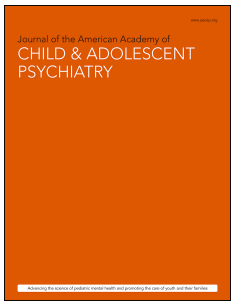
Journal Pre-proof



A Multisite Randomized Controlled Trial Comparing the Effects of Intervention Intensity and Intervention Style on Outcomes for Young Children With Autism

Sally J. Rogers, PhD, Paul Yoder, PhD, Annette Estes, PhD, Zachary Warren, PhD, John McEachin, PhD, Jeff Munson, PhD, Marie Rocha, PhD, Jessica Greenson, PhD, Lisa Wallace, MS, Elizabeth Gardner, PhD, Geraldine Dawson, PhD, Catherine A. Sugar, PhD, Gerhard Hellemann, PhD, Fiona Whelan, MS

PII: S0890-8567(20)31350-2

DOI: https://doi.org/10.1016/j.jaac.2020.06.013

Reference: JAAC 3147

To appear in: *Journal of the American Academy of Child & Adolescent Psychiatry*

Received Date: 18 December 2019

Revised Date: 30 April 2020

Accepted Date: 16 June 2020

Please cite this article as: Rogers SJ, Yoder P, Estes A, Warren Z, McEachin J, Munson J, Rocha M, Greenson J, Wallace L, Gardner E, Dawson G, Sugar CA, Hellemann G, Whelan F, A Multisite Randomized Controlled Trial Comparing the Effects of Intervention Intensity and Intervention Style on Outcomes for Young Children With Autism, *Journal of the American Academy of Child & Adolescent Psychiatry* (2020), doi: https://doi.org/10.1016/j.jaac.2020.06.013.

A Multisite Randomized Controlled Trial Comparing the Effects of Intervention Intensity and Intervention Style on Outcomes for Young Children With Autism

RH = Early Intervention Style and Intensity

Editorial
Supplemental Material

Sally J. Rogers, PhD, Paul Yoder, PhD, Annette Estes, PhD, Zachary Warren, PhD, John McEachin, PhD, Jeff Munson, PhD, Marie Rocha, PhD, Jessica Greenson, PhD, Lisa Wallace, MS, Elizabeth Gardner, PhD, Geraldine Dawson, PhD, Catherine A. Sugar, PhD, Gerhard Hellemann, PhD, Fiona Whelan, MS

Accepted August 17, 2020

Drs. Rogers and Rocha are with the University of California, Davis MIND Institute, Davis. Drs. Yoder, Warren, Wallace, and Gardner are with Vanderbilt University, Nashville, Tennessee. Drs. Estes, Munson, and Greenson are with the University of Washington, Seattle. Dr. McEachin is with Autism Partnership Foundation, Seal Beach, California. Dr. Dawson is with Duke University, Durham, North Carolina. Drs. Sugar, Hellemann, and Ms. Whelan are with the University of California, Los Angeles.

This study was supported by the National Institute of Mental Health (NIMH)/the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) award number 1R01MH100030 (Sally J. Rogers, PI) as part of the Autism Centers of Excellence (ACE) Treatment Network and by the MIND Institute Intellectual and Developmental Disabilities Research Center U54HD079125 (Leonard Abbeduto, PI). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agency.

Dr. Hellemann served as the statistical expert for this research.

Author Contributions
Conceptualization: Rogers, Yoder
Data curation: Munson, Greenson, Whelan
Formal analysis: Yoder, Munson, Sugar, Hellemann
Funding acquisition: Rogers
Investigation: Rogers, Estes, McEachin, Wallace, Gardner, Sugar
Methodology: Rogers, Yoder, Estes, McEachin, Greenson, Dawson, Hellemann
Project administration: Rogers, Estes, Warren, Rocha, Greenson, Wallace, Gardner, Sugar
Resources: Rogers, Estes, Warren
Software: Munson, Hellemann, Whelan
Supervision: Rogers, Estes, Warren, McEachin, Rocha, Greenson, Wallace, Gardner, Sugar, Hellemann
Writing – original draft: Rogers, Yoder, Hellemann
Writing – review and editing: Rogers, Yoder, Estes, Warren, McEachin, Munson, Rocha, Greenson, Wallace, Gardner, Dawson, Sugar, Hellemann

ORCID
Sally J. Rogers, PhD: https://orcid.org/0000-0003-0755-2900
Paul Yoder, PhD:
Annette Estes, PhD: https://orcid.org/0000-0003-2687-4155
Zachary Warren, PhD: https://orcid.org/0000-0001-9677-9386
John McEachin, PhD:
Jeff Munson, PhD:
Marie Rocha, PhD: https://orcid.org/0000-0001-7018-6427
Jessica Greenson, PhD:
Lisa Wallace, MS:
Elizabeth Gardner, PhD: https://orcid.org/0000-0003-2785-1385
Geraldine Dawson, PhD: https://orcid.org/0000-0003-1410-2764
Catherine A. Sugar, PhD: https://orcid.org/0000-0003-2804-7612
Gerhard Hellemann, PhD:
Fiona Whelan, MS:

The authors wish to acknowledge members of the MIND Institute staff and research participants including the work of the therapists and families involved in this study and for the help of Shannon Barber, MPH, and Diane Larzelere, BA, of University of California, Davis, and Catherine Bush, MA, of Vanderbilt University, with the manuscript.

Disclosure: Dr. Rogers has received a University of British Columbia BC subcontract, grant funding from NICHD, the US Health Resources Services Administration (HRSA), the US Department of Education, and NIMH, royalties for published materials from Guilford Press, APPI Press, Springer Press, Wiley Press, and UC Extension Service, and honoraria for academic talks. Dr. Yoder has received grant funding from the National Institute on Deafness and Other Communication Disorders (NIDCD) and honoraria for academic talks unrelated to article content. Dr. Estes has received grant funding/the National Institutes of Health (NIH) subcontracts from NIMH, NICHD, and the NIH Brain Initiative Fellowship. She has served as a consultant to Motivity and has received honoraria for academic talks. Dr. Warren has received grant funding from the Centers for Disease Control and Prevention, the Centers for Medicare and Medicaid Services (CMS), the US Department of Defense, HRSA, the Institute for Education Sciences (IES), NICHD, NIMH, Autism Speaks-the Autism Treatment Network, the Simons Foundation, and the Tennessee Department of Education. He has served as a consultant to Roche and Adaptive Technology Consulting and has received honoraria for academic talks unrelated to article content. Dr. McEachin has received book royalties from DRL Books and honoraria for invited addresses. Dr. Munson has received grant funding from NIMH, the National Institute of Allergy and Infectious Diseases (NIAID), and Children's Healthcare of Atlanta. Dr. Rocha has received grant support from IES, NICHD, NIMH, and HRSA and has served as a consultant for clinical services. Dr. Gardner has received grant funding from NICHD. Dr. Dawson has served as a member of the Scientific Advisory Boards of Janssen Research and Development, Akili, Inc., LabCorp, Inc., Roche Pharmaceutical Company, and Tris Pharma. She has served as a consultant to Apple, Gerson Lehrman Group, Guidepoint, Inc., Axial Ventures, Teva Pharmaceuticals, and is CEO of DASIO, LLC. She has received book royalties from Guilford Press, Oxford University Press, Springer Nature Press. She has held the following patent applications: 1802952, 1802942, 15141391, and 16493754. She has developed technology that

has been licensed and she and Duke University have benefited financially. She has received research grants from NICHD, NIMH/the National Institute of Neurological Disorders and Stroke (NINDS), NIMH, the Simons Foundation Award, the Stylli Translation Neuroscience Award, and the Marcus Foundation Award. Dr. Sugar has received research grant support from NIH through multiple divisions, including NIMH, the National Institute of General Medical Sciences (NIGMS), NICHD, NIAID, and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), HRSA, the Food and Drug Administration (FDA), and the US Department of Veterans Affairs. She has served on technical expert panels for CMS and Data Safety and Monitoring Boards for both academic institutions and Kaiser Permanente. Ms. Wallace has received grant funding from the US Department of Defense and the Tennessee Department of Education. Ms. Whelan has received grant support from NIMH. Drs. Greenson and Hellemann have reported no biomedical financial interests or potential conflicts of interest.

Correspondence to Sally J. Rogers, PhD, University of California, Davis, MIND Institute, 2825 50th Street, Sacramento, CA 95817; e-mail: sjrogers@ucdavis.edu

Abstract

Objective: This randomized, multisite, intent-to-treat study tested the effects of two levels of treatment intensity (number of hours) and two treatment styles on progress of young children with autism spectrum disorder (ASD). We predicted that initial severity of developmental delay or autism symptoms would moderate the effects of intensity and style on progress in four domains: autism symptom severity, expressive communication, receptive language, and nonverbal ability.

Method: Eighty-seven children with ASD, mean age 23.4 months, were assigned to one of two intervention styles (naturalistic developmental-behavioral or discrete trial teaching), each delivered for either 15 or 25 hours per week of 1:1 intervention for 12 months by trained research staff. All caregivers received coaching twice monthly. Children were assessed at four timepoints. Examiners and coders were naive to treatment assignment.

Results: Neither style nor intensity had main effects on the four outcome variables. In terms of moderating effects of initial severity of developmental delay and of autism symptom severity, neither moderated the effects of treatment style on progress in any of the four domains. In terms of treatment intensity, initial severity moderated effect of treatment intensity on only one domain - change in autism symptom severity, and in a secondary analysis, this effect was found in only one site.

Conclusion: Neither treatment style or intensity had overall effects on child outcomes in the four domains examined. Initial severity did not predict better response to one intervention style than another. We found very limited evidence that initial severity predicted better response to 25 versus 15 hours per week of intervention in the domains studied.

Clinical trial registration information: Intervention Effects of Intensity and Delivery Style for Toddlers With Autism: https://clinicaltrials.gov/; NCT02272192

Key words: autism spectrum disorder, early intervention, Early Start Denver Model, early intensive behavioral intervention, treatment intensity

INTRODUCTION

For young children with autism spectrum disorders (ASD), high quality single-blind randomized controlled trials of well-defined, manualized interventions delivered 1:1 at intended fidelity have demonstrated significant effects on IQ, expressive language, receptive language, and autism severity.[1-5] While these different approaches have all demonstrated efficacy, their manuals and publications describe large differences among them in (1) intervention style (e.g., adult-directed didactic vs naturalistic developmental behavioral), and (2) intervention intensity (i.e., hours of intervention per week). While caregivers and practitioners need information about what style and intensity of early intervention is optimal for a given child, there have been no rigorously controlled comparative studies addressing these questions, which have major policy, practice, and funding implications for intervention delivery systems, professional training and practice, caregiver choice and learning, and children's lives. Multiple previous intervention studies reported that autism symptom severity, severity of developmental delays, and treatment intensity predicted child outcomes. However, none of these studies used an RCT design.[6-9] While it is common for some interventions to be delivered 1:1 at intensities as high as 35-40 hours a week in keeping with Lovaas's (1987) initial report, the costs and scarcity of such intensive treatment require supportive high quality evidence, providing one of the rationales for this study.[10]

Because rigorous assessment of relative efficacy of treatment style and intensity requires a study of adequate power and the control of other factors, we designed a randomized, multi-site, intent-to-treat design with naive examiners and coders to compare effects of two intervention styles and two intervention intensities on outcomes in multiple domains of toddlers with ASD.[11]

The styles compared were 1:1 discrete trial teaching (Early Intensive Behavioral Intervention, EIBI) and 1:1 naturalistic developmental-behavioral intervention (Early Start Denver Model, ESDM). The intensity levels were 15 hours or 25 hours per week delivered for 12 months.[12,13]

Given the similarity of outcomes from well-controlled RCTs testing very different intensities and styles we did not predict main effects of intensity or style.[1,2,4] However, based on the slower learning rates and need for more learning repetitions of children with lower developmental quotients (DQs) than those with higher DQs, we expected baseline DQ to moderate effects of treatment intensity and style on children's outcomes. And based on the decreased social responsivity of children with more severe versus milder autism symptoms, we predicted that baseline ASD severity would moderate child response to treatment style and intensity.

The study hypotheses were as follows:

Hypothesis 1: Initial degree of developmental delay and autism severity will moderate *effects of treatment style* on growth trajectories of expressive communication, receptive language, nonverbal ability, and autism symptom severity, with milder degree of delay and/or autism symptoms predicting greater progress in response to ESDM versus EIBI treatment.

Hypothesis 2: Initial degree of developmental delay and autism severity will moderate *effects of treatment intensity* on growth trajectories of expressive communication, receptive language, nonverbal ability, and autism symptom severity, with more severe initial developmental delay and/or autism symptoms predicting greater progress in response to 25 versus 15 hours of treatment.

METHOD

Procedure

Trial design: We conducted an intent-to-treat, single-blind RCT at three universities. The study period began in 2013 and ended in 2019, depicted in Figure 1.

Children were recruited, screened, qualified, consented, stratified by DQ and age, and randomized to one of four cells (15 or 25 hours of ESDM; 15 or 25 hours of EIBI). Treatment was delivered for 12 months in homes (most children) and/or childcare settings by research staff. Additionally, all families received two 1.5 hour sessions of caregiver coaching monthly in use of the assigned intervention. An independent data coordinating center (DCC) conducted the randomization of participants using a computer program allocation. All assessor-administered measures were conducted by staff researchers naïve to group assignment. Observational measures were coded by observers who were naïve to group assignment. Caregivers were not naïve to assignment and caregiver reports were used in some measures. Interventionists were trained across sites to acceptable levels of fidelity of implementation (FOI) before beginning to treat children, were supervised by expert professional therapists regularly, and were overseen by developers of the two treatments. The treatment manuals were followed rigorously; ongoing training and FOI monitoring occurred across sites throughout treatment. See Supplement 1, available online, for more information.

All children were scheduled for four clinic-based assessments conducted by qualified, trained staff members across a 24-month period: time of enrollment, 6 months following enrollment (mid intervention phase), 12 months following enrollment (end of intervention phase), and 24 months following enrollment (follow-up phase). Caregivers were provided verbal and written reports and referral to community-based services. Given the intent-to-treat design, all children's data were included in the original assignment group, including those who dropped out

before the study's end and all were assessed at all timepoints as possible. The project was approved by appropriate Institutional Research Boards. Design and data were routinely reviewed by an independent Data Safety and Monitoring Board. Main modifications to the design after commencement of the study involved reducing the minimum age to 12 months, adding inclusion/exclusion criteria described above, some additions of measures not being reported here, timing of measure administration, personnel changes, and addition of follow-up procedures.

Recruitment, enrollment, and randomization

Children were recruited from pediatricians, developmental disability settings, and university website postings using IRB-approved materials. DCC randomization was performed using a computer algorithm based on two pre-specified blocks: 12-20 months or 20+ months and DQ <60 or DQ >60. Blocks were monitored and a difference of 3 between ESDM and EIBI groups resulted in a compensatory assignment to maximize equal samples.

Participants

We screened 128 toddlers referred to the study using two different published, age-appropriate autism risk screeners and we conducted diagnostic assessments on all those who screened positive according to screener criteria. We enrolled 87 with Autism Spectrum Disorder (DSM-5) recruited from three sites: 28 from Site 1, 30 from Site 2, and 29 from Site 3. Both genders were recruited; the group was ethnically/racially diverse. See CONSORT Table, Figure 1, for the flow of participants through the study.

Inclusion criteria: 12-30 months of age at time of assessment; ambulatory and without impairments affecting hand use; meets criteria for Autism Spectrum Disorder on the APA Diagnostic and Statistical Manual, 5th revision criteria and on the Autism Diagnostic Observation Schedule for Toddlers clinical consensus of ASD diagnosis by 2 independent staff (including a

licensed psychologist) based on observation as well as record review; overall developmental quotient of ≥35 on Mullen Scales of Early Learning; normal hearing and vision screen; and caregiver agreement to comply with all project requirements, including regular videotaping at home with provided equipment.[14-16]

Exclusion criteria: English not a primary language spoken at home; absence at 2 or more appointments without prior notice during the intake assessment, more than 10 hours per week of 1:1 ABA based treatment, other health or genetic conditions (i.e. fragile X syndrome, seizures, prematurity).

Sample Size. Power analyses were calculated using longitudinal mixed-level models on data from measures used in Dawson et al.[1], estimated for 108 participants. In the Dawson study, longitudinal analyses of MSEL data showed a 17.6% group difference, between-subject variances of 23.7, and within-subject correlation of repeated measurements of 0.45.[1] We had 90% power to detect main effects of the treatment intensity and the style on average MSEL score of **2.0-2.25%** when between-subject variance was 30 and within-subject correlation of repeated measurements was 0.4. We had over 90% detection power when this difference increased to 2.5%, as hypothesized for those with more severe initial developmental and autism symptoms, and also when the within-subject correlation increased to 0.5.

Treatment:

Children were scheduled to receive either 15 or 25 hours a week of 1:1 treatment in their homes or care/preschool settings, delivered by trained interventionists (ITs), for 12 months. Sessions were typically 1.5 or 2.5 hour blocks, 10 blocks per week, generally one AM and one PM block, fitted around sleep and family schedules.

ITs were employed full-time by the study and were trained and overseen by full-time supervisors with graduate degrees who monitored the treatment at all sites via internet and live viewing and met regularly within and across sites to ensure that all procedures were carried out consistently. Treatment procedures followed the published manuals and implementation guidelines. Primary caregiver(s) received coaching in the assigned model by their team supervisor every two weeks throughout their enrollment. See Supplement 1, available online, for further information.

Intervention intensity: 15 vs. 25 hours per week. All delivered hours across the 48 treatment weeks (out of 52) were recorded daily and reported weekly. Cancelled treatment hours were made up as possible. We quantified the actual number of hours of treatment per week for each participant using schedulers' weekly attendance logs. Both groups received an average of 83% of the hours scheduled: means were 12.42 hrs/wk (*SD* = 1.45) and 20.82 hrs/wk (*SD* = 1.40) for low and high intensity groups, respectively, which differed significantly by design (Cohen's between-group $d = 5.9$).

Fidelity of treatment implementation. The Leaf and McEachin (1999) manual, **A Work in Progress,** defined the EIBI approach and curriculum.[12] Supervisors and ITs received quarterly ongoing onsite training and ongoing consultation from (xxxx, blinded). An FOI tool was developed to measure correct implementation of 9 components used a 5-point Likert scale applied to randomly selected 20 minute sections of recorded treatment sessions (Yoder, P., McEachin, J., Wallace, E., Leaf, R., 2014, unpublished. Discrete Trial Training Fidelity of Treatment Rating.) During instruction, children had blocks of teaching trials interspersed with short breaks that included therapist interactions and a 10-15 minute break half way through each session.

ESDM intervention was carried out as per manual instructions and measured by its published FOI tool.[13] Instruction followed the ESDM curriculum and was embedded in typical early childhood activities, carried out within the joint activity structure first defined by Ratner and Bruner (1978), using everyday play objects and activities and routines from everyday life.[18] The FOI tool measured quality of instruction on a 14 item, 5 point Likert based scale from videos of 30 minutes or more of activities.[13] ESDM credentialed therapists supervised the ITs and their work was overseen by certified ESDM Trainers.

Staff were trained to fidelity on all procedures before beginning to work with children and maintained 80% or better fidelity scores measured twice a month throughout the project, assessed via 20-minute random videotaped segments of therapy gathered quarterly. We quantified FOI by using the average of the key items from each FOI rating scale. With "5" as the score that indicated highest compliance, means were 4.15 (*SD* = 1.3) for EIBI (83% of possible total) and 4.3 (*SD* = .15) for ESDM (86%). There were no significant differences between them (Cohen's between-group *d* = -.16; 95% CI [-.59, .27]). If fidelity scores dropped below 80%, the supervisor retrained the IT until acceptable fidelity was regained. See Supplement 2, available online, for additional detail.

We created used a Treatment Modification instrument to measure the extent to which each child's intervention may have deviated from the core EIBI or ESDM principles as a way of individualizing the intervention approach for a particular child's learning needs in ways other than what was captured in the FOI instruments (Yoder, 2014, unpublished). For the EIBI group, the instrument assessed whether four naturalistic modifications were made. For the ESDM group, the instrument assessed whether four EIBI-type modifications were made. Each quarter, supervisors rated the percent modifications made in EIBI or ESDM using a Likert-like scale (i.e.,

1 = 1 – 25%, 2 = 26 – 50%, 3 = 51 - 75%, 4 = 76 – 100%). The degree to which EIBI treatment incorporated naturalistic characteristics was associated with Time 1 DQ, $r = .50$, $p < .05$, and Time 1 autism severity, $r = -.35$, $p < .05$; thus adding ESDM principles to EIBI was related to initial milder severity. The degree to which ESDM treatment incorporated EIBI characteristics was correlated with Time 1 developmental delay, $r = -.47$, $p < .05$, and Time 1 autism severity, $r = .34$, $p < .05$, indicating that the addition of discrete trial methods to ESDM was done for children with initial greater severity.

Measures

Screening measures:

Early Screening of Autistic Traits Questionnaire (ESAT), an autism risk parent-report questionnaire validated in several large studies, was used for children 12-15 months. The inclusion criterion required 3 or more negative responses.[19]

Infant Toddler Checklist (ITC), is a standardized parent questionnaire with screening cutoffs and standard scores for children 6-24 months based on a normative sample of over 2,188 children.[20] Criterion was a score at or below the 5th percentile.

Modified Checklist for Autism in Toddlers (M-CHAT)**,** a 23-item autism screening questionnaire validated in two large studies, was used for children 16-24 months.[22] Criterion was failure on 2 critical items or any three items.

Constructing the composite scores. The primary outcome and moderating variables were composites representing expressive language, receptive language, nonverbal performance, and autism severity constructed from variables of the component measures described below. See Table 1 for a list of the composites, the variables included in each composite, and relevant statistics. We computed composite scores from multiple measures of a construct to increase

construct validity and to minimize the number of significance tests required to test predictions, thereby reducing familywise error. To build the composite scores, we first examined whether the proposed component variables were correlated (for 2 component variables) or factor loaded (for 3 or more component variables) at or above .3. See Supplement 3, available online, for Chronbach's alpha data on relationships among these composites. Because a single time period's mean and SD is required so that composites can show change over time, we used Time 4 means and SD to compute the z scores.

Measures contributing to composite dependent variables or moderators

The ADOS calibrated severity score was calculated from the Autism Diagnostic Observation Scale (ADOS 2) assessment. The appropriate module of the standardized ADOS based on each child's language level was administered.[23] Assessors naive to assignment were trained to 85% reliability on the full range of scores and all met research criteria. Inter-observer reliability at individual sites was assessed on at least 15% of interviews, and any deviation from standards led to retraining.

The PDD Behavior Inventory is a rating scale filled out by caregivers and teachers designed to assess response to intervention in children with PDD.[24] Subscales measure both maladaptive and adaptive behaviors as well as a summary score that reflects overall severity. Factor analyses have confirmed the structure of the scale. Correlations with psychometrically sound instruments that assess autism severity, adaptive behavior, and maladaptive behavior reveal high correlations and document construct validity. Scores used in composite dependent variables and one of the moderators are listed in Table 1.

The Communication Sample Procedure is a 15-minute videotaped lab-based communication sample using a standard toy set that provides semi-structured free-play with

interspersed opportunities for the child to respond to an adult's topic change and to request clarification. The examiner's interaction style is guided by specific principles described in the procedure manual, available from the authors. Using a timed event sampling method, trained observers unaware of group coded number of different words and weighted frequency of intentional communication from video. Reliability of the coded variables was assessed on a random selection of 20% of the sessions, with coders kept naive to session dates. The intraclass correlation coefficient using a two-random analysis and absolute agreement method averaged .85 for the weighted frequency of intentional communication and averaged .83 for the number of different words. Variables used in composites included weighted frequency of intentional communication and number of different word roots.

Mullen Scales of Early Learning are standardized measures of early child development across four domains: expressive and receptive language, fine motor and visual perception.[16] These were carried out by experienced assessors naive to group assignment employed and trained at the research sites. Reinforcers for child attention and cooperation were used as needed to support children's motivation to perform. Because the standard scores on the MSEL have set floors and many of the participants had scores at the floor, we used the age equivalent scores from each subdomain to construct the composite dependent variables and the initial DQ for a possible moderator**.**

The MacArthur-Bates Communicative Developmental Inventories information from caregivers about expressive and receptive words observed in their children in the recent past.[25] The expressive vocabulary lists from the Infant and Toddler inventories were combined into one caregiver checklist and endorsed items were summed to form the expressive raw score variable.

The Vineland Adaptive Behavior Scales – second edition is a standardized caregiver questionnaire that gathers information across four domains: expressive and receptive language, daily living skills, and motor skills. Caregivers completed the caregiver form during assessment visits.[26]

Overall developmental quotient (DQ): This variable was constructed by averaging the age equivalents of the expressive and receptive language, fine motor and visual perception subscales from the MSEL and dividing by chronological age.

Measure of Intervention hours received outside of the project

Amount of therapeutic intervention of all types was examined at each timepoint and calculated as per procedures described in Rogers et al. (2019).[3] There were no significant differences in number of treatment hours delivered across the sites and across the four randomized experimental groups. Groups received very little outside treatment weekly in year 1, as shown in Table 2, averaging only 1-2 hours of additional treatment per week. In year 2, after project-delivered treatment had ended, there were no significant differences in treatment received across the four randomized experimental groups; there was a significant difference in number of treatment hours received in year 2 across the sites.

Table 2 goes about here.

Measure of Caregiver Satisfaction with Treatment

1. The **TADPOLE Project Intervention Evaluation for Parents** (Estes, A. 2013. TADPOLE Project Intervention Evaluation for Parents) was a 16 item Likert- based caregiver satisfaction questionnaire constructed to capture caregiver experiences across the two different treatment styles and intensities and used to examine caregiver

acceptability of each treatment and intensity. Total raw score was used to quantify caregiver satisfaction with the treatment that their child received.

Analysis Plan

The four primary outcomes (autism severity, receptive language, expressive communication and non-verbal development) were analyzed using a general linear mixed model (GLMM). The GLMM is a full information method that included information from all randomized (intent to treat) participants, including those that provide only partial data due to drop out or other reasons. We modelled the participants' development over time by using a linear spline multilevel model to account for the specific structure of the data, involving an observation range of 24 months that included two distinct phases, an active treatment phase from the beginning of the study to month 12, and a follow-up phase from month 12 to the end of the study at month 24. The linear spine model simultaneously modelled an overall linear trajectory from the beginning of the study to the end of the study and a deviation term that modelled the degree to which the transition from active treatment to the follow-up period led to changes in the trajectories. We used a random intercept to account for individual differences in the dependent variable at baseline. Site was included as a covariate when that predictor and interactions with that predictor were significant. Time was parameterized as months since study entry and was exact for each participant so that included potential effects due to assessment appointments that were not exactly on schedule were included.

The two independent variables in this model were treatment intensity (high vs low hours) and treatment style (ESDM vs. EIBI). A second set of analyses addressed whether effects of these treatment characteristics were moderated by baseline DQ or autism severity. Due to the complex model and the sample size, we carried out a hierarchical series of models that included

site as a predictor only if site or if interactions with site were significant in the primary analyses. A final set of analyses were conducted to help interpret higher order statistical interactions.

RESULTS

Analytic Methods and Baseline Data

Baseline data: Baseline data for all demographic measures are presented in Table 2. There were significant site differences on several variables: child age, race, DQ, and maternal education, as expected given geographic locations. However, there were no significant differences among the four randomized groups on any variable.

Change on outcome measures

As seen in Table 3, all groups of children made significant gains on all composite outcome variables regardless of assignment group. Effect sizes (i.e., Cohen's *d*) for the within-group change ranged from |.5| to |2.2|. All mean change scores are significantly higher than zero.

Effects of Treatment Style

Overall there was no evidence that treatment style had effects on the trajectories for any of the dependent variables over time: autism severity (overall trajectory: $F(1,225)=0.9$, $p=.76$), spline: $F(1,224)=0.2$, $p=.16$)); expressive language (overall trajectory: $F(1,224)=0.2$, $p=.78$, spline: $F(1,223)=0.5$, $p=.48$)); receptive language (overall trajectory: $F(1,224)=0.3$, $p=.59$, spline: $F(1,223)=0.0$, $p=.87$)); and nonverbal ability (overall trajectory: $F(1,222)=1.7$, $p=.18$, spline: $F(1,221)=0.4$, $p=.67$)). There was no significant interaction of style by time by site.

Moderating Effects of Initial Severity on Effect of Treatment Style

Baseline autism severity did not moderate the effects of treatment style on trajectories of any dependent variable: autism severity (overall: $F(1,228)=0.1$, $p=.77$, spline: $F(1,228)=0.0$,

p=.92); expressive language (overall: F(1,226)=0.3, p=.58, spline: F(1,226)=0.0, p=.98); receptive language (overall: F(1,223)=2.8, p=.09, spline: F(1,223)=1.0, p=.32); nonverbal ability(overall: F(1,224)=0.0, p=.83, spline: F(1,224)=0.8, p=.37).

Similarly, baseline DQ did not moderate the effect of treatment style on trajectories of the dependent variables: autism severity (overall: F(1,231)=0.5, p=.46, spline: F(1,229)=0.7, p=.74), expressive language (overall: F(1,229)=0.2, p=.66, spline: F(1,228)=0.1, p=.90), receptive language (overall: F(1,229)=0.7, p=.42, spline: F(1,227)=0.9, p=.33), nonverbal ability (overall: F(1,227)=0.2, p=.42, spline: F(1,226)=0.0, p=.97).

Effects of Treatment Intensity

Overall, there was no main effect of treatment intensity on trajectories of any dependent variable: expressive language (overall change: F(1,225)=0.8, p=.36, spline: F(1,223)=0.7, p=.36)), receptive language (overall change: F(1,222)=0.0, p=.96, spline: F(1,221)=0.0, p=.92), nonverbal ability (overall change: F(1,223)=0.4, p=.54, spline: F(1,211)=0.4, p=.69), or autism severity (F(1,226)=0.8, p=.38, spline (F(2,224)=0.1, p=.78).

For autism severity, but not for the three developmental variables, there was a significant higher order interaction with time and site (F(2,226)=4.1, p=.02). When probing this three - way interaction effect, we found that Site 1 showed no time x intensity interaction effect on autism severity (F(1,230)=0.0, p=.87), Site 2 showed a significant time x intensity interaction effect on autism severity (F(1,223)=8.0 with greater improvement resulting from higher intensity intervention, and Site 3 showed no time x intensity interaction effect on autism severity (F(1,225)=1.2, p=.28).

Moderating Effects of Initial Severity on Effect of Treatment Intensity

Baseline autism severity did not moderate the effects of treatment intensity on the trajectories of the three developmental dependent variables: expressive language (overall: F(1,226)=1.1, p=.30, spline: F(1,226)=0.6, p=.44), receptive language (overall: F(1,223)=0.0, p=.93, spline F(1,223=1.3, p=.26), nonverbal ability (overall: F(1,224)=0.0 p=.84, spline: F(2,223)=1.6, p=.21).

Similarly, baseline DQ did not moderate effects of treatment intensity on the trajectories of the three developmental dependent variables: expressive language (overall: F(1,229)=3.5, p=.06, spline: F(1,227)=0.8, p=.37), receptive language (overall: F(1,228)=1.1, p=.30, spline: F(1,227)=1.4, p=.24), and non-verbal development (overall: F(1,228)=1.1, p=.30, spline: F(1,227)=1.4, p=.24).

In examining moderation effects on autism severity trajectory, we included site in the analyses because of the significant effects of site that we described above in the primary analyses of intensity effects. When exploring baseline DQ as a moderator to the primary model, we found significant main effects of treatment intensity on the trajectory of autism severity (overall: F(1,226)=5.1, p=.02, spline: F(1,222)=3.9, p=.05). This effect is conditional on the significant moderation effect of DQ (time x intensity x baseline DQ) when predicting the effects of intensity on the trajectory of autism severity (overall: F(1,226)=6.5, p=.01, spline: F(1,222)=4.2, p=.04).

Similarly, when exploring baseline autism severity as a moderator to the primary model, we found significant main effects of treatment intensity on the trajectory of autism severity (F(1,221)=4.1, p=.04). This effect is conditional on the significant moderation effect of baseline autism severity (time x intensity x baseline autism severity) when predicting the effects of intensity on the trajectory of autism severity (F(1,221)=4.5, p=.03). However, increasing the complexity of these results, there is also a significant three way moderation among site, baseline

autism severity, and time. That is, the moderation effects of baseline autism severity on the effects of treatment intensity on trajectory of autism severity differed by site (F(2,226)=3.6, p=.03). We did not probe these site effects due to lack of power (see Supplement 4, available online). The model is shown in Table 4.

Caregiver Satisfaction

Caregivers were highly satisfied with the treatment their children received, regardless of the style or intensity their child was assigned to. With "5" meaning "highly satisfied" on the caregiver satisfaction measure, means for the four cells were 4.7 (*SD* = .31 for EIBI low intensity), 4.8 (*SD* = .34 for EIBI high intensity), 4.8 (*SD* = .30 for ESDM low intensity), and 4.8 (*SD* = .19 ESDM high). Means were not significantly different among groups, $F(1,78) = 0.2$, $p = .73$.

DISCUSSION

What is the significance of the main findings?

While the science of early intervention for autism has a long history, dating back to the 1960's, and has been dominated by two different treatment approaches, those derived from the principles of ABA, and treatment derived from the principles of developmental science, few studies have been designed to provide answers to the questions that parents receiving a first diagnosis typically ask: what kind of treatment should I seek and for how many hours per week. Previous studies have documented that both types of treatment can result in significant child gains, but we have little comparative information from well-controlled studies. On the question of treatment intensity, we have no data from past experimental studies on which to base decisions; thus, recommendations for specific level of intensity have been based on assumptions, correlates, opinions, and common sense rather than on experimental evidence.

We conducted this trial to provide objective answers about relationships among initial child characteristics, treatment styles, treatment intensities, and child progress over time on key developmental and symptom domains. We used a rigorous, randomized, controlled multi-site study design, rigorous fidelity of implementation methods, and outcome assessments conducted by raters naïve to group membership. The four composite outcome variables: receptive and expressive language ability, nonverbal development, and autism severity were composed to represent constructs that have been associated with long-term outcomes in ASD [27].

We hypothesized that differences in child outcomes related to treatment style and intensity would be associated with initial severity of children's disability reflected in baseline DQ and autism symptom severity, as has been previously reported using quasi-experimental designs. The results did not support our primary hypothesis that initial degree of developmental delay and autism severity would moderate effects of *treatment style* on growth trajectories of children's outcomes. The severity of children's initial DQ and autism symptoms did not differentially affect the outcomes of EIBI versus ESDM intervention on children's progress in receptive language, expressive communication, nonverbal ability, and autism symptom change.

Although our rigorous fidelity of implementation methods ensured that the treatments were delivered as intended, one possible reason for the lack of treatment style effect is the tendency of both treatments to modify delivery style in specific, converging ways according to initial severity of disability. In the EIBI treatment, naturalistic teaching involving play, dyadic engagement, child initiation, and generalization occurred during child breaks from direct instruction. However, the direct instruction episodes were carried out at high levels of fidelity of implementation, as reported. In ESDM treatment, additional structure was added to by choosing and managing materials that would result in massed practice, by working at a table, and by

adding external cues of temporal structure to work times and play times. ESDM fidelity of implementation principles were maintained by: allowing for children's choice – making and initiation among the materials available, adult sensitivity and responsivity to child communications, adult use of developmentally appropriate language, management of child attention, modulation of child affect and arousal, use of turn-taking/dyadic engagement, adult positive affect, adult communications involving multiple pragmatic functions, smooth transitions between activities, working on language or social objectives within every activity, and theme and variation format within planned activities. Thus, the modifications were carried out within the basic fidelity principles and practices of each treatment.

Our second hypothesis predicted that initial degree of developmental delay and autism severity would moderate effects of *treatment intensity* on children's outcomes. When examining this hypothesis, we included site as an additional moderator because we found a significant higher order interaction involving site predicted change in the outcome variables. The hypothesis was supported for only one of the four outcome variables, change in autism severity, and it was conditional by site. We were under-powered to conduct within-site tests to determine the initial severity of disability values at which intensity levels mattered.

Does our finding indicate that 12-15 hours a week of comprehensive intervention is sufficient and that the NAS (2001) recommendation of 20+ hours is not supported?[28] It does not because, for at least one important variable – degree of autism symptoms – 25 hours of intervention was found to be more efficacious than 15 hours for improving core autism symptoms in one site. Furthermore, the study focused on toddler-age children, and it is possible that different findings could emerge for preschool age children or those with more years of intervention. Finally, because this is the first study to address these questions in a controlled fashion, replication is

necessary before practice alterations might be addressed. Replications should also address the weaknesses in this study associated with lack of power to conduct multiple within-site tests, as well as the lack of a no-treatment group, needed to assure that both treatments are responsible for child gains that occur. It is also important for future studies to examine the question of similarities and/or differences of underlying mechanisms of child learning involved in these two approaches. Further studies are needed to help us understand the relationships among individual differences in child characteristics, type of intervention delivery, and child outcomes.

Recommendations for clinicians:

Clinicians are challenged by caregivers' questions about their young ASD children's potential and the course of treatment to pursue. One recommendation from our data addresses caregiver questions about predicting outcomes for young children. Children receiving intensive early intervention can change substantially in their developmental rates and autism severity over time, and a better picture of the future will result from examining child progress and status after a few years of early childhood intervention, rather than at the time of diagnosis.

A second recommendation has to do with caregivers' questions about recommendations for 35-40 hours per week of behavioral treatment. It may be reassuring for caregivers to know that the field lacks high quality evidence that such a high number of hours provides greater gain for children, even those with more severe difficulties.

Finally, the "brand name" of interventions may be less important than more general characteristics of high quality intervention, which both interventions tested in this study met:

(1) a manualized, evidence-based approach that assesses children frequently and teaches to developmental and behavioral needs across all domains;

(2) incorporation of developmental and behavioral science;

(3) delivered consistently at fidelity across multiple environments, and

(4) integrated into everyday activities by caregivers regularly coached by program staff.

Perhaps the most important message that clinicians can provide early on is that, for both people with ASD and people without disability, quality of life across the life span is not determined by scores on tests, years of education, or income, but rather by quality of relationships with others and of engagement in personally satisfying activities, both at work and in play.

References

1. Dawson G., Rogers S., Munson J., et al. Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model. *Pediatrics*. 2010;125(1): e17-e23.
2. Smith, T., Groen, A., Wynn, J. (2000). Randomized trial of intensive early intervention for children with pervasive developmental disorder. *Am J Ment Retard*, 105(4), 269 – 285.
3. Rogers, S., Estes, A., Lord, C., Munson, J., Rocha, M., Winer, J., Greenson, J., Colombi, C., Dawson, G., Vismara, L., Sugar, C.A., Helleman, G., Whelan, C., Talbott, M. A multisite randomized controlled two-phase trial of the Early Start Denver Model compared to treatment as usual. *J Am Acad Child Adolescent Psychiatry* 2019; 58(9), 853-865.
4. Wetherby, A., Guthrie, W., JM Schatschneider, C., Holland, R., Morgan, L., Lord, C. Parent-implemented social intervention for toddlers with autism: an RCT. *Pediatrics*, 2014;134(6), 1084 – 1093.
5. Pickles, A., LeCouteur, A., Leadbitter, K., et al. Parent-mediated social community therapy for young children with autism (PACT); long-term follow-up of a randomized controlled trial. *The Lancet*, 2016;388(10059), 2501-2509.
6. Fossum, K., Williams, L., Garon, N., Bryson, S., Smith I. Pivotal response treatment for preschoolers with autism spectrum disorder: defining a predictor profile. *Autism Research*, 2018;11 (1), 153-165.
7. Tiura, M., Kim, J., Detmers, D., Baldi, H. Predictors of longitudinal ABA treatment outcomes for children with autism: A growth curve analysis. *Res Dev Disabil*, 2017;70, 185-197.
8. Smith, T., Klorman, R., Mruzek, D. Predicting outcome of community-based intensive behavioral intervention for children with autism. *J Abnorm Child Psychol*, 2015;43(7), 1271 – 1282.
9. Linstead, E., Dixon, D.,Hong, E., Burns, C., French, R., Novack, M., Granpeesheh, D. *Transl Psychiatry*, 2017;7(9); e1234.

10. Lovaas O. Behavioral treatment and normal educational and intellectual functioning in young autistic children. *J Consul Clin Psychol*. 1987;55: 3–9.
11. Yoder, P. J., Woynaroski, T. How to study the influence of intensity of treatment on generalized skill and knowledge acquisition in students with disabilities. *J Behav Edu*,2014; 1-15. PMID:25914513. PMC4405899.
12. Leaf, R., McEachin, J.*A Work in Progress*. New York, DRL publishing;1999.
13. Rogers S., Dawson G. The Early Start Denver Model for Young Children with Autism: Promoting language, learning, and engagement. NY: Guilford; 2010.
14. American Psychiatric Association. Diagnostic and statistical manual of mental disorders, 5th edition (DSM-5). Washington, DC: American Psychiatric Association; 2013.
15. Luyster R, Gotham K, Guthrie W, et al. The Autism Diagnostic Observation Schedule—Toddler Module: A New Module of a Standardized Diagnostic Measure for Autism Spectrum Disorders. *J Autism and Dev Disorders*. 2009;39(9):1305-1320. doi:10.1007/s10803-009-0746.
16. Mullen E. Mullen Scales of Early Learning: AGS edition. Circle Pines, MN: *American Guidance Service*; 1995.
17. Ratner, N., Bruner, J. Games, social exchange and the acquisition of language. *J Child Language*, 1978,5(3),391-401.
18. Swinkels S., Dietz C., Van Daalen E., Kerkhof I., van Engeland H., Buitelaar JK. Screening for autistic spectrum in children aged 14 to 15 months. I: The development of the Early Screening of Autistic Traits Questionnaire (ESAT). *JADD*. 2006;36(6):723-732.
19. Dietz C, Swinkels S, van Daalen E, van Engeland H, Buitelaar JK. Screening for autistic spectrum disorder in children aged 14–15 months. II: Population screening with the Early Screening of Autistic Traits Questionnaire (ESAT). Design and general findings. *J Autism Dev Disorder*. 2006;36(6):713-722.
20. Wetherby, A., Prizant, B. Communication and symbolic behavior scales developmental profile – First normed edition. Baltimore, MD; Brookes;2002.
21. Kleinman J., Robins D., Ventola P., et al. The modified checklist for autism in toddlers: A follow-up study investigating the early detection of autism spectrum disorders. *J Autism Dev Disorder*. 2008;38(5);827-839.
22. Robins D., Fein D., Barton M., Green J.The Modified Checklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *J Autism Dev Disorder*. 2001;31(2):131-144.
23. Gotham K., Pickles A., Lord C. Trajectories of autism severity in children using standardized ADOS scores. *Pediatrics*. 2012;30e1278-e1284.
24. Cohen, I., Schmidt-Lackner, S., Romanzyk, R., Sudhalter,V. The PDD Behavior Inventory: a rating scale for assessing response to intervention in children with pervasive developmental disorder. *J Autism Dev Disorder*, 2003; 33(1),31-45.
25. Fenson, L., Marchman, V., Thal, D., Dale, P., Reznick, S., Bates, E. MacArthur Bates Communicative Development Inventories MB-CDI *The MacArthur-Bates Communicative Development Inventories, User's Guide and Technical Manual*, Second Edition, Baltimore, MD: Brookes Publishing Group;2006.

26. Sparrow SS, Cicchetti DV, Balla, DA. *Vineland adaptive behavior scales (Vineland-II). 2nd ed.* San Antonio, TX: Pearson;2005.
27. Magiati, I., Tay, X.W., & Howlin, P.Cognitive, language, social, and behavioural outcomes in adults with autism spectrum disorders: A systematic review of longitudinal follow-up studies in adulthood. Clinical Psychology Review, 2014; 34(1), 73-86.
28. National Academy of Sciences -National Research Council, Washington, DC. Educating children with autism. Washington, DC: ERIC Clearinghouse; 2001.

Table 1. Components of the Primary and Moderating Composite Dependent Variables

composite DV	Component Variables
Autism symptom severity	
	ADOS-2 [a] calibrated severity score
	PDDB-I [b] expressive receptive social communication composite reflected raw score
Expressive communication	
	15-min communication sample; wfic [c]
	15-min communication sample; number of different word roots
	MSEL [d] expressive language age equivalency
	VABS-2 [e] expressive communication age equivalency
	MB-CDI [f] expressive raw score
	PDDB-I expressive social communication abilities composite raw score
	PDDB-I expressive language raw score
Nonverbal abilities	
	MSEL fine motor age equivalency
	MSEL visual reception age equivalency
	VABS-2 daily living skills age equivalency
	VABS-2 motor skills age equivalency
	VABS-2 socialization age equivalency
Receptive language	
	MSEL receptive language age equivalency
	VABS-2 receptive language age equivalency
DQ	MSEL expressive language age equivalency
	MSEL receptive language age equivalency
	MSEL fine motor age equivalency
	MSEL visual reception age equivalency

Note: ADOS-2 = Autism Diagnostic Observation Schedule, Second Edition; DQ = developmental quotients; DV = ??; MB-CDI = MacArthur-Bates Communicative Development Inventories; MSEL = Mullen Scales of Early Learning; PDD = pervasive developmental disorder; PDDB-I = PDD Behavior Inventory; VABS-2 = Vineland Adaptive Behavior Scales, Second Edition.

[a] Autism Diagnostic Observational Schedule version 2.
[b] Pervasive Developmental Disorders Behavior Inventory-parent rating form.
[c] Weighted frequency of intentional communication.
[d] Mullen Early Learning Scales.
[e] Vineland Adaptive Behavior Scales version 2.
[f] MacArthur-Bates Communicative Development Inventory, combined infant and toddler forms.

Table 2. Demographic Descriptions of the Participants

Variable	Total	Site A	Site B	Site C	Statistics	EIBI 25 hr	EIBI 15 hr	ESDM 25 hr	ESDM 15 hr	Statistics
Race, n (%)[a]					p=.04					p=0.93
African American	8.0% (7)	10.7% (3)	6.7% (2)	6.9% (2)		8.7% (2)	13.6% (3)	4.8% (1)	4.8% (1)	
American Indian / Alaska Native	1.1% (1)	3.6% (1)	0.0% (0)	0.0% (0)		4.3% (1)	0.0% (0)	0.0% (0)	0.0% (0)	
Asian	10.3% (9)	7.1% (2)	23.3% (7)	0.0% (0)		4.3% (1)	13.6% (3)	9.5% (2)	14.3% (3)	
Caucasian	54.0% (47)	39.3% (11)	50.0% (15)	72.4% (21)		60.9% (14)	40.9% (9)	57.1% (12)	57.1% (12)	
Multi	23.0% (20)	32.1% (9)	20.0% (6)	17.2% (5)		17.4% (4)	27.3% (6)	23.8% (5)	23.8% (5)	
Pacific Islander	1.1% (1)	3.6% (1)	0.0% (0)	0.0% (0)		0.0% (0)	0.0% (0)	4.8% (1)	0.0% (0)	
Unknown	2.3% (2)	3.6% (1)	0.0% (0)	3.4% (1)		4.3% (1)	4.5% (1)	0.0% (0)	0.0% (0)	
Ethnicity, n (%)[a]					p=.17					p=82
Hispanic/Latino	19.5% (17)	28.6% (8)	16.7% (5)	13.8% (4)		17.4% (4)	18.2% (4)	19.0% (4)	23.8% (5)	
Non-Hispanic	73.6% (64)	60.7% (17)	83.3% (25)	75.9% (22)		69.6% (16)	72.7% (16)	76.2% (16)	76.2% (16)	
Unknown	6.9% (6)	10.7% (3)	0.0% (0)	10.3% (3)		13.0% (3)	9.1% (2)	4.8% (1)	0.0% (0)	
Sex, n (%)[a]					p=.13					p=.99
Female	24.1% (21)	17.9% (5)	16.7% (5)	37.9% (11)		21.7% (5)	27.3% (6)	23.8% (5)	23.8% (5)	
Male	75.9% (66)	82.1% (23)	83.3% (25)	62.1% (18)		78.3% (18)	72.7% (16)	76.2% (16)	76.2% (16)	
Age (months), mean (SD)					$F_{2,84}$ = 3.6 p = .03					$F_{3,83}$ = 0.9 p = .46
Time 1	23.4 (4.0)	25 (3.6)	22.7 (4.2)	22.6 (3.8)		22.6 (3.5)	23 (4.7)	24.4 (4)	23.8 (3.6)	
Time 2	30.1 (4.1)	31.8 (3.6)	29.3 (4.3)	29.2 (4)		29.2 (3.8)	29.3 (4.7)	31.4 (4)	30.4 (3.8)	
Time 3	36.2 (4.1)	38.1 (3.4)	35.5 (4.3)	35.2 (3.9)		35.7 (3.4)	35.2 (4.8)	37.6 (4.2)	36.5 (3.7)	
Time 4	48.2 (4.3)	50.2 (3.9)	47.8 (4.4)	46.6 (3.9)		47.8 (3.8)	46.9 (5.1)	49.6 (4.4)	48.7 (3.8)	
Mullen Composite DQ										
Time 1	64.35 (18.55)	65.89 (17.43	73.13 (20.83)	54.36 (12.1)	$F_{2,84}$ = 9.07	66.96 (19.28)	64.25 (17.88)	61.36 (15.79)	65.05 (22.05)	$F_{3,83}$ = .20

		)			p<0.001					p=0.89
Time 4	77.88 (28.18)	75.75 (23.05	88.79 (29.59)	64.05 (25.41)	$F_{2,64}$ = 4.98 p=0.01	78.35 (29.96)	79.44 (29.07)	73.87 (25.55)	79.5 (29.68)	$F_{3,63}$ = 0.13 p=0.94
ADOS Severity Score										
Time 1	8.18 (1.73)	8.14 (1.75)	7.97 (1.75)	8.42 (1.7)	$F_{2,83}$ = .57 p=.57	8.19 (1.86)	8.38 (1.28)	8.29 (1.54)	7.82 (2.22)	$F_{3,82}$ = .76 p=0.52
Time 4	6.78 (2.25)	6.78 (2.09)	6.42 (2.42)	7.25 (2.24)	$F_{2,66}$ = .75 p=0.47	6.68 (2.06)	6.94 (2.54)	6.35 (2.37)	7.19 (2.17)	$F_{3,65}$ = 0.41 p=0.75
Outside Tx (avg hrs/wk) Year 1	1.90 (3.89)	2.18 (2.88)	1.57 (3.41)	1.97 (5.12)	$F_{2,84}$ = .18 p = .84	0.91 (2.17)	2.43 (3.94)	2.74 (5.87)	1.58 (2.62)	$F_{3,83}$ = 1.01 p = .39
Outside Tx (avg hrs/wk) Year 2	3.18 (6.77)	5.98 (9.78)	1.45 (3.91)	2.26 (4.56)	$F_{2,84}$= 3.88 P=0.02	3.53 (5.93)	4.20 (7.98)	3.13 (8.84)	1.78 (3.15)	$F_{3,83}$= 0.48 P=0.70
Income ($), mean (SD)	89439.7 (33573.8)	81018. 5 (34802)	99137.9 (30128.3)	62500 (35355.3)	$F_{2,55}$ = 2.88 p = .06	85000 (29730.9)	91428.6 (36554.9)	96833.3 (32956.5)	84285.7 (36956.7)	$F_{3,54}$ = 0.4 p = .72
Mother's Education (y), mean (SD)	16.2 (2.7)	16.8 (2.8)	17.1 (2.4)	14.8 (2.3)	$F_{2,81}$ = 7.1 p = .001	15.4 (2.8)	16.6 (3.1)	16.3 (2)	16.6 (2.8)	$F_{3,80}$ = 0.9 p = .44

Note: ADOS = Autism Diagnostic Observation Schedule; Avg = average: DQ = developmental quotient; EIBI = early intensive behavioral intervention; ESDM = Early Start Denver Model; hr = hours; n = number; TX=treatment

Table 3. Means, Standard Deviations, and Effect Sizes of the Primary and Moderating Variables by Site for the 4 Treatment Groups Including Effect Sizes at all Time Points

Composite variable label	Time Point	EIBI 25 hr	EIBI 15 hr	ESDM 25 hr	ESDM 15 hr
Aut_sx_comp	T1	1.27 (0.71)	1.07 (0.57)	1.24 (0.46)	1.08 (0.81)
Expr_com_comp	T1	-1.54 (0.27)	-1.34 (0.41)	-1.50 (0.22)	-1.48 (0.29)
Nv_dev_comp	T1	-1.52 (0.34)	-1.43 (0.42)	-1.45 (0.33)	-1.45 (0.24)
Recp_lang_comp	T1	-1.75 (0.41)	-1.62 (0.51)	-1.63 (0.47)	-1.55 (0.43)
Aut_sx_comp	T2	0.83 (0.81)	0.43 (0.74)	0.45 (0.92)	0.37 (0.72)
Expr_com_comp	T2	-1.13 (0.65)	-0.80 (0.48)	-0.99 (0.58)	-0.92 (0.57)
Nvdev_comp	T2	-1.00 (0.46)	-0.90 (0.45)	-0.85 (0.51)	-0.94 (0.33)
Recp_lang_comp	T2	-0.97 (0.81)	-0.84 (0.63)	-0.90 (0.67)	-0.87 (0.69)
Aut_sx_comp	T3	0.27 (0.69)	0.24 (0.79)	0.24 (0.90)	0.22 (0.56)
Expr_com_comp	T3	-0.69 (0.72)	-0.41 (0.62)	-0.65 (0.68)	-0.51 (0.61)
Nvdev_comp	T3	-0.58 (0.63)	-0.51 (0.65)	-0.59 (0.63)	-0.51 (0.45)
Recp_lang_comp	T3	-0.51 (0.81)	-0.41 (0.54)	-0.50 (0.85)	-0.47 (0.61)
Aut_sx_comp	T4	0.03 (0.93)	-0.11 (0.96)	-0.07 (0.94)	0.01 (0.85)
Expr_com_comp	T4	-0.01 (1.09)	0.07 (0.88)	-0.20 (0.95)	-0.00 (0.81)
Nvdev_comp	T4	0.03 (0.86)	0.04 (1.12)	-0.14 (0.76)	0.04 (0.71)
Recp_lang_comp	T4	-0.03 (0.95)	0.04 (0.99)	-0.09 (0.93)	-0.03 (0.93)

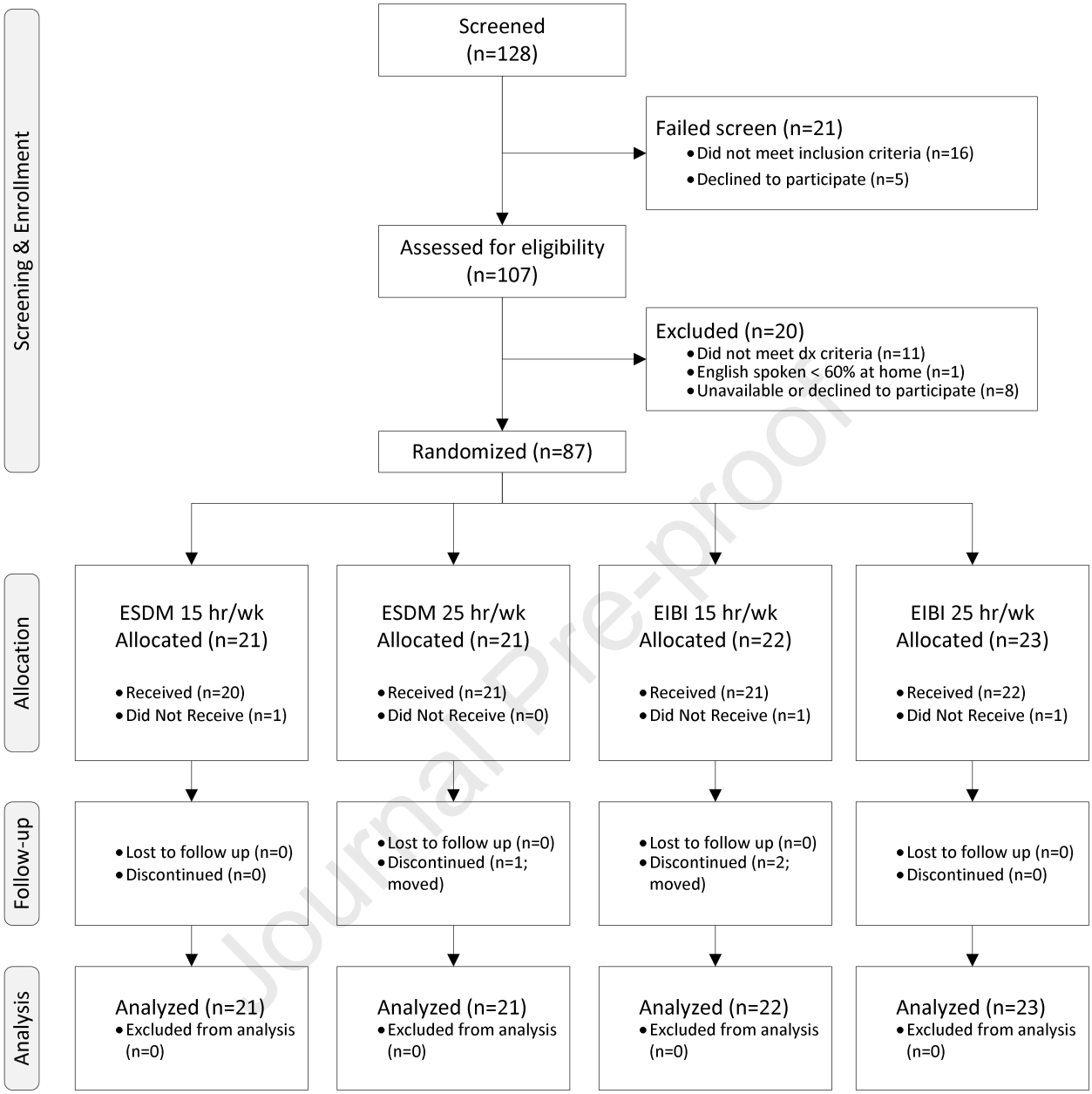
Note: All composite scores are based on the Time 4 reference point. Aut_sx_comp = Autism Symptomology Composite; EIBI = early intensive behavioral intervention; ESDM = Early Start Denver Model; Expr_lang_comp = expressive language composite: Nv_dev_comp = Nonverbal developmental composite: Recp_lang_comp = receptive language composite.

Table 4**.** The Moderation Model: Testing Moderation of Initial Autism Severity on Effects of Treatment Intensity on Trajectory of Change in Autism Severity over 24 Months

Dependent Variable: Autism Severity					
Type III Tests of Fixed Effects					
Effect	DFb	DFw	F	p	f^2
Intercept	1	146	0.0	0.84	0.00
Slope	1	221	6.1	0.02	0.03
Spline	1	221	0.5	0.50	0.00
Intensity	1	146	0.0	0.98	0.00
Intensity * Slope	1	221	4.1	0.04	0.02
Intensity * Spline	1	221	2.0	0.16	0.01
Site	2	151	0.2	0.84	0.00
Site * Slope	2	226	2.3	0.10	0.02
Site * Spline	2	224	1.9	0.16	0.02
Intensity * Site	2	151	0.1	0.91	0.00
Intensity * Site * Slope	2	226	3.6	0.03	0.03
Intensity * Site * Spline	2	224	2.0	0.14	0.02
Autism Severity	1	146	97.3	<.01	0.67
Slope * Autism Severity	1	221	19.6	<.01	0.09
Spline * Autism Severity	1	221	15.3	<.01	0.07
Intensity * Autism Severity	1	146	0.0	0.94	0.00
Intensity * Slope * Autism Severity	1	221	4.5	0.04	0.02
Intensity * Spline * Autism Severity	1	221	3.1	0.08	0.01

Note: Highlighted rows indicate significant interaction effects.

Figure 1. CONSORT Table



Supplement 1

Ensuring consistency across sites. Specific assessment administration and scoring practices involved a self-instructional course on administration techniques for all measures, requirement to achieve 85% of better agreement on procedural fidelity checklist and scoring reliability before collecting data, continuous 10% -15% of assessments fidelity coded, and deviations followed up with further training. In addition, the cross-site evaluation team met quarterly or more via internet conferencing to review difficult cases, administration problems, and other issues. We used a multi-method approach to measure and ensure FOI involving clearly defined assessment and teaching procedures, detailed protocols designed by the originators of each intervention delivery, extensive training, and frequent supervision and fidelity checks.[30]

Maintaining Blindness of Evaluation Team. The Evaluation Team and the Treatment Teams worked independently, including separate training, supervision, and establishment of reliability. At each phone call or contact with families by the Evaluation team, the family was reminded not to reveal anything about their child's treatment. Separate participant files were kept and protected by the two teams. Any Evaluation Team member for whom the randomization code was broken, a rare event, was replaced for that child's assessment.

Treatment Teams. Site interdisciplinary Treatment Teams included a speech/language pathologist, one or more behavior analysts, special educators, and a developmental or clinical child psychologist. The teams reviewed each child's progress, helped plan children's treatment objectives at regular intervals, and provided ongoing consultation to the treatment supervisor. The daily planning of each child's treatment was conducted by the treatment supervisor, who was one of the interdisciplinary, professional, graduate-educated team members and was accredited in the assigned approach. Therapy assistants (TAs) delivered the treatment in the

children's homes or care settings. TAs typically had Bachelor's degree and previous experience in discrete trial teaching in early autism. They received one to two months of full-time training: readings, homework, observations, hands-on training, and in vivo feedback and met rigorous clinical competency in either EIBI or ESDM before they began to deliver the intervention independently in homes. Both TAs and supervisors worked in only one approach. TAs were supervised live, by video recordings, or via live-streaming every 1-2 weeks.

An individual intervention plan was constructed for each child following the same overall process. After randomization, each child was assessed on the specified treatment curriculum. For the ESDM group, treatment objectives were written for each set of skills to be taught in the quarter based on the first failed or inconsistent passes on the curriculum.[13] Each quarterly objective was task broken down into a sequence of teaching steps. These teaching steps for each objective were condensed onto daily data sheets for recording performance data at 15-minute intervals during each session. For the EIBI group, treatment goals for each quarter were based on skills identified as not mastered from the curriculum assessment published in the treatment manual, "A Work in Progress" goals were broken into phases, as outlined in the EIBI manual.[12] Performance data on treatment goals were recorded trial by trial on data sheets during sessions.

Supplement 2

The Leaf and McEachin (1999) **A Work in Progress** manual defined the EIBI approach and the skills taught to each child**.**[12] Supervisors and ITs received ongoing training and consultation from one of the manual authors, including direct observation, feedback, and training within child sessions in clinic and home both live and via video recordings. The FOI tool used a 5-point Likert scale applied to randomly selected 20 minute sections of recorded treatment sessions to measure correct implementation of 9 components of the intervention style: behavior management; discreteness; discriminative stimulus; consequence; positive consequence; prompting; organization of session; response opportunity; and communication/attending temptations.[17] During instruction, children had brief play breaks interspersed between blocks of teaching trials, and a longer 10-15 minute break half way through each session during which the interventionist continued to interact with the child.

ESDM instruction, carried out as per manual instructions and measured by its published FOI tool, followed developmental and ABA principles.[1,13] Instruction followed the ESDM curriculum and was embedded in typical early childhood activities, carried out within the joint activity structure first defined by Ratner & Bruner (1978).[18] Adults followed children's choices into activities, embedded teaching objectives inside the activity using the activity or activity ending as the external reward. Activities were those that typically occur in the daily life of a toddler: toy play, social play, books, snacks, pretend play, handwashing, and outdoor play. Activities occurred in a separate room or a separate area of a room in the house or other setting, and they occurred in different locations on the floor, at a table, in a beanbag chair, etc. The FOI tool measured quality of instruction on a 14 item, 5 point Likert based scale from videos of 30 minutes or more of activities.[1,13]

Supplement 3

Relationships among the four composite variables measured with Cronbach's alpha.

tp	Aut sx comp[a]	expr com comp[b]	nvdev comp[c]	recp lan comp[d]
T1	0.44	0.86	0.80	0.73
T2	0.58	0.93	0.77	0.77
T3	0.48	0.95	0.85	0.79
T4	0.65	0.96	0.88	0.83

[a]Autism symptom severity composite. [b]Expressive communication composite. [c].Nonverbal abilities composite. [d].Receptive language composite.

Supplement 4

Preliminary examination of the Interaction of Baseline Disability Severity x Intensity Predicting Change in Autism Severity Within Site.

While we lacked sufficient power to test these results confidently, we conducted a preliminary exploration in order to provide some guidance for future study designs in this area. To do so, we calculated the balance points of the moderation effects. The following can be interpreted in this way: If the balance point falls outside of the confidence bounds, then initial differences do not moderate the effect of the treatment data. If the balance point falls inside the confidence bounds, this suggests that for one group, one intensity is better, and for the others the other intensity is better for outcomes. To aid interpretation, the percentile of the moderator's balance point and its confidence bounds are provided. We first describe site effects involving the baseline DQ moderator, and then we describe site effects involving the baseline autism severity moderator.

The baseline developmental quotient (DQ) moderator of intensity effects on autism severity change

Site 1: The balance point on initial DQ is 64 (54th %). The lower bound of the confidence band on initial DQ is 27, which is outside the site's range for baseline developmental quotient. The upper bound of the confidence band on initial DQ is 97 (95th %), indicating that those few participants with a baseline DQ of 97 or above for this site showed larger decreases in autism severity over time when receiving high intensity compared to low intensity treatment.

Site 2: The balance point on initial DQ is 45.5 (15th %). The lower bound of the confidence band is an illegal value for DQ (-60) and is thus uninterpretable. The upper bound of the confidence band on initial DQ is 68 (63rd %), indicating that participants with a baseline of 68 or above show more reduction in autism symptoms over time in the high intensity treatment than in the lower intensity treatment.

Site 3: The balance point on initial DQ is 68 (63rd %). The lower bound of the confidence band of initial DQ is 45 (15th %), and the upper bound of the initial confidence band of DQ is 133, which beyond the range of the developmental quotient for this sample. In this site, the children who initially scored lower than DQ 45 benefited more from the lower intensity treatment than the high intensity treatment.

The baseline autism severity moderator of intensity effects on autism severity change

Site 1: The balance point on initial autism severity - the level of baseline autism severity where lower and higher intensity groups are equal - is 0.9, the 25th percentile of autism severity. In this site, the lower and upper confidence bounds on autism severity are -2.8 and 3.1, which correspond to < 0th percentile, and >100th percentile. For children in this site, baseline autism severity does not show any significant effects of treatment intensity on trajectory of change in autism severity over time.

Site 2: The balance point on initial autism severity is 2.4, which is outside of the moderator range for this site. The lower boundary of the confidence band on initial autism severity, below which the effect size of the moderator favors higher intensity, is 1.3 (50^{th} percentile). The upper boundary of the confidence band, above which there is significant evidence that lower intensity is more beneficial, is 18.5, which is outside the range of the sample. In this site, children who scored below the 50^{th} percentile for the site on baseline autism severity (the milder end of the range of severity) are estimated to benefit significantly more from high intensity than low intensity treatment in terms of trajectory of change in autism severity.

Site 3: The balance point on initial autism severity is .05 (7^{th} percentile). In this site, the lower confidence bound on initial autism severity is -10 (i.e., < 0^{th} percentile) uninterpretable for this site's sample. The upper bound is 1.6 (77^{th} percentile), indicating that children in this site who scored above the 77^{th} percentile on baseline autism severity (e.g. the more severe end of the range of severity), are estimated to benefit significantly more from low intensity than high intensity treatment in terms of trajectory of change in autism severity.

With the exception of Site 2, for most participants, their baseline autism severity did not significantly influence effects of treatment intensity on change in autism severity over time.

Thus, for most participants, neither their baseline severity of DQ nor their baseline severity of autism symptoms significantly influenced effects of treatment intensity on change in autism severity over time. And where influences existed, they were in the opposite direction of the hypothesis.