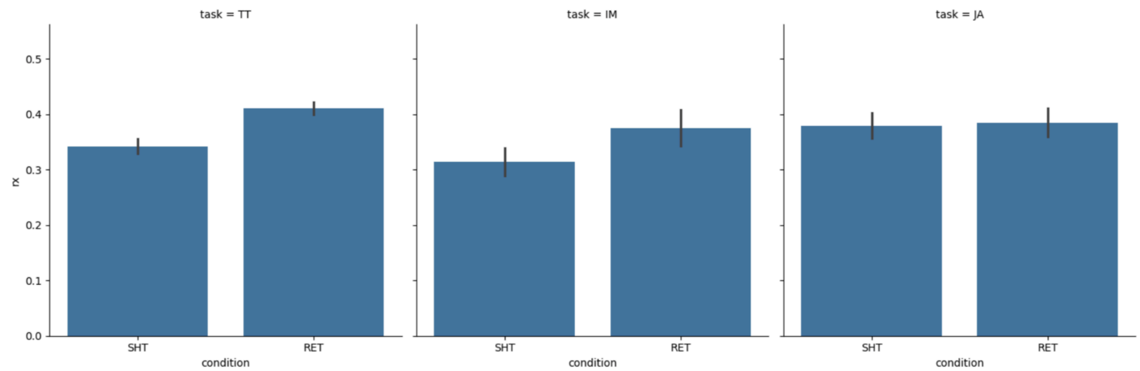


*Figure 4: Shows the standard deviation of the eye gaze data for each type of task script on the x-axis across participants in their final intervention sessions. Illustrates SEM bars to account for statistical significance. TT stands for the "Turn Taking" task script, IM stands for the "Imitation" task script, and "JA" stands for the "Joint-Attention" task script.*

Similarly, the bar graph among the y-axis revealed that the standard deviation of eye gaze was significantly greater for RET in comparison to SHT for the Turn-Taking task script. Figure 5 depicts these results. This was verified quantitatively using a linear regression that modeled the standard deviation of the eye gaze in the y dimension as a function of the SHT or RET condition and task used. I found that the standard deviation of the eye gaze data in the y dimension was lower for the SHT relative to the RET condition, $t$ = -2.2, $p$ = .035, $B$ = -0.027, 95% $CI$ = [-0.052, -0.002]. However, there was no significant difference between the standard deviation of the eye gaze data between the two conditions for the Join-Attention and Imitation task scripts.

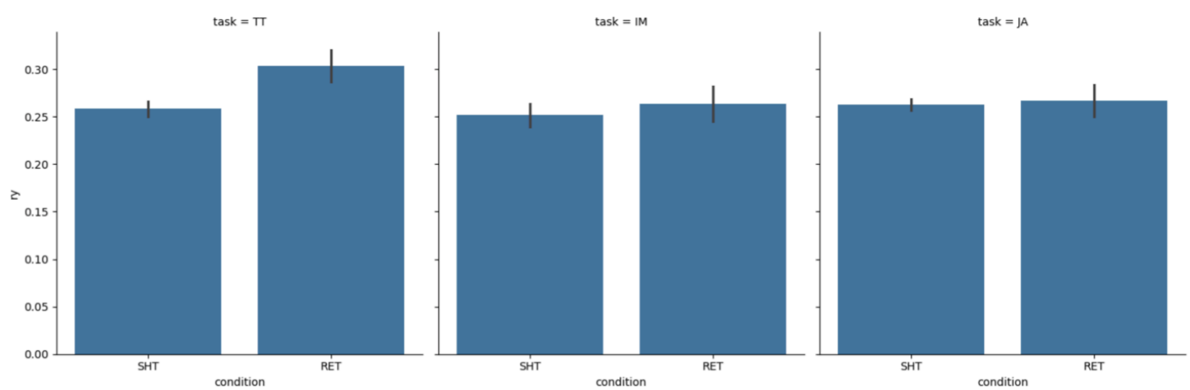


*Figure 5: Shows the standard deviation of the eye gaze data for each type of task script on the y-axis across participants in their final intervention sessions. Illustrates SEM bars to account for statistical significance. TT stands for the "Turn Taking" task script, IM stands for the "Imitation" task script, and "JA" stands for the "Joint-Attention" task script.*

Finally, the bar graph among the z-axis revealed that the standard deviation of eye gaze was significantly greater for RET in comparison to SHT for the Turn-Taking and Imitation task scripts. Figure 6 depicts these results. This was verified quantitatively using a linear regression that modeled the standard deviation of the eye gaze in the z dimension as a function of the SHT or RET condition and task used. I found that the standard deviation of the eye gaze data in the z dimension was lower for SHT relative to the RET condition, $t = -2.6$, $p = .011$, $B = -0.041$, 95% $CI = [-0.071, -0.010]$. However, there was no significant difference between the standard deviation of the eye gaze data between the two conditions for the Join-Attention task script.

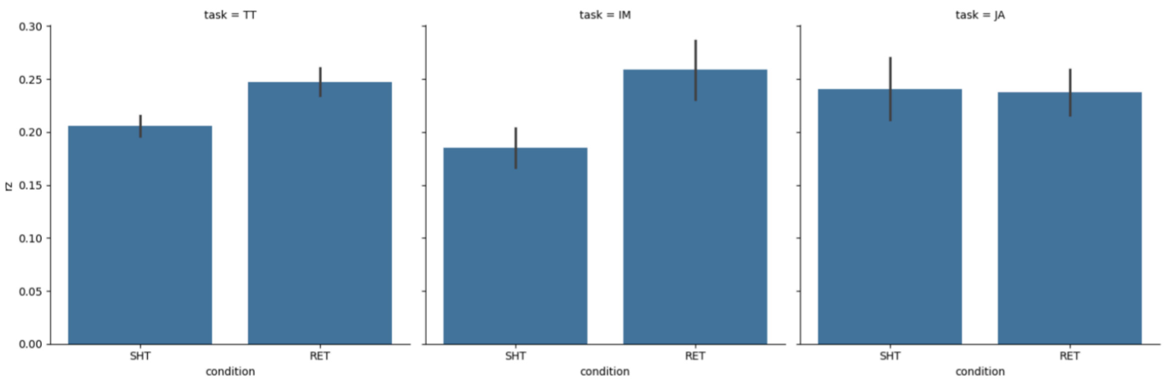


*Figure 6: Shows the standard deviation of the eye gaze data for each type of task script on the z-axis across participants in their final intervention sessions. Illustrates SEM bars to account for statistical significance. TT stands for the "Turn Taking" task script, IM stands for the "Imitation" task script, and "JA" stands for the "Joint-Attention" task script.*

In all, participants exhibited greater standard deviation of eye gaze across all axes for the "turn-taking" task script in the RET Condition. Similarly, for the "imitation" task script, RET has a significantly greater standard deviation of eye gaze among participants across the x-axis and z-axis. However, for the "joint-attention" task scripts, there is no statistically significant difference between SHT and RET across all axes.

This qualitative and quantitative analysis shows that, largely, RET has a significantly greater standard deviation of eye gaze across participants in their final intervention sessions.

## Discussion

The results of this do not support the hypothesis that the RET intervention method is less attention-grabbing than SHT. Robots could serve as better therapists than robots for many reasons, including that robots do not pressure children with autism to maintain eye contact (unlike with humans), are more adaptable due to their ability to collect quantitative data, and are customizable to adapt to individual children.

Although these findings were contrary to the original hypothesis, there are several caveats to be considered that were not addressed in this study due to the lack of resources to be able to do so accurately. Firstly, standard deviation only serves as an indirect approximation of engagement. For example, it is possible that the child was distracted, causing the standard deviation of the eye gaze to increase. This increase would ultimately serve as a contradiction to the correlation of a high standard deviation to high engagement and attention to the task script. Secondly, standard deviation of the eye gaze does not provide any conclusion about task performance or real-world outcomes. In other words, even if the child is engaged with the interaction partner as per the high standard deviation of the eye gaze across their session, this does not mean that the child will execute the task script correctly.

There are, however, next steps that could be taken in response to these caveats and as a way to take this study further and identify effective intervention methods. First, there could be a control for session duration. As seen in Figure 7, the session duration is not the same for both RET and SHT across all interventions. RET's longer session duration is likely caused by there being more attempts required for the task scripts in this method in comparison to SHT. Regardless, this difference could severely skew the results. In this case, this skew would be caused by providing children who participated in RET more of a chance to improve than those in SHT. This could ultimately shift the results by causing the standard deviation of the eye gaze in RET intervention sessions to be higher by giving the child more of a chance to engage with the robotic interaction partner in comparison to the human counterpart. A figure of the difference in session duration is shown below:

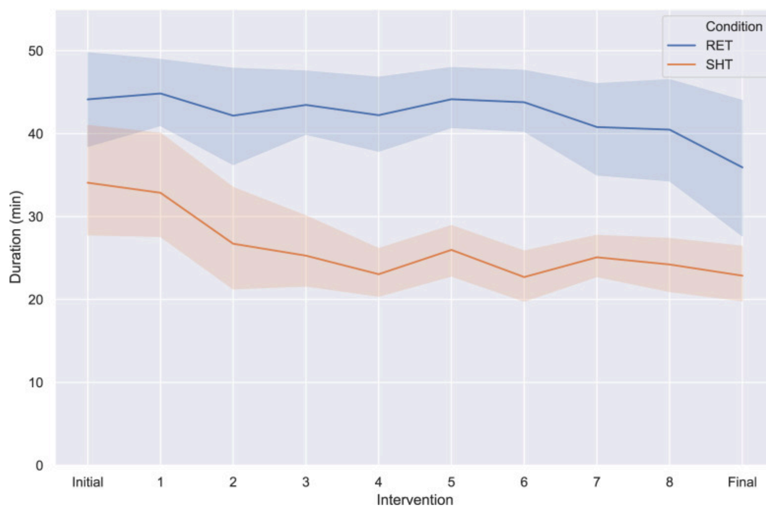


*Figure 7: Demonstrates the average session duration across each type of intervention session (initial, eight intermediary, and final). Figure adapted fromBilling, Erik, et al. "The Dream Dataset: Supporting a Data-Driven Study of Autism Spectrum Disorder and Robot Enhanced Therapy." PloS One, U.S. National Library of Medicine, 21 Aug. 2020, pmc.ncbi.nlm.nih.gov/articles/PMC7444515/.*

Next, the correlation between ADOS score and standard deviation of the eye gaze data could be analyzed. This is because a higher ADOS score is correlated with increased severity of behaviors characteristic of ASD and, as a result, would make it harder for children with higher scores to maintain their attention. Therefore, the standard deviation of their eye gaze (not taking into account distractions) could be far lower. Specifically, the higher severity of behaviors characteristic of ASD would reduce the ability to maintain attention. As a result, the standard deviation of the eye gaze would be lower since the child with the high ADOS score will be less able to remain engaged in the task script and will look elsewhere. Hence, it is necessary to take into account the average ADOS score of children in each intervention method and how potential differences between these intervention methods may skew the results.

Additionally, the change in the standard deviation of the eye gaze across the eight intervention sessions could be analyzed. Specifically, whether the robot or human interaction partner is able to stimulate engagement faster across the intervention sessions could be analyzed by looking at how much the standard deviation of the eye gaze changes from session to session. This would be a very effective analysis to conduct to arrive at a more concrete answer of whether robots or humans serve as better therapists by stimulating growth in children with ASD at a faster rate. Based on the standard deviation of eye gaze and how higher values can be correlated with increased engagement, it can be postulated that human therapists, due to their ability to retain attention of the child, will stimulate faster growth than their robotic counterparts.

Finally, to further measure the effectiveness of SHT in comparison to RET, analysis of task performance (change in the number of attempts required across the eight intervention sessions) could also be taken into account. In other words, including how task performance changes, which could be measured by understanding how the number of attempts required by the child changes across the eight intervention sessions, could further push a conclusion on whether SHT or RET is a better intervention method.

Understanding whether humans or robots might serve as better therapists is the first step to narrowing down to better intervention methods. This is because, as mentioned earlier, the debate over whether human or robot therapists are better has largely not been settled yet. Hence, by doing so, a barrier in producing effective interventions can be removed. This would allow further studies to be conducted to narrow down specific methods within types of interventions that could serve as the most effective way of combating the symptoms associated with ASD.

Real-world applications of better intervention methods include school curriculums, training in the workplace, and parenting. Schools could use better intervention methods as a means of helping children with ASD, and related conditions that lead to similar symptoms, learn the concepts taught better. It could also be applied in the workplace to enable employees who experience the symptoms of ASD to be able to enhance their skills and contribute more meaningfully. Parents could also use better intervention methods as a way of getting themselves involved in the child’s life and teaching them to handle the symptoms of ASD at an early age, potentially enabling the child to overcome these symptoms.

## Bibliography

Al-Beltagi, M. (2021). Autism medical comorbidities. *World Journal of Clinical Pediatrics*, *10*(3). https://doi.org/10.5409/wjcp.v10.i3.15

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). https://doi.org/10.1037/0012-1649.40.2.271

*DREAM2020/data: A public release of the dream dataset, comprising anonymous behavioural data from autism therapy*. GitHub. (n.d.). https://github.com/dream2020/data

Du, G., Guo, Y., & Xu, W. (2024, October 17). *The effectiveness of applied behavior analysis program training on enhancing autistic children's emotional-social skills - BMC psychology*. SpringerLink. https://link.springer.com/article/10.1186/s40359-024-02045-5

Hamad, N. I., Bakry, S. G., & Moawad, O. G. (2025). Quality of life of mothers having children with autism spectrum disorder. *Alexandria Scientific Nursing Journal*, *27*(1), 108–120. https://doi.org/10.21608/asalexu.2025.414958

Ingersoll, B. (2008). The effect of context on imitation skills in children with autism. *Research in Autism Spectrum Disorders*, *2*(2), 332–340. https://doi.org/10.1016/j.rasd.2007.08.003

Lord, C., Elsabbagh, M., Baird, G., & Veenstra-Vanderweele, J. (2018). Autism spectrum disorder. *The Lancet*, *392*(10146). https://doi.org/10.1016/s0140-6736(18)31129-2

Wimpory, D., Hobson, R., Williams, J., & Nash, S. (2000). Are infants with autism socially engaged? A study of recent retrospective parental reports. *Journal of Autism and Developmental Disorders*, *30*(6). https://doi.org/10.1023/a:1005683209438

World Health Organization. (2025, September 17). *Autism*. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/autism-spectrum-disorders