## Supplementary Materials

*Outcome measures*

*Developmental profiles: Griffiths Mental Development Scales - Edition Revised (GMDS-ER)*

The Griffith Mental Development Scales-Edition Revised (GMDS-ER, (Luiz et al., 2006)) are standardized scales with satisfactory validity and reliability administered to the child by trained psychologists in a laboratory setting through semi-structured activities designed to evaluate different aspects of mental development in infants and children. The scale provides standardized Z scores (M = 100; SD = 15) relative to 6 developmental domains: Locomotion; Personal-Social; Hearing and Language; Eye-Hand Coordination; Performance, and Practical Reasoning. The subscales provide a General Quotient (GQ) and a Developmental Age-Equivalent (measured in months) that allow the comparison with what would be expected during TD and to track trajectories considering the time elapsed in-between the reference assessments. Further, since each subscale is independent from the others, it is also possible to evaluate the degree of homogeneity of developmental profiles between the different developmental dimensions, and within each of them, since the tasks proposed to the child are arranged in an age-related sequence. Therefore, it is possible to qualitatively assess whether a child shows, for example, good competencies in some tasks for his/her age, but still misses some ground abilities that are typically acquired earlier in development in the TD population. These aspects are particularly relevant to design an effective personalized intervention tailored to the specific needs of every single child.

*Measuring trajectories of change over time: the Learning Rate (LR)*

The main developmental outcome measure used as a dependent variable for the study of treatment response trajectories consisted in the Learning Rate (LR, (Klintwall et al., 2015), i.e., *the ratio between the variation in Developmental age-equivalents over time and the time elapsed between the assessments*. Individual LRs are particularly suitable to assess changes over time with respect to TD trajectories, also in the presence of differences with respect to the intervals of time elapsed between the assessments, as well as baseline chronological ages. Confronting ASD response trajectories with TD in terms of developmental outcomes is crucial, since it is expected that autistic children may exhibit a certain degree of spontaneous acquisitions with respect to developmental milestones, regardless of intervention. Thus, employing LRs allows the assessment of whether, during the time interval of intervention, the developmental trajectory reflects the gains expected during TD, or whether the child is narrowing (or widening) the gap between developmental and chronological age, taking into account the effect of time. LRs are computed between the GMDS-ER Developmental age-equivalents at baseline (TPre) and at the end of treatment (TPost):

$$LR = \frac{DevAge_{\text{TPost}} - DevAge_{\text{TPre}}}{TPost - TPre}. \tag{1}$$

Mathematically, the LRs represent the slope of the line that intersect two points on the plane with coordinates reflecting chronological age on the first axis and developmental age on the second axis. A LR = 1 represents changes over time that resemble a TD trajectory,

i.e. for each chronological month in the considered time interval, the child increased the developmental age equivalent by the same amount. Therefore, A LR $>$ 1 indicates that the child is narrowing the gap between mental and chronological age, i.e. the child gained more developmental months than the chronological months elapsed. Finally, a LR $<$ indicates an increasing developmental delay, and may reflect situations where the child may probably respond less or at all, an important risk condition that should raise clinician awareness, and that probably indicates that something in the intervention for that particular child is not working as expected. LRs can also account for the variability in developmental trajectories generally observed also in TD (Klintwall et al., 2015).

Employing LRs for treatment monitoring can therefore also help to promptly introduce the necessary changes in intervention design, objectives, or modalities (Eikeseth et al., 2012). LRs may also be preferable to standardized developmental measures for different reasons. First, children who learn slowly may have decreasing standard scores. This is because, especially in early development, TD children increase fast in their standard score, and reference norms frequently change based on chronological age, making it more difficult to obtain the same score. This may in turn lead to misinterpret this situation as a developmental regression, whereas in this situation a child would present a LR $<$ than 1, indicating an increasing developmental delay without neglecting the progression (even if small) he showed in the reference time interval.

A main disadvantage of standard scores with respect to intervention is that they may mask differences in changes based on the initial level of functioning. For example, research supports the idea that children may differentially respond to intervention, based on a variety of features like, for instance, the developmental age-equivalent at intake. However, given the differences in chronological ages, such differential response may be masked. LRs and developmental trajectories may be useful and numerically consistent for outcome studies aiming to evaluate not only efficacy of interventions and to monitor developmental trajectories, but also to quantitatively investigate predictive factors, as well as mediators of the process and the specific trajectories of change (Eikeseth et al., 2012). In fact, from a clinical standpoint, it is fundamental to shed light on the process of change as the ultimate goal to promote and support child development in the best way possible. As a main limitation, age-equivalents may be more prone to the risk of non-normal distributions, challenging their interpretation in research studies (Maloney & Larrivee, 2007). However, considering the age-equivalent / age-chronological ratio may help in mitigating this effect.

*Symptom severity: Autism Diagnostic Observation Schedule - Second Edition (ADOS-2)*

The Autism Diagnostic Observation Schedule-2 (ADOS-2, (Lord et al., 2012)) is a golden standard instrument for the diagnosis of ASD and the assessment of symptoms severity with high validity and reliability. The administration of this tool is carried out by trained psychologists. The instrument provides 5 different modules according to the child's chronological age (from 18 months to adulthood) and expressive level of language (from no-words to fluent language). In this study, modules Toddler, 1, and 2 were used, according to the age and the linguistic level of each child. Each module gives a Total Score for the autism–autism spectrum–non spectrum classification. For the purpose of this study, we considered the

ADOS-2 Social Affect (SA) and the Restricted Repetitive Behaviors (RRB) scores separately, since they provide specific quantitative information from the two main clusters of ASD core symptoms (Association, 2013). For diagnostic purposes, the instrument also provides a Comparison Score derived from the Total Score, to compare different modules. We referred to the raw scores since they provide more variable and specific numerical information about the intensity of symptoms. Finally, for the administration of the ADOS-2 clinicians need to undergo an official course ensuring the reliable administration of the instrument.

*Process measures*

*Interaction Coding System (ICS)*

The Interaction Coding System (ICS, (Bertamini et al., 2021)) is a quantitative observational coding scheme aiming to characterize child-therapist interactions during video-recorded sessions of intervention. The rationale behind the idea concerns the centrality of the child-therapist interpersonal relationship during therapy. In fact, the entire process of intervention unfolds fully immersed in the dyadic interaction, which represents the main mediator of the process and the base for scaffolding working objectives to promote learning. However, interaction and relationship based aspects are still under-investigated in the context of Autism intervention. In our perspective, the investigation of the child-therapist interaction may be important to shed light on key aspects of a successful intervention. To argue this hypothesis we refer mainly to two lines of evidence. First, experience-mediated relationship-dependent learning is fundamental for child-development, as a multitude of developmental psychology studies demonstrated in the context of early infant-caregiver interactions and emotional communication (Feldman, 2017, 2007; Siegel, 2001). Further, patient-clinician relational aspects have been already investigated in other clinical contexts like psychotherapy showing, through rigorous quantitative methods, that constructs like therapeutic alliance and interpersonal synchrony have an impact on therapeutic outcomes (Kang et al., 2021; Goldstein et al., 2020; Flückiger et al., 2018; Koole & Tschacher, 2016; Green, 2006). Recently, Naturalistic Developmental Behavioral Interventions (NDBIs, (Schreibman et al., 2015)) emphasized the role of interactional aspects in Autism intervention, and current research identified broad constructs and process aspects, mostly evaluated through qualitative observational grids, that may impact outcomes (Frost et al., 2020). The child-therapist interaction is schematized in Figure 1.

The Interactional Coding System (ICS) focuses on a child's intentionality and social motivation, and on therapist's responsiveness, dyadic reciprocal adaptation and cohesion. The aim is to detect complex relationships between temporally related events characterized by multiple dimensions. The single codes refer to social units that can be multimodally coordinated and expressed by different intersubjective channels. Further, the code aims to produce data for quantitative analysis in terms of functional patterns, durations and latencies. The scheme has been designed to catch the bidirectional nature of the interchange in terms of structure and dynamics. The set of behaviors consists of 15 observational codes (13 point events and one goal state with a numeric attribute) that describe the interplay as modeled in two interconnected and recurrent scenarios: Interaction Units (IUs) and the Shared Activity (SA) state. IUs provide an annotation set mainly focused on the initial phase of agreement

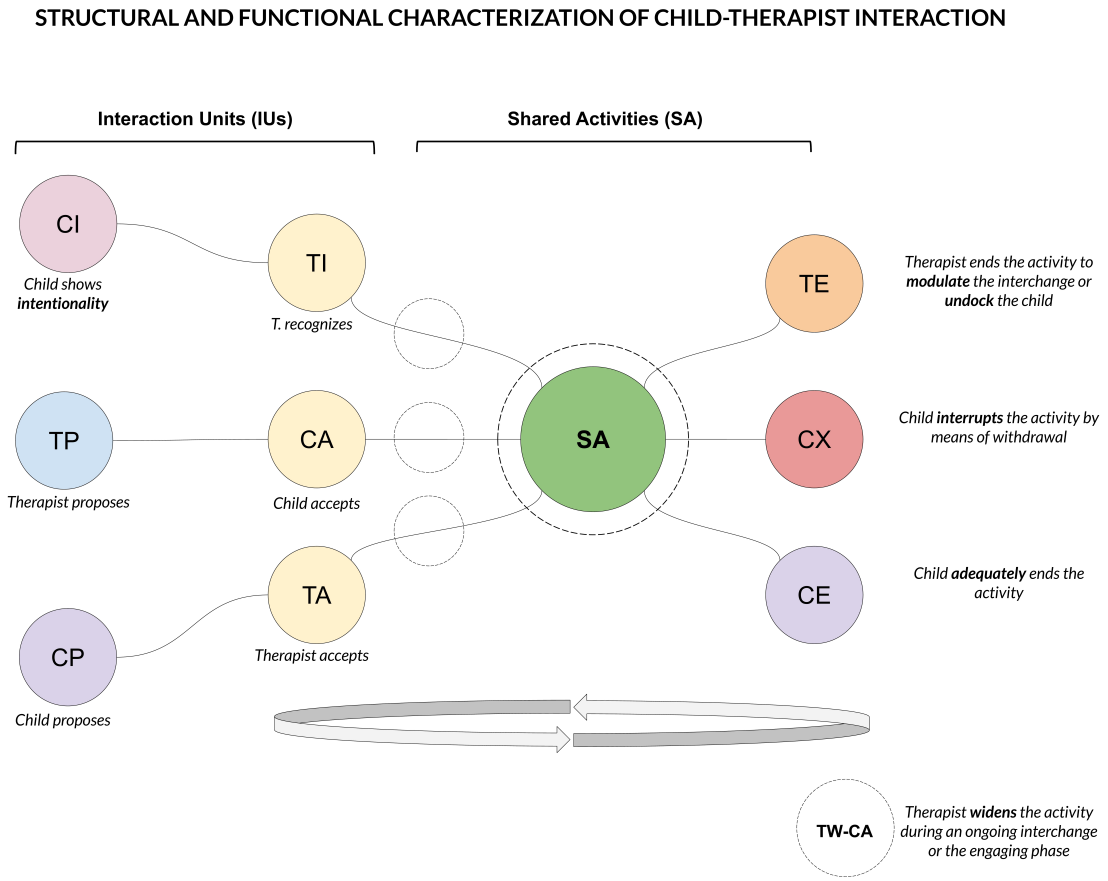


Figure 1: **Structural and functional characterization of the child-therapist interaction** - The child-therapist interplay is schematized by means of an engagement precursor, consisting in different types of Units of Interaction (UIs), like therapist's or child's proposals and child's intentionality signals, that may lead towards the onset of a sustained and reciprocal interpersonal exchange (Shared Activity state). In turn, the interplay can end by means of different scenarios, including dyadic agreement, therapist's decision, and child's withdrawal

between the social partners in terms of different types of proposals and consist of 3 couples of paired codes. From such states, child and therapist may achieve the onset of a SA that presents the characteristics of interpersonal synchrony. Therefore, such sequences of IUs may represent building blocks to scaffold and start the interaction, and may be a useful representation for the study of both interaction structure and dynamics. Additionally, three codes describe the conclusion of the shared activity. The annotations identify different types of typical social behaviors and signals mutually exchanged during the session, e.g., the therapist tries to involve the child by proposing a play or social routine (TP: therapist proposes). In turn, the child can accept (CA: child accepts), refuse (CR: child refuses) or ignore these attempts (no code). Another interactive scenario concerns child's intentionality (CI: child intentionality), which often represents the starting point to scaffold a play routine and it is important for the therapist to catch (TI: therapist intentionality). Therefore, the IUs assemble themselves in sequences of start and response behaviors, which are the structural patterns giving birth to the actual interplay (e.g., TP-CA, CI-TI and CP-TA). They reflect the reciprocal dynamic of the interplay in terms of engagement, involvement and agreement between the dyad through interactive patterns, which can be evaluated with specific metrics in terms of type, duration, latency and outcome. Notably, beside structure and dynamics, the coding scheme includes annotations to catch a child's emotional reactions to the degree of social stimulation, like signals of dysregulation (CD: child dysregulation) in response to therapist attempts of engagement or during intensively playful activities. Symmetrically, the ability of the therapist to recognize the child's emotional states is coded (TR: therapist recognizes child's emotional state). To design the scheme, we referred to the literature in developmental research that highlighted the importance of early interactions for infant development and in ASD, and in TD. (Association, 2013; Leclère et al., 2014; Feldman, 2007). As well, NDBI interventions integrate the developmental perspective in the design process of ASD early interventions focusing on the quality of the interaction, incidental learning, intrinsic motivation and shared affective states (Minjarez et al., 2020; Wong et al., 2019; Vivanti et al., 2020; Mazefsky et al., 2013; Zwaigenbaum et al., 2015). Behavioral events and Intraclass Correlation Coefficients (ICCs) are summarized in Table 1. Cohen's k coefficient for inter-rater reliability indicated substantial reliability and ranged between 0.73 and 0.85 (M = 0.78; SD = 0.04). In the initial validation phase, the ICS showed good reliability and both construct convergent validity and predictive validity, with different behavioral descriptors significantly correlated with the Emotional Availability Scales dimensions of child's involvment and responsiveness (Biringen, 2008; Bertamini et al., 2021).

Table 1: The Interaction Coding System (ICS): List of behavioral events and Intraclass Correlation Coefficient (ICC)

| Code | Description | ICC (alternative hypothesis r0 > 0.8) |
|---|---|---|
| **TP** | **Therapist proposes** | **0.958 F(9,5.100) = 5.400; p = 0.038 [0.764 - 0.990]** |
| **TW** | **Therapist widens** | **0.953 F(9,9.090) = 4.620; p = 0.016 [0.824 - 0.988]** |

| | | |
|---|---|---|
| **CA** | **Child accepts** | **0.952 F(9,5.580) = 4.600; p = 0.039 [0.765-0.989]** |
| CR | Child refuses | 0.627 F(9,9.49) = 0.477; p = 0.858 [0.096 - 0.889] |
| **CI** | **Child's intentionality signal** | **0.940 F(9,7.990) = 3.640; p = 0.042 [0.766 - 0.985]** |
| **TI** | **Therapist recognizes intentionality** | **0.957 F(9,5.820) = 5.250; p = 0.030 [0.786 - 0.990]** |
| CP | Child's proposes | 1 |
| TA | Therapist shares | 1 |
| SA | Social interplay / Shared activity | 1 |
| **CX** | **Child's withdrawal** | **0.992 F(9,10) = 28.800; p < 0.001 [0.971 - 0.998]** |
| TE | Therapist ends activity | 1 |
| **CE** | **Child ends activity** | **0.975 F(9,10) = 8.700; p = 0.001 [0.908 - 0.994]** |
| CD | Child's emotional dysregulation | 0.571 F(9,9.870) = 0.401; p = 0.907 [0.014 - 0.869] |
| TR | Therapist recognizes child's emotional state | 0.667 F(9,9.960) = 0.556; p = 0.805 [0.138-0.904] |
| ENG | 1 - low engagement | 0.943 F(9,5.91) = 3.9; p = 0.057 |
| | 2 - medium engagement | 0.741 F(9,6.56) = 0.738; p = 0.672 |
| | 3 - high engagement | 1 |

The IUs mostly represent precursors of the actual social exchange, and may or may not start a social routine. If engagement results in being successful (e.g., TP-CA, CI-TI and CP-TA), the therapist scaffolds and modulates the interaction with the child, coordinating actions and sharing affective states. During SA states the social partners are engaged in behavioral patterns that have to be reciprocal and coordinated in order to maintain the shared experience, and thus show some degree of interpersonal synchrony. SAs require the onset of a social or play routine that actively involves both the child and the therapist. They could be object-mediated, e.g., throwing a ball to each other in turns, or either involve primary-intersubjectivity exchanges, e.g., a tickling routine in which the child asks for continuation. Therefore, the key element to code for shared activities is the observation of an established routine participated and modulated by both the partners. Social exchanges are also characterized by an internal dynamic in which the therapist modulates the activity to promote children's social abilities by widening the interplay (TW: therapist widens). If the child accepts (CA) and if the therapist's widenings are proximal to the child's abilities, the complexity of the interchange and its social demands may progressively increase during the shared experience. In general, the exchange eventually ends by a decision of the

therapist (TE: therapist ends) or after an agreement with the child to support his or her communicative intentionality. Otherwise, the child may be able to autonomously request the end of the exchange with adequate communicative modalities addressed towards the social partner (CE: child ends). Finally, the exchange may also be unilaterally disrupted by child withdrawal (CX: child's withdrawal). Measures based on these three alternatives may be relevant to monitor longitudinal changes in child's and dyadic interactive behavior. The child may also show different levels of engagement (ENG): low (1), moderate (2) and high (3). The rate is assigned based on the involvement of the child in the activity, the degree of active participation and motivation through verbal and non-verbal cues and shared affect.

*Naturalistic Developmental Behavioral Intervention*

The intervention applied by the ODFLab integrates empirically validated scientific principles in a developmental frame, together with guidelines in accordance with the Public Health System (Ministero della Salute, 2011). , and adheres to the NDBI framework (Wong et al., 2019; Schreibman et al., 2015). It is individualized, comprehensive and integrates behavioral, developmental and relationship-based principles. In order to foster intentionality and reciprocity, the aim of the therapist in this intervention protocol is to create a pleasant relationship and establish intersubjective routines by starting from the child's own pleasure, behavior, and intrinsic motivation to scaffold exchanges based on shift alternation. It also promotes Intentionality, by giving to the child's behavior a communicative value so that he/she experiments that an action influences others behavior to foster reciprocity. In fact, the key aspect is providing a communicative meaning to children's behaviors, which is stronger when interactive situations are started from their own choice or based on shift alternation (Vivanti et al., 2020). Besides a specific work focused on restoring effective interactions and promoting emotional communication (Feldman, 2007), the intervention protocol aims at the acquisition of specific functional competencies through playful psychoeducational activities such as cognitive activities and emotional, social, and symbolic play. During the intervention sessions the therapist constantly monitors intervention goals, updating them on the basis of the child's developmental improvements, and gradually increasing the complexity in terms of both cognitive loads and social requests (Vivanti & Zhong, 2020; Apicella et al., 2013). During the intervention sessions, the therapist constantly monitors intervention objectives with reference to a specific observational grid administered at regular time intervals. Objectives are regularly updated based on the child's developmental improvements, and their complexity is gradually increased in terms of both cognitive loads and social requests, reflecting developmental milestones and emerging abilities. All the therapists are regularly supervised by expert psychotherapists to guarantee their adherence to the intervention protocol, to the specific objectives defined for the single patient, and to assess whether the objectives are adequate for the children's current developmental level. The intervention was delivered by licensed psychologists after receiving specific training on developmental models of intervention for autistic children. All the psychologists have completed at least the introductory course to the Early Start Denver Model (Fuller & Kaiser, 2020), and/or JASPER (Waddington et al., 2021), and have more than one year of professional experience in Autism intervention for preschool children.

*Results*

*Variable selection*

After computing all the variations, variables were standardized (N = 133, i.e., 13 baseline variables excluding the dependent variable, plus 4*12 interaction variables measured at the 4 time points, and 6*12 computed in-between variations) using the robust standardization described in the statistical analysis. At this stage, N = 8 variables were excluded from the analysis because of numerical problems (variations that were set to infinite after the computation). Table 2 reports the initial feature set.

Table 2: Initial feature set

| **INITIAL FEATURE SET** | |
|---|---|
| **Feature** | **Description** |
| **BASELINE VARIABLES** | |
| Age (months) | Chronological age |
| Developmental age (months) | GMDS-ER Developmental age |
| rate_DC | Developmental / Chronological age ratio |
| Time (months) | Time elapsed between the two clinical assessment |
| Hpw (hours) | Hours per week, intervention intensity |
| Q | GMDS-ER Developmental quotient |
| QA | GMDS-ER Locomotor Quotient |
| QB | GMDS-ER Personal-Social Quotient |
| QC | GMDS-ER Hearing and Language Quotient |
| QD | GMDS-ER Eye-Hand Coordination Quotient |
| QE | GMDS-ER Performance Quotient |
| AS | ADOS-2 Social Affect Score |
| CRR | ADOS-2 Restricted Repetitive Behavior Score |
| TOT | ADOS-2 Total Score |
| **INTERACTION FEATURES** | |
| P_CX | Proportional frequency of child's withdrawals |
| R_SYNC | Rate of synchrony code pairs over total code pairs |
| R_CX | Rate of shared activities interrupted by child's withdrawal |
| R_SA_TPCA | Rate of therapist's proposals that actually allowed for the initiation of a shared activity |
| LATENCY_SA (s) | Mean latency between IUs and the actual start of the shared activity |
| LATENCY_TPCA (s) | Mean latency between therapist's proposals and child's acceptances |
| R_TPCA | Rate of therapist's proposals over the total number of IUs |
| DURATION_SA (s) | Mean duration of shared activities |
| P_CA | Proportional frequency of child's acceptance |
| LATENCY_SYNC (s) | Mean latency between synchrony codes |
| ENG_SA | Mean engagement level during shared activities |
| R_SA | Rate of IUs that actually led to the initiation of a shared activity |

Interaction features are computed at four time points (T0, T1, T2, T3). Variations between time points were also calculated (T1-T0, T2-T1, T3-T2, T2-T0, T3-T0)

The initial set of candidate predictors consisted in N = 125 variables.

*Selection between correlated variables*

The first variable selection step was based on eliminating highly correlated predictors based on Pearson's correlation coefficient. At first, the correlation matrix between all the variables was computed. Afterwards, variables with $\mid r \mid > 0.55$ were excluded with a conservative threshold. Coherently with our predictive aims, we prioritized variables measured at earlier time points in the selection. To investigate variations over time in behavioral descriptors, we also prioritized variation between descriptors). As well, baseline variables were retained if correlated with interaction descriptors, and subquotients of clinical variables were preferred with respect to more global indices to analyze the impact of specific domains. After this step, a set of N = 29 variables were included in the subsequent analysis. Table 3 reports the selected variables after correlation-based selection.

Table 3: Selected variables after the first step of correlation-based exclusion ($|r| >= 0.55$)

| **RETAINED VARIABLES AFTER FIRST SELECTION** |
|---|
| **Variable** |
| Age |
| Time (T0-T3) |
| rate_DC |
| QA.T0 |
| AS.T0 |
| CRR.T0 |
| hpw (hours per week) |
| therapist change (T0-T3) |
| therapist change (T0-T1-T2-T3) |
| therapist change (T0-T1) |
| P_CX_DELTA10 |
| R_SYNC_DELTA10 |
| LATENCY_SA_DELTA10 |
| LATENCY_TPCA_DELTA10 |
| R_TPCA_DELTA10 |
| DURATION_SA_DELTA10 |
| LATENCY_SYNC_DELTA10 |
| ENG_SA_DELTA10 |
| R_SA_DELTA10 |
| ENG_SA.T0 |
| R_SA_TPCA.T1 |
| LATENCY_TPCA.T1 |
| R_TPCA.T1 |
| R_CX_DELTA20 |
| DURATION_SA_DELTA21 |
| R_CX.T2 |
| LATENCY_TPCA.T2 |
| LATENCY_SYNC.T2 |
| R_TPCA.T2 |
| DURATION_SA_DELTA32 |
| LATENCY_SA_DELTA30 |
| ENG_SA_DELTA30 |
| R_CX.T3 |

*Random Forest predictor importance*

We employed a Machine Learning method, i.e., Random Forest, to assess variables importance at this stage through a single model fitting with optimized parameters. We considered both the %IncMSE and the permutation importance as an importance metric to perform feature selection, intersecting the first N = 10 most important variables. The final feature set consisted of N = 8 predictors. In this phase, RF predictor importance excluded the effect of therapist and therapist change. Table 4 summarizes importance-based RF feature selection. Figure 2 graphically shows the two metrics of variable importance.

Table 4: Random Forest variable selection based on %IncMSE and Permutation importance

| **Variable** | **%IncMSE** | **Permutation imp** |
|---|---|---|
| rate_DC | 0.049 | 11.380 |
| Age | 0.018 | 6.604 |
| LATENCY_TPCA.T1 | 0.014 | 6.745 |
| QA | 0.006 | 3.530 |
| LATENCY_SA_DELTA | 0.006 | 4.183 |
| LATENCY_SYNC_DELTA | 0.005 | 3.146 |
| P_CX_DELTA | 0.003 | 1.761 |
| LATENCY_TPCA_DELTA | 0.002 | 2.890 |

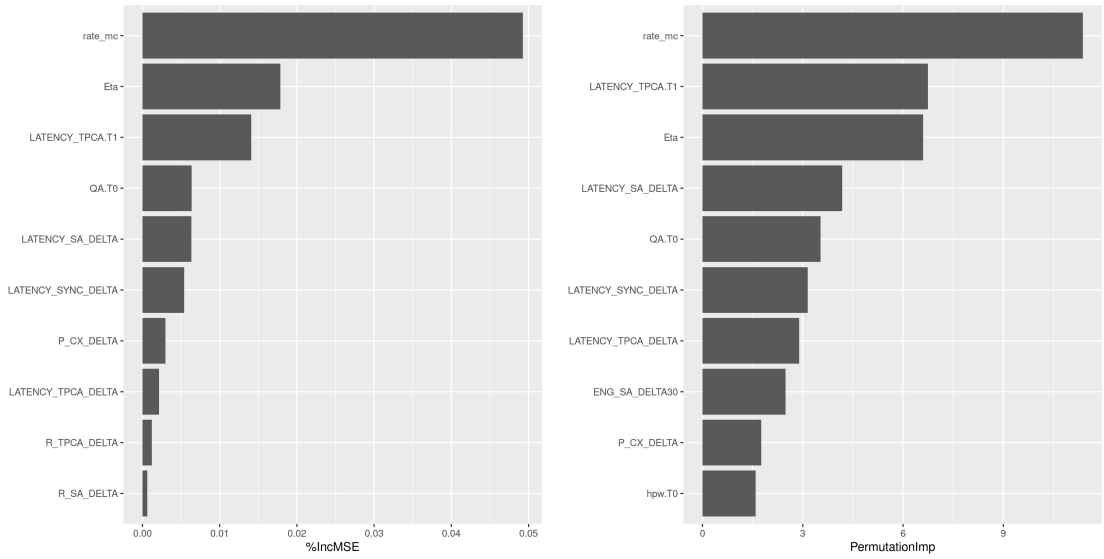


Figure 2: **Random Forest variable importance** - N = 10 most important variables with associated percentage of increase in Mean Standard Error (MSE) and permutation importance metric

*Final variable selection with Benjamini-Hochberg correction for multiple comparisons*

The first two step of variable selection reduced the pull of predictors from N = 125 initial variables to N = 29 after the correlation-based selection, and to N = 8 after the RF importance selection. At this stage, we performed a final variable selection step based on Pearson's

correlations between the candidate predictors and the outcome variable, i.e., the Learning Rate, applying Benjamini-Hochberg correction to control for multiple comparisons (FDR = 0.10 (Shang et al., 2022; Hu et al., 2010)). After correction, N = 4 candidate predictors were selected as significantly correlated with the LR and were considered to build the three models. We added one more interaction predictor to the analysis, LATENCY_TPCA.T1, i.e., the mean latency between a therapist proposal and child acceptance. We included this additional predictor, despite not being significantly correlated with the outcome variable, for two reasons: (i) it has been selected by RF importance as one of the most important predictors; and (ii) it is actually a clinically relevant parameter to be considered. Correlations between the predictors have also been explored. Table 6 reports the selected model formulas.

*Longitudinal changes in clinical outcome variables*

Table 5: Sample descriptive statistics and changes over time in outcome measures

| **Variable** | **T0 mean (ds) [range]** | **T3 mean (ds) [range]** | **test** | **p-value** | **d** | **BF** |
|---|---|---|---|---|---|---|
| Chronological age (C) (months) | 38.50 (10.18) [23-57] | 57.73 (12.17) [33-75] | | | | |
| GMDS-ER (D) Developmental age (months) | 26.19 (7.25) [14-45] | 39.60 (11.07) [18-63] | t(24)=-9.66 | 0.001 | 1.93 | 100 |
| D / C age ratio (RATE_DC) | 0.71 (0.19) [0.38-1.30] | 0.77 (0.26) [0.38-1.58] | t(24)=-2.73 | 0.012 | 0.56 | 4.23 |
| GMDS-ER General Quotient (Z) | 71.96 (15.22) [48-103] | 78.28 (22.29) [39-115] | t(24)=-2.16 | 0.041 | 0.43 | 1.5 |
| GMDS-ER Locomotion (Z) | 78.32 (18.37) [48-111] | 79.40 (18.87) [45-112] | t(24)=-0.42 | 0.677 | 0.09 | 0.23 |
| GMDS-ER Personal-Social (Z) | 68.88 (20.85) [33-120] | 74.36 (21.53) [28-114] | t(24)=-1.61 | 0.121 | 0.32 | 0.65 |
| GMDS-ER Language (Z) | 55.20 (25.56) [25-120] | 73.24 (35.11) [21-135] | V=60.50 | 0.011 | 0.66 | 13.89 |
| GMDS-ER Eye-hand coordination (Z) | 73.72 (19.13) [35-110] | 77.20 (22.46) [42-124] | t(24)=-1.22 | 0.234 | 0.24 | 0.41 |
| GMDS-ER Performance (Z) | 88.08 (23.20) [33-133] | 88.76 (24.14) [42-131] | t(24)=-0.19 | 0.851 | 0.04 | 0.21 |
| ADOS-2 Total Score | 16.16 (4.12) [9-24] | 14.04 (3.84) [8-21] | t(24)=2.94 | 0.007 | 0.61 | 7.73 |
| ADOS-2 Social Affect (AS) | 12.32 (3.29) [6-19] | 10.24 (3.23) [4-16] | t(24)=3.13 | 0.005 | 0.63 | 9.23 |
| ADOS-2 Restricted Repetitive Behaviors (RBB) | 3.84 (1.57) [1-7] | 3.76 (1.74) [1-7] | t(24)=0.20 | 0.84 | 0.04 | 0.22 |
| Learning Rate (LR) | 0.94 (0.49) [0.22 - 2.20] | | | | | |

Outcome variables description with means, standard deviations, and ranges, pre- and post- intervention. The appropriate paired statistical test is reported along with p-value, Cohen's d (d) and Bayes Factor (BF). The Learning Rate of each child was computed with the formula LR = (Dev_age(T3) - Dev_age(T0)) / (T3-T0)

*Model formulas*

Table 6: Selected model formulas

| SELECTED MODEL FORMULAS | |
|---|---|
| **Model** | **Formula** |
| **Baseline model** | |
| M1 | LR ~ rate_DC |
| **Interaction models** | |
| M2 | LR ~ P_CX_DELTA10 + LATENCY_SA_DELTA10 |
| M3 | LR ~ LATENCY_SA_DELTA10 + LATENCY_TPCA.T1 |
| **Combined models** | |
| M4 | LR ~ rate_DC + P_CX_DELTA10 |
| M5 | LR ~ rate_DC + P_CX_DELTA10 + LATENCY_SA_DELTA10 |
| M6 | LR ~ rate_DC + LATENCY_SA_DELTA10 + LATENCY_TPCA.T1 |

LR: Learning Rate
rate_DC: Developmental / Chronological age ratio
P_CX_DELTA10: Variation in proportional frequency of child's withdrawal between T1 and T0
LATENCY_SA_DELTA10: Variation in mean latency to start the interchange between T1 and T0
LATENCY_TPCA.T1: Mean latency for acceptance of therapist's proposals by the child

*Bootstrapped model coefficients*

Table 7: Bootstrapped model coefficients and Adjusted-$R^2$ with N = 1000 repetitions. The original estimates are reported along with the Bootstrap mean, Standard Error (SE), and the 95% Confidence Interval (CI).

| | **Original** | **Bootstrap (mean)** | **Bootstrap (SE)** | **25% C.I.** | **97.5% C.I.** |
|---|---|---|---|---|---|
| | | | **M1** | | |
| Intercept | 0.16 | 0.16 | 0.07 | 0.03 | 0.31 |
| rate_DC | 0.62 | 0.62 | 0.15 | 0.29 | 0.84 |
| Adj-$R^2$ | 0.48 | 0.49 | 0.17 | 0.12 | 0.79 |
| | | | **M2** | | |
| Intercept | 0.43 | 0.41 | 0.11 | 0.22 | 0.63 |
| LATENCY_SA_DELTA10 | 0.41 | 0.45 | 0.20 | -0.03 | 0.78 |
| P_CX_DELTA10 | -0.37 | -0.36 | 0.17 | -0.77 | -0.08 |
| Adj-$R^2$ | 0.32 | 0.37 | 0.13 | 0.03 | 0.54 |
| | | | **M3** | | |
| Intercept | 0.37 | 0.36 | 0.09 | 0.23 | 0.57 |
| LATENCY_TPCA.T1 | -0.46 | -0.46 | 0.18 | -0.73 | -0.12 |
| LATENCY_SA_DELTA10 | 0.64 | 0.64 | 0.17 | 0.17 | 0.64 |
| Adj-$R^2$ | 0.44 | 0.47 | 0.17 | 0.007 | 0.70 |

| | | | | | |
|---|---|---|---|---|---|
| **M4** | | | | | |
| Intercept | 0.31 | 0.31 | 0.08 | 0.16 | 0.46 |
| rate_DC | 0.57 | 0.59 | 0.11 | 0.30 | 0.74 |
| P_CX_DELTA10 | -0.31 | -0.32 | 0.15 | -0.65 | -0.07 |
| Adj-$R^2$ | 0.58 | 0.60 | 0.13 | 0.32 | 0.82 |
| **M5** | | | | | |
| Intercept | 0.22 | 0.22 | 0.08 | 0.06 | 0.37 |
| rate_DC | 0.51 | 0.51 | 0.10 | 0.30 | 0.68 |
| LATENCY_SA_DELTA10 | 0.29 | 0.25 | 0.15 | -0.07 | 0.57 |
| P_CX_DELTA10 | -0.30 | -0.30 | 0.15 | -0.58 | 0.02 |
| Adj-$R^2$ | 0.63 | 0.66 | 0.12 | 0.36 | 0.84 |
| **M6** | | | | | |
| Intercept | 0.19 | 0.18 | 0.08 | 0.02 | 0.34 |
| rate_DC | 0.44 | 0.45 | 0.13 | 0.24 | 0.69 |
| LATENCY_TPCA.T1 | -0.32 | -0.31 | 0.16 | -0.52 | 0.05 |
| LATENCY_SA_DELTA10 | 0.43 | 0.43 | 0.16 | 0.05 | 0.63 |
| Adj-$R^2$ | 0.64 | 0.67 | 0.16 | 0.19 | 0.85 |

*Model comparison*

The ANOVA between the nested models M1 and M4 showed a significant (F = 6.77; p = 0.016) improvement in model fit in favor of M4 and coherently with the BF analysis (BF(M4/M1) = 3.39). Coherently, the ANOVA between the nested models M1 and M5 showed a significant (F = 5.73; p = 0.010) improvement in model fit in favor of M5, also with BF(M5/M1) = 4.26. The result was even clearer with respect to M6 (F = 6.30; p = 0.007; BF(M6/M1) = 5.87).

The BF analysis comparing M2 against M1 indicated evidence in favor of M1 (BF(M2/M1) = 0.04), as well as between M3 against M1 (BF(M3/M1)=0.19)

The ANOVA between the two nested models M4 and M5 failed to detect a statistically significant difference (F = 3.83; p = 0.064; BF(M5/M4) = 1.26). However, both the AICc and the Adj-$R^2$ showed improvements in favor of the more complex model. The BF analysis comparing M6 against M5 indicated evidence in favor of M6 (BF(M6/M5) = 1.38). Table 8 summarizes the Cross-Validation and Bootstrap results for the selected models.

Considering the other evaluation metrics, Table 8 shows coherent results in favor of M6. In fact, it performed better than the other models in terms of AICc, BIC, BF, as well as during cross-validation, with the lowest MAE, RMSE, and the highest CV-$R^2$. Further M6 showed the most prominent improvement both in $R^2$ and in predictive performance with respect to any other model considered. Bootstrap performance for both model coefficients and Adjusted-$R^2$ are reported in Table 8, whereas Bootstrapped model evaluation metrics are reported in Supplementary Materials, with results suggesting an increase of stability from M1 to M6 models.

Table 8: Evaluation metrics of selected models: F-statistic (F), p-value, Adjusted-$R^2$, Corrected Akaike Information Criterion (AICc), Bayesian Information Criterion (BIC), Bayes Factor (BF), Mean Absolute Error (MAE), Root-Mean-Square Error (RMSE), Cross-Validated Adjusted-$R^2$ (CV-$R^2$), and Bootstrapped metrics.

| **Model** | **Statistics and estimators** | | | | | | **Cross-validation** | | | **Bootstrap** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **F** | **p-value** | **Adj-$R^2$** | **AICc** | **BIC** | **BF** | **MAE** | **RMSE** | **CV-$R^2$** | **MAE** | **RMSE** | **Adj-$R^2$** |
| M1 | 23.16 | 0.001 | 0.48 | 17.99 | 20.51 | 5.57 | 0.27 | 0.30 | 0.41 | 0.28 (0.06) | 0.34 (0.07) | 0.49 (0.22) |
| M2 | 6.718 | 0.005 | 0.32 | 26.35 | 29.22 | 2.25 | 0.33 | 0.37 | 0.20 | 0.39 (0.07) | 0.41 (0.09) | 0.34 (0.19) |
| M3 | 10.27 | 0.001 | 0.44 | 21.78 | 24.66 | 3.9 | 0.28 | 0.34 | 0.36 | 0.31 (0.08) | 0.37 (0.09) | 0.42 (0.08) |
| M4 | 17.87 | 0.001 | 0.58 | 14.15 | 17.02 | 6.79 | 0.25 | 0.28 | 0.51 | 0.26 (0.06) | 0.32 (0.07) | 0.55 (0.20) |
| M5 | 14.72 | 0.001 | 0.63 | 13.12 | 16.05 | 7.02 | 0.25 | 0.27 | 0.52 | 0.27 (0.06) | 0.32 (0.07) | 0.57 (0.20) |
| M6 | 15.49 | 0.001 | 0.64 | 12.24 | 15.18 | 7.34 | 0.22 | 0.24 | 0.60 | 0.24 (0.08) | 0.30 (0.09) | 0.55 (0.24) |

*Discussion*

*Changes in outcome measures*

In terms of outcome variables, in our dataset, the developmental profiles showed significant improvements over time. More specifically, such change was driven by the linguistic domain. On the contrary, the other developmental dimensions appeared to be more stable. This trend seems to be consistent with the literature on ASD treatment response highlighting the specificity by which the language domain may be affected by intervention (Panganiban & Kasari, 2022). Concerning symptom severity, significant reductions specific to the area of social affect were found. In comparison, restricted repetitive behaviors and interests appeared to remain more stable (Wetherby et al., 2018). High variability in outcome measures also emerged, confirming that treatment response can be highly variable on individual base (Sandbank et al., 2020). Therefore, we applied a careful model selection plan to identify optimal models of treatment response and relevant predictors.

## References


Apicella, F., Chericoni, N., Costanzo, V., Baldini, S., Billeci, L., Cohen, D., & Muratori, F. (2013). Reciprocity in interaction: A window on the first year of life in autism. *Autism Research and Treatment*, *2013*, 1–12. doi:10.1155/2013/705895.

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders*. American Psychiatric Association. doi:10.1176/appi.books.9780890425596.

Bertamini, G., Bentenuto, A., Perzolli, S., Paolizzi, E., Furlanello, C., & Venuti, P. (2021). Quantifying the child–therapist interaction in ASD intervention: An observational coding system. *Brain Sciences*, *11*, 366. doi:10.3390/brainsci11030366.

Biringen, Z. (2008). The Emotional Availability (EA) Scales manual: Part 1. Infancy/Early Childhood Version.

Eikeseth, S., Klintwall, L., Jahr, E., & Karlsson, P. (2012). Outcome for children with autism receiving early and intensive behavioral intervention in mainstream preschool and kindergarten settings. *Research in Autism Spectrum Disorders*, *6*, 829–835. doi:10.1016/j.rasd.2011.09.002.

Feldman, R. (2007). Parent–infant synchrony. *Current Directions in Psychological Science*, *16*, 340–345. URL: https://doi.org/10.1111/j.1467-8721.2007.00532.x. doi:10.1111/j.1467-8721.2007.00532.x.

Feldman, R. (2017). The Neurobiology of Human Attachments. *Trends Cogn Sci*, *21*, 80–99. URL: https://doi.org/10.1016/j.tics.2016.11.007. doi:10.1016/j.tics.2016.11.007.

Flückiger, C., Del Re, A., Wampold, B., & Horvath, A. (2018). The alliance in adult psychotherapy: A meta-analytic synthesis. *Psychotherapy*, *55*, 316. doi:10.1037/pst0000172.

Frost, K. M., Brian, J., Gengoux, G. W., Hardan, A., Rieth, S. R., Stahmer, A., & Ingersoll, B. (2020). Identifying and measuring the common elements of naturalistic developmental behavioral interventions for autism spectrum disorder: Development of the ndbi-fi. *Autism*, *24*, 2285–2297. URL: https://doi.org/10.1177/1362361320944011. doi:10.1177/1362361320944011.

Fuller, E., & Kaiser, A. (2020). The effects of early intervention on social communication outcomes for children with autism spectrum disorder: A meta-analysis. *Journal of autism and developmental disorders*, *50*, 1683–1700. doi:10.1007/s10803-019-03927-z.

Goldstein, P., Losin, E. A. R., Anderson, S. R., Schelkun, V. R., & Wager, T. D. (2020). Clinician-patient movement synchrony mediates social group effects on interpersonal trust and perceived pain. *The Journal of Pain*, *21*, 1160–1174. doi:10.1016/j.jpain.2020.03.001.

Green, J. (2006). Annotation: the therapeutic alliance–a significant but neglected variable in child mental health treatment studies. *Journal of child psychology and psychiatry, and allied disciplines*, *47*, 425–435. doi:10.1111/j.1469-7610.2005.01516.x.

Hu, J. X., Zhao, H., & Zhou, H. H. (2010). False discovery rate control with groups. *Journal of the American Statistical Association*, *105*, 1215–1227. doi:10.1198/jasa.2010.tm09329.

Kang, E., Gioia, A., Pugliese, C., Islam, N., Martinez-Pedraza, F., Girard, R., & Lerner, M. (2021). Alliance-outcome associations in a community-based social skills intervention for youth with autism spectrum disorder. *Behavior Therapy*, *52*, 324–337. doi:10.1016/j.beth.2020.04.006.

Klintwall, L., Eldevik, S., & Eikeseth, S. (2015). Narrowing the gap: effects of intervention on developmental trajectories in autism. *Autism : the international journal of research and practice*, *19*, 53–63. doi:10.1177/1362361313510067.

Koole, S., & Tschacher, W. (2016). Synchrony in Psychotherapy: A Review and an Integrative Framework for the Therapeutic Alliance. *Frontiers in psychology*, *7*, 862. doi:10.3389/fpsyg.2016.00862.

Leclère, C., Viaux, S., Avril, M., Achard, C., Chetouani, M., Missonnier, S., & Cohen, D. (2014). Why synchrony matters during mother-child interactions: A systematic review. *PLoS ONE*, *9*, e113571. URL: https://doi.org/10.1371/journal.pone.0113571. doi:10.1371/journal.pone.0113571.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation*. (schedule–second edition (ados-2 ed.). Los Angeles: Western Psychological Services.2012.

Luiz, D., Barnard, A., Knoesen, N., Kotras, N., Horrocks, S., McAlinden, P., & O'Connell, R. (2006).

*Griffiths Mental Development Scales—Extended Revised: Two to Eight*. Years: Administration Manual. Hogrefe.

Maloney, E. S., & Larrivee, L. S. (2007). Limitations of age-equivalent scores in reporting the results of norm-referenced tests. *Contemporary Issues in Communication Science and Disorders*, *34*, 86–93. doi:10.1044/cicsd_34_f_86.

Mazefsky, C., Herrington, J., Siegel, M., Scarpa, A., Maddox, B., Scahill, L., & White, S. (2013). The role of emotion regulation in autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*, 679–688. doi:10.1016/j.jaac.2013.05.006.

Minjarez, M., Bruinsma, Y., & Stahmer, A. (2020). Considering NDBI Models.+.

Panganiban, J., & Kasari, C. (2022). Super responders: Predicting language gains from JASPER among limited language children with autism spectrum disorder. *Autism Research*, *1*, 11. doi:10.1002/aur.2727.

Ministero della Salute, I. S. d. S. (2011). Il trattamento dei disturbi dello spettro autistico nei bambini e negli adolescenti. *Linea Guida 21, Sistema Nazionale per le Linee Guida*, .

Sandbank, M., Bottema-Beutel, K., Crowley, S., Cassidy, M., Dunham, K., Feldman, J., Crank, J., Albarran, S., Raj, S., Mahbub, P., & Woynaroski, T. (2020). Project AIM: Autism intervention meta-analysis for studies of young children. *Psychological bulletin*, *146*, 1–29. doi:10.1037/bul0000215.

Schreibman, L., Dawson, G., Stahmer, A., Landa, R., Rogers, S., McGee, G., & Halladay, A. (2015). Naturalistic developmental behavioral interventions: Empirically validated treatments for autism spectrum disorder. *Journal of autism and developmental disorders*, *45*, 2411–2428. doi:10.1007/s10803-015-2407-8.

Shang, P., Ho, A. M.-C., Tufvesson-Alm, M., Lindberg, D. R., Grant, C. W., Orhan, F., Eren, F., Bhat, M., Engberg, G., Schwieler, L., Fatouros-Bergman, H., Imbeault, S., Iverson, R. M., Dasari, S., Piehl, F., Cervenka, S., Sellgren, C. M., Erhardt, S., & Choi, D.-S. (2022). Identification of cerebrospinal fluid and serum metabolomic biomarkers in first episode psychosis patients. *Translational Psychiatry*, *12*. doi:10.1038/s41398-022-02000-1.

Siegel, D. J. (2001). Toward an interpersonal neurobiology of the developing mind: Attachment relationships,"mindsight," and neural integration. *Infant Mental Health Journal: official publication of the world association for infant mental health*, *22*, 67–94. doi:10.1002/1097-0355(200101/04)22:1.

Vivanti, G., Bottema-Beutel, K., & Turner-Brown, L. (Eds.) (2020). *Clinical guide to early interventions for children with autism*. Berlin/Heidelberg, Germany: Springer. doi:10.1007/978-3-030-41160-2.

Vivanti, G., & Zhong, H. (2020). Naturalistic Developmental Behavioral Interventions for Children with Autism. In G. Vivanti, K. Bottema-Beutel, & L. Turner-Brown (Eds.), *Clinical Guide to Early Interventions for Children with Autism* (pp. 93–130). Cham: Springer International Publishing. URL: https://doi.org/10.1007/978-3-030-41160-2_6. doi:10.1007/978-3-030-41160-2_6.

Waddington, H., Reynolds, J. E., Macaskill, E., Curtis, S., Taylor, L. J., & Whitehouse, A. J. (2021). The effects of jasper intervention for children with autism spectrum disorder: A systematic review. *Autism*, *25*, 2370–2385. URL: https://doi.org/10.1177/13623613211019162. doi:10.1177/13623613211019162. arXiv:https://doi.org/10.1177/13623613211019162. PMID: 34348479.

Wetherby, A., Woods, J., Guthrie, W., Delehanty, A., Brown, J., Morgan, L., Holland, R., Schatschneider, C., & Lord, C. (2018). Changing Developmental Trajectories of Toddlers With Autism Spectrum Disorder: Strategies for Bridging Research to Community Practice. *Journal of speech, language, and hearing research : JSLHR*, *61*, 2615-2628. doi:10.1044/2018_JSLHR-L-RSAUT-18-0028.

Wong, V., Fuller, E. A., & Rogers, S. J. (2019). Time to make a change. early start denver model for young children with asd. *Enfance*, *1*, 83. URL: https://doi.org/10.3917/enf2.191.0083. doi:10.3917/enf2.191.0083.

Zwaigenbaum, L., Bauman, M., Choueiri, R., Kasari, C., Carter, A., Granpeesheh, D., & Pierce, K. (2015). Early intervention for children with autism spectrum disorder under 3 years of age: recommendations for practice and research. *Pediatrics*, *136*, 60- 81. doi:10.1542/peds.2014-3667E.