2017

Community-Based Early Intervention For Children With Autism Spectrum Disorder

Allison S. Nahmias
University of Pennsylvania, allison.nahmias@gmail.com

Community-Based Early Intervention For Children With Autism Spectrum Disorder

Abstract
One in 68 children has been identified with Autism Spectrum Disorder (ASD), a disorder defined by 1) deficits in social-communication and social interactions and 2) restricted, repetitive patterns of behavior, interests or activities. Research has shown that children with ASD who receive high-quality early intervention (EI) services in university-based research trials can make large gains in cognitive, communication, and adaptive behaviors skills, with positive long term effects. However, less is known about the outcomes for the over 50,000 children who receive EI in community settings. This dissertation provides initial evidence of the current state of community-based EI for children with ASD. Chapter 1 presents a meta-analysis of cognitive, communication, social, and adaptive behavior outcomes for children with ASD in community-based EI programs, and demonstrates that the gains made in the community are much smaller than those observed in university-based trials. In Chapter 2, prospective, longitudinal data collected from a local EI system is studied to understand which characteristics of preschool EI predict cognitive gains for 79 preschoolers with ASD that received publicly-funded services in classroom placements. The best predictor of gains was the utilization of recommended intervention practices to support the development of social and peer relationships. Chapter 3 discusses measurement of executive functioning (EF) among preschoolers with ASD, as executive functioning skills likely play an important role in response to EI. However existing EF measures have not been validated for use with low-functioning, nonverbal preschoolers with ASD. Results are presented from the development and the validation of a battery of nonverbal, performance-based EF tasks. These measures can be utilized in future community-based treatment trials.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Psychology

First Advisor
David S. Mandell

Keywords
Autism Spectrum Disorder, community settings, early intervention, executive functioning, meta-analysis, preschool

Subject Categories
Clinical Psychology

COMMUNITY-BASED EARLY INTERVERVENTION FOR CHILDREN WITH AUTISM SPECTRUM DISORDER

Allison S. Nahmias

A DISSERTATION

in

Psychology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

____

David Mandell, ScD
Professor of Psychiatry

Graduate Group Chairperson

____

Sara Jaffee, PhD
Professor, Psychology

Dissertation Committee

David Mandell, ScD	Professor of Psychiatry
Robert Schultz, PhD	Professor of Psychiatry
Sara Jaffee, PhD	Professor, Psychology

ACKNOWLEDGMENTS

I am eternally grateful for the mentorship that I have received from my advisor, David Mandell. David opened the door for me to the complicated, often messy, but important world of community-based research and has inspired my career in implementation science. He has stretched my thinking and pushed me to become a better researcher, while always reminding me that "perfect is the enemy of good." I am also grateful for the guidance that I have received from my committee members throughout graduate school: Sara Jaffee, Robert Schultz, and Dianne Chambless. Sara, Bob, and Dianne have all provided feedback and mentorship that has fostered my growth as a clinical scientist.

I also sincerely thank my colleagues and collaborators at the Center for Mental Health Policy and Services Research and Center for Autism Research for all of their help and support with this research and throughout graduate school. It truly takes a village to complete community-based research and I could not have done it without them, especially the following current and past lab members, externs, and student workers: Samantha Crabbe, Erica Blanch, Jill Locke, Heather Nuske, Julia Parish-Morris, Melanie Pellecchia, Keiran Rump, Ben Yerys, Ashley Brightwell, Ligia Antezana, Fam Baguio, Jen Bertollo, Anjeli Elkins, Emily Ferguson, Rachel Ouellete, Steve Marcus, Elizabeth Ayad, Lauren Brumley, Kevin Donley, Mina Kim, Simone Milaresis, Joseph Pereira, Kiri Baga, Elise Dadourian, Lauren Ashley Daley, Amanda Dennis, Jordan Doman, Emma Finkel, Stephanie Gregory, Helen Jin, Molly Johnson, Cary Kane, Vaishali Mahalingam, Mark Miller, Yoon Ji Moon, Rachel Orlinsky, Chaereen Pak, and Liza Tomczuk.

I am so grateful to my fellow graduate students for their friendship and unconditional support over the years, including Kelly Allred, Moriah Brier, Gabi Kattan Khazanov, Gwen Lawson, Lorenzo Lorenzo-Luaces, Sarah Morris, Eliora Porter, Annie Roepke, Laura Sockol, Lisa Yankowitz, and Hana Zickgraf. I especially want to thank my amazing "lab-sister" Caitlin Clements, who always goes above and beyond. You each inspire me. I also thank my internship cohort and mentors at the University of California of Los Angeles, especially Sima Finy, Rowena Ng, and Karin Best, for their unconditional support and frequent reminders of what is possible. I am also forever grateful to my family for their patience, unwavering support, love, and for always believing in me.

Thank you to the children, families, teachers, and early intervention providers that participated in the MOCA project. I appreciate your willingness to allow us into your homes and schools for the shared goal of improving community-based early intervention for children with autism spectrum disorder.

Finally, I would like to acknowledge my funding sources. The research reported here was supported in part by the Institute of Education Sciences, U.S. Department of Education, through grant #R305B090015 to the University of Pennsylvania. The opinions expressed are those of the authors and do not represent the views of the Institute or the U.S. Department of Education. This research was also supported by Autism Speaks grant #8874 (PI: Mandell) and a McMorris Autism Foundation grant (PIs: Nahmias and Yerys).

ABSTRACT

COMMUNITY-BASED EARLY INTERVENTION FOR CHILDREN WITH AUTISM SPECTRUM DISORDER

Allison S. Nahmias

David S. Mandell

One in 68 children has been identified with Autism Spectrum Disorder (ASD), a disorder defined by 1) deficits in social-communication and social interactions and 2) restricted, repetitive patterns of behavior, interests or activities. Research has shown that children with ASD who receive high-quality early intervention (EI) services in university-based research trials can make large gains in cognitive, communication, and adaptive behaviors skills, with positive long term effects. However, less is known about the outcomes for the over 50,000 children who receive EI in community settings. This dissertation provides initial evidence of the current state of community-based EI for children with ASD. Chapter 1 presents a meta-analysis of cognitive, communication, social, and adaptive behavior outcomes for children with ASD in community-based EI programs, and demonstrates that the gains made in the community are much smaller than those observed in university-based trials. In Chapter 2, prospective, longitudinal data collected from a local EI system is studied to understand which characteristics of preschool EI predict cognitive gains for 79 preschoolers with ASD that received publicly-funded services in classroom placements. The best predictor of gains was the utilization of recommended intervention practices to support the development of social and peer

relationships. Chapter 3 discusses measurement of executive functioning (EF) among preschoolers with ASD, as executive functioning skills likely play an important role in response to EI. However existing EF measures have not been validated for use with low-functioning, nonverbal preschoolers with ASD. Results are presented from the development and the validation of a battery of nonverbal, performance-based EF tasks. These measures can be utilized in future community-based treatment trials.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Chapter 2

CHAPTER 1: Effectiveness of Community-based Early Intervention for Children with Autism Spectrum Disorder: A Meta-Analysis

Abstract

The present study comprises a meta-analysis of the effectiveness of community-based early intervention (EI) programs for children with autism spectrum disorder (ASD). While university-based trials of EI programs for children with ASD generally produce medium-to-large gains on average, less is known about the results from community-based intervention. A systematic search identified 40 groups from 29 studies assessing change in cognitive, communication, social, or adaptive behavior skills from pretreatment to posttreatment. There was significant improvement in each of the domains, however, the gains were small. Uncontrolled effect sizes (Hedges *g*) ranged from 0.21 for adaptive behavior to 0.31 for communication outcomes, after removal of outliers and correction for publication bias. "Model" EI programs (e.g., those associated with universities and hospitals) were generally superior to other community EI program types across all four outcomes. Only communication outcomes demonstrated increasingly larger effect sizes in more recent years. These results suggest that there remains a large gap between research and community practice. Implications of the findings for clinical practice and future research are discussed.

Keywords: Autism Spectrum Disorder, meta-analysis, early intervention, community settings

Effectiveness of Community-based Early Intervention for Children with Autism Spectrum Disorder: A Meta-Analysis

A growing body of research demonstrates that early intervention provided in university-based research settings by experts can result in large gains in cognition, communication, social skills, and adaptive behavior for young children with ASD. The interventions with most evidence are early intensive behavior interventions (EIBI) and applied behavior analysis (ABA). For example, a recent systematic review identified moderate to large effects in improving cognitive functioning and language skills for early interventions based on high-intensity ABA, usually delivered in a university-based setting (Weitlauf et al., 2014). The results of a meta-analysis by Ospina and colleagues (2008) indicated statistically and clinically significant positive effects on intellectual functioning, adaptive behavior, communication and language when high-intensity Lovaas-based ABA intervention was compared with either low-intensity Lovaas (standardized mean difference = 0.92) or special education (standardized mean difference = 0.95). A Cochrane Collaboration systematic review and meta-analysis of EIBI compared with treatment as usual in the community found medium to large significant positive effects for adaptive behavior (g = 0.69), language (g = 0.50 - 0.57), daily communication skills (g = 0.74), IQ (g = 0.76), socialization (g = 0.42), and daily living skills (g = 0.55) (Reichow, Barton, Boyd, & Hume, 2012). Other meta-analyses of ABA and EIBI also report medium to large positive gains on outcomes such as intellectual ability, adaptive behavior, and communication (Eldevik et al., 2009; Makrygianni & Reed, 2010; Reichow, 2012; Strauss, Mancini, & Fava, 2013; Virués-Ortega, 2010).

Naturalistic Developmental Behavioral Interventions, which incorporate ABA principles, also have been identified as evidence-based practices for young children with ASD (Schreibman et al., 2015). A meta-analysis of one such intervention, joint attention interventions, found significant positive effects on joint attention. Overall Hedges's g ranged from 0.53 to 0.76, depending on the type of control group (Murza, Schwartz, Hahs-Vaughn, & Nye, 2016).

These studies speak to the gains possible with highly structured (usually) university-based early intervention delivery, but do not speak to what is probable in "treatment as usual" or standard care received outside the context of research. Evidence-based interventions rarely make their way into community practice (Hess, Morrier, Heflin, & Ivey, 2008; Stahmer, Collings, & Palinkas, 2005). Although research has begun to demonstrate that community providers can be trained in evidence-based behavioral interventions with positive child outcomes (Shire & Kasari, 2014; Shire et al., 2017), much less research has examined what gains children with ASD make when receiving typical early intervention in the community. Only a few studies have reported on child outcomes associated with community-based intervention and these studies suggest that the gains are not as large as those seen in university-based settings. Often these studies include data only from a single site or from a small sample, which limits the generalizability of the results (Ben-Itzchak, Watson, & Zachor, 2014; Magiati, Charman, & Howlin, 2007).

To date, there has been no systematic, empirical review or meta-analysis of community outcomes for young children with autism. The present study takes advantage of the fact that most comparative trials of early intervention for children with autism provide evidence of community outcomes, in that the "treatment as usual" or control group comprises an assessment of the effectiveness of community intervention. Here we combine them to provide a more rigorous assessment of the state of community-based interventions and explore patterns in the results. In addition, we use the variation in these studies to examine the effects of placement characteristics on child outcome.

Quantifying these outcomes using meta-analytic techniques serves several purposes. First, it provides a benchmark against which other community programs can be measured. Second, it has the potential to identify models of excellence that can be emulated. Third, it allows us to examine program characteristics that may be associated with positive outcomes, which is important for program development. Fourth, it allows us to explore whether there have been changes over time in the effectiveness of intervention provided in the community. Finally, it serves as an assessment of the penetration of research to practice.

Methods

Search Procedures and Selection of Studies

Studies were eligible for inclusion in the meta-analysis if they met the following criteria:

a) Published study, written in English
b) Prospective study that utilized pre-test, post-test group design
c) Presented outcomes for children identified with ASD separately
d) More than 10 children with ASD receiving community-based intervention

e) Child age at study intake was less than 73 months, which corresponds with the typical age of early intervention in the United States.

f) The study provided information on outcomes of educational or behavioral services available in the community or treatment as usual (could be a group of participants within a study). Groups that received intervention provided by researchers were excluded.

g) Outcome measures included at least one of the following, reported as standard scores or developmental quotients (standard scores were required to partially account for potential maturation effects):

    a. Cognitive: Early Learning Composite from the Mullen Scales of Early Learning (MSEL, Mullen (1995)), or Full Scale IQ. Studies that only included non-verbal IQ or a cognitive measure that only assessed non-verbal IQ (e.g., Merril-Palmer (Roid & Sampers, 2004), Leiter International Performance Scale-Revised (Roid & Miller, 1997)) were excluded to minimize measurement differences.

    b. Communication: Vineland Adaptive Behavior Scales Communication domain (Sparrow, Balla, & Cicchetti, 1985; Sparrow, Cicchetti, & Balla, 2005)

    c. Social: Vineland Adaptive Behavior Scales Socialization domain (Sparrow et al., 1985, 2005)

d. Adaptive Behavior: Vineland Adaptive Behavior Composite (Sparrow et al., 1985, 2005)

h) Reported unadjusted pre- and post- intervention means and standard deviations for outcome measures (based on recommendation from the What Works Clearinghouse (2014)) and so that all effect sizes were calculated utilizing the same method

i) Studies only reporting follow-up data were excluded

j) For studies with overlapping (or potentially overlapping) samples, the study published with the largest sample for each outcome was utilized

A systematic search of research databases was initially conducted through August 2015 to identify relevant studies. Databases available through the University of Pennsylvania Library, including PsycINFO and Medline were searched for terms related to *autism* and *intervention* (see Appendix A for a sample search strategy). The reference list of retrieved articles, existing reviews, and meta-analysis were also examined for eligible studies. The search was then updated, searching through January 2017. As the Medline search did not provide any unique studies that met inclusion criteria, it was not included in the search update.

Study selection was conducted in three stages. Studies were first screened for eligibility based on the title and abstract using the following exclusion criteria: a) did not include children with ASD, b) $n < 10$ children with ASD, c) article was written in a language other than English, d) participants were outside the age range (i.e., older than 6 years old) or the study did not analyze children less than six years of age separately, and

e) presented the results of a drug or medication study that did not also include a behavioral intervention. Screening was conducted by the first author and a coder trained to reliability. Studies then underwent full-text review for eligibility by the first author and two coders trained to reliability. The first author then completed final review of all articles and subgroups within the articles based on final inclusion criteria (described above). See Figure 1 for the PRISMA flow diagram (Liberati et al., 2009).

INSERT FIGURE 1 HERE.

Coding of Studies

The following data items regarding participant, intervention, and study characteristics were coded from each article: percentage of male participants, percentage of non-Caucasian participants, mean age of participants, baseline IQ of participants, country of intervention, years during which the intervention took place, EI duration, intensity of intervention (e.g., hours/week), baseline and post-intervention means and standard deviations for cognitive, communication, social, and adaptive behavior outcomes, and name of cognitive measure utilized. One study (Cohen, Amerine-Dickens, & Smith, 2006) did not provide the unadjusted post-treatment standard deviations in the original paper, however, they were reported in a recent Cochrane systematic review and meta-analysis (Reichow et al., 2012) and so this information was extracted from that paper. The category of EI was also coded based on the following criteria: a) "Model" programs were defined as intervention programs providing intervention in the community associated with universities and/or hospitals, b) "Treatment As Usual" (TAU) programs were defined as specific treatment as usual program, treatment as usual from local school/agency, or standard educational provisions, and c) the "Variable EI" category included participants in a wait-list group, services as usual in the community where participants received an unclear variety of different services and some participants may not have gotten any services. Data was extracted from articles by the first author.

Analyses

All outcome data were continuous. Changes between baseline and posttreatment assessments were assessed utilizing standardized mean gain scores. Positive values reflect improvements in cognitive, communication, social, and adaptive skills over the course of treatment. Uncontrolled effect sizes standardized mean gain scores were utilized as the principal summary score. Effect sizes were calculated by dividing the mean change from baseline to post-treatment by the pooled standard deviation of the difference score. The effect sizes were transformed to Hedges's *g* estimates (Hedges, 1981) to correct a potential bias due to small sample sizes.

As no included studies reported the pretest-posttest correlation for the selected outcome measures, or provided the data needed for these values to be calculated, per the recommendation of Lipsey and Wilson (2001), test-retest reliabilities from test manuals and published papers were utilized as a proxy. The average was utilized when multiple test-retest scores were reported. As the test-retest reliability may overestimate the pretest-posttest correlations, sensitivity analyses with *r* values of 0.3, 0.5, and 0.8 as estimates of low, medium, and high correlations were conducted. Overall effect sizes were similar, so the test-retest reliabilities were determined to be acceptable approximations.

Potential outliers were detected using the sample-adjusted meta-analytic deviancy (SAMD) statistic, as a failure to exclude extreme studies may result in overestimation of the true variability (Huffcutt & Arthur, 1995). A conservative cutoff of the absolute value of 2.58 was utilized to consider groups for exclusion from analyses. As extreme values can result both from error and true population variability, the ability to assess the role of moderators is limited when outliers whose effects represent true population variability are removed (Beal, Corey, & Dunlap, 2002). The SAMDs were rank-ordered and scree plots were examined to confirm the outlier status of groups with SAMDs above the 2.58 cutoff.

Calculations of weighted mean effect sizes, heterogeneity, moderators, and publication bias statistics were conducted using Comprehensive Meta-Analysis Version 2.2.064 (Borenstein, Hedges, Higgins, & Rothstein, 2005). Separate random effects model meta-analyses were conducted for cognitive, communication, social, and adaptive behavior outcomes to assess the effects of community-based EI for different domains of functioning, and were chosen over fixed effects models for conceptual reasons, as recommended by Bornstein, Hedges, Higgins, and Rothstein (2009) and Lipsey and Wilson (2001). Fixed effect models assume that variability in effect sizes is due to random error within studies, and that there is a common true effect size across all studies. The overall effect size represents the estimate of the true effect size for the population of studies, but is not generalizable beyond the sample of included studies. In contrast, random effects models assume that variability in effect sizes is due to both random error within studies and systematic variability between studies, and the true effect size is allowed to vary across studies. Overall effect size in a random effects model represents the estimated average of the true effect sizes, and the results can be generalized to studies not included in the analysis. Random effects analyses were used to model two aspects of the observed variance: random within-study variance and systematic between-study variance. Each effect size was weighted to account for its relative precision based on the standard error of the effect size (within-study variance) and tau-squared (between-study variance) using the reciprocal of the squared standard error plus tau-squared. Study quality was not used to weight effect sizes as the study characteristics to assess study quality were inconsistently reported.

Heterogeneity of effect sizes was examined using the Q statistic and the I^2 statistic. The Q statistic tests the hypothesis that the observed variance in effect sizes is no greater than that expected by sampling error alone. A significant Q statistic indicates that the observed range of effect sizes is significantly larger than would be expected based on within-study variance. While a significant Q statistic indicates heterogeneous effect sizes, nonsignificant Q statistics should be interpreted with caution, as heterogeneous effect sizes may yield a nonsignificant Q value due to low power. The I^2 value indicates the proportion of variance in effect sizes accounted for by between-study variance and has a range from 0 to 100 (Higgins & Thompson, 2002). The I^2 values of 25, 50, and 75 are interpreted as low, moderate, and high levels of heterogeneity, respectively (Higgins, Thompson, Deeks, & Altman, 2003). An I^2 in the low range suggests that the effect sizes are homogeneous relative to the precision of the individual studies.

Exploratory moderator analyses were conducted for models with a significant *Q* statistic or an I^2 at or above 50. Categorical moderators were examined using an analysis-of-variance (ANOVA) of mixed-effect models for each variable hypothesized to moderate the overall effect size. Meta-regression analyses were used to examine continuous moderators. Variables related to participant (e.g., age) and intervention (e.g., country, duration) characteristics were included in the moderation analyses. Due to the relatively small number of studies included in this meta-analysis, only one potential moderator was included in the meta-regression at a time.

To assess publication bias for all four outcomes, funnel plots and Duval and Tweedie's (2000) trim-and-fill procedure were calculated. First, funnel plots were created by plotting each study's effect size against its standard error. An asymmetric distribution suggests missing studies due to publication bias. Duval and Tweedie's (2000) trim-and-fill procedure provides an effect size estimate that corrects for the number and assumed location of the missing studies when asymmetry in the funnel plot is indicated. The overall estimates for the model were calculated using the trim-and-fill correction when this test indicated significant asymmetry in the funnel plot.

Results

Study Characteristics

Table 1 displays sample characteristics of the 40 groups from 29 studies included in the analysis. Participants were predominantly male (mean percentage across groups 84.9 (SD = 8.5)). Seventeen studies (59%) reported gender by group, while seven studies

(24%) only reported the gender of the entire study sample. Nine studies (31%) reported sufficient data on the race and ethnicity of participants. Six studies (21%) reported this information by group, while three studies (10%) only reported racial information of the entire study sample. The percentage of participants identified as non-Caucasian ranged from 24.5% (Rogers et al., 2012) to 72.6% (Baker-Ericzén, Stahmer, & Burns, 2007), with a mean of 37.3% and a standard deviation of 14.8%. Ten groups (25%) from nine studies only included children less than three years old (Ben-Itzchak et al., 2014; Carter et al., 2011; Dawson et al., 2010; Klintwall, Macari, Eikeseth, & Chawarska, 2015; Rogers et al., 2012; Schertz, Odom, Baggett, & Sideris, 2013; Stahmer, Akshoomoff, & Cunningham, 2011; Turner-Brown, Hume, Boyd, & Kainz, 2016; Zachor & Ben-Itzchak, 2010). Two groups (5%) from two studies only included children between three and six years old (Baker-Ericzén et al., 2007; Rickards, Walstab, Wright-Rossi, Simpson, & Reddihough, 2007). The rest of the groups included children between 18 and 72 months old. Across the groups the mean age of participants was 37.4 months (SD = 9.7). Twenty-four groups (60%) reported cognitive outcomes eligible for inclusion in the meta-analysis, 23 (58%) reported communication outcomes, 23 (58%) reported social outcomes, and 24 groups (60%) reported adaptive behavior outcomes eligible for inclusion in the meta-analysis.

INSERT TABLE 1 HERE

Table 2 displays characteristics of the interventions in the 40 included groups. Studies described interventions that took place in a variety of countries. Fifteen groups (37.5%) from 12 studies occurred in the United States, nine groups (22.5%) from six studies occurred in the United Kingdom, six groups (15%) from four studies occurred in Australia, three (7.5%) groups from two studies occurred in Israel, two groups (5%) from two studies occurred in Italy, two groups (5%) from one study occurred in Norway, two groups (5%) from one study occurred in Sweden, and one group (2.5%) from one study occurred in Canada. Only 19 groups (47.5%) from 11 studies reported the years over which the intervention occurred. Intervention years ranged from 1995 to 2003 (Cohen et al., 2006) to 2012 to 2014 (Turner-Brown et al., 2016). Intervention duration ranged from three months (Anan, Warner, McGillivary, Chong, & Hines, 2008; Baker-Ericzén et al., 2007; Rogers et al., 2012) to 36 months (Cohen et al., 2006), with a mean of 14.1 months and a standard deviation of 8.0. Thirty-three groups (82.5%) from 24 studies reported some information regarding intervention intensity. The type of community-based EI provided varied between the groups. Twenty-four groups (60%) from 16 studies described treatment as usual EI programs. Nine groups (22.5%) from eight studies described Model treatment programs associated with hospitals or universities. Seven groups (17.5%) from seven studies reported outcomes for children receiving a variety of EI services that varied in the amount and type of intervention received.

INSERT TABLE 2 HERE

Uncontrolled Effect Sizes

Tables 3 and 4 present the uncontrolled effect sizes and the results of the random effects models for cognitive, communication, social, and adaptive behavior outcomes, representing results from 40 groups from 29 studies. These values should be interpreted with caution as they reflect within-study change and cannot differentiate changes that resulted due to the intervention as opposed to the passage of time.

INSERT TABLE 3 HERE

INSERT TABLE 4 HERE

Cognitive. Twenty-four groups from 17 studies (744 participants) reported the results from eligible cognitive outcomes. There was variability in the effect of community-based EI on participants' cognitive scores, with Hedges's *g* ranging from -0.43 to 1.50. Sixteen groups (66.7%) demonstrated significant positive effects, indicating improvement over baseline cognitive scores. Two groups (8.3%) demonstrated positive effects that were marginally significant ($p < .08$, the PACTS group from Reed et al (2010) and the control group from Tonge et al (2014)). One group (4.1%) from one study demonstrated a significant negative effect, indicating a decline in cognitive scores over the course of treatment (Rickards et al., 2007). Four groups (16.7%) from four studies reported cognitive scores that did not significantly change over the course of the intervention. The early intensive behavior analytic treatment group (IBT) from one study (Howard, Sparkman, Cohen, Green, & Stanislaw, 2005) had a SAMD value greater than 2.58, so this group was excluded from subsequent analyses. The average effect size excluding this outlier was small (0.30, 95% CI 0.20 - 0.40, $p < .001$), see Appendix B Figure 1B for the forest plot without this outlier.

The *Q* statistic indicated that there was significant heterogeneity among the cognitive effect sizes ($p < .001$). The I^2 value indicated a high level of heterogeneity, with 83% of the variance in effect sizes attributable to between-study variance. The funnel plot was slightly asymmetric (see Appendix C Figure 1C). Trim-and-fill procedures suggested that three studies with effect sizes to the left of mean were missing, suggesting a publication bias that overestimates the true effect size of community-based early intervention on cognitive results. The corrected average effect size was 0.24 (95% CI 0.13-0.35).

Communication. Twenty-three groups from 17 studies (797 participants) reported the results from eligible communication outcomes. Although Ben-Itzchak (2014) also reported communication outcomes, this group was excluded from the analyses due to the potential overlapping sample with Zachor and Ben-Itzchak (Zachor & Ben-Itzchak, 2010). There was variability in the effect of community-based EI on participants' social scores, with Hedges's *g* ranging from -0.26 to 0.70. Seventeen groups (73.9%) demonstrated significant positive effects, indicating improvement over baseline communication scores. Six groups (26%) from six studies reported communication scores that did not significantly change over the course of the intervention. The average effect size was small (0.31, 95% CI 0.22 - 0.41, $p < .001$). See Appendix B Figure 2B for forest plot. No outliers were identified.

The Q statistic indicated that there was significant heterogeneity among the communication effect sizes ($p < .001$). The I^2 value indicated a high level of heterogeneity, with 85% of the variance in effect sizes attributable to between-study variance. The funnel plot was symmetric (see Appendix C Figure 2C) and trim-and-fill procedures did not suggest any missing studies.

Social. Twenty-three groups from 17 studies (857 participants) reported the results from eligible communication outcomes. Although Ben-Itzchak (2014) also reported social outcomes, this group was exclude from analyses due to the potential overlapping sample with Zachor and Ben-Itzchak (2010). There was variability in the effect of community-based EI on participants' social scores, with Hedges's *g* ranging from -0.96 to 0.75. Sixteen groups (69.6%) demonstrated significant positive effects, indicating improvement over baseline social scores. Six groups (26%) from 5 studies reported social scores that did not significantly change over the course of the intervention. One study (Dawson et al., 2010) had a SAMD value less than -2.58, so this study was excluded from subsequent analyses. The average effect size excluding this outlier was small (0.26, 95% CI 0.14 - 0.37, $p < .001$), see Appendix B Figure 3B for forest plot.

The *Q* statistic indicated that there was significant heterogeneity among the communication effect sizes ($p < .001$). The I^2 value indicated considerable heterogeneity, with 88% of the variance in effect sizes attributable to between-study variance. The funnel plot was symmetric (see Appendix C Figure 3C) and trim-and-fill procedures did not suggest any missing studies.

Adaptive Behavior. Twenty-four groups from 19 studies (1,028 participants) reported results from eligible adaptive behavior outcomes. There was variability in the effect of community-based EI on participants' adaptive behavior scores, with Hedges's *g* ranging from -1.25 to 0.95. Fourteen groups (60.9%) demonstrated significant positive effects, indicating improvement over baseline social scores. Two groups (8.7%) from 2 studies reported adaptive behavior scores that significantly decreased over the course of the intervention. The adaptive behavior scores for the other eight groups (33.3%) did not significantly change over the course of the intervention. One study (Dawson et al., 2010) had a SAMD value less than -2.58, so this study was excluded from subsequent analyses. The average effect size excluding this outlier was small (0.21, 95% CI 0.11 - 0.30, $p < .001$). See Appendix B Figure 4B for the forest plot without this outlier.

The Q statistic indicated that there was significant heterogeneity among the adaptive behavior effect sizes ($p < .001$). The I^2 value indicated a high level of heterogeneity, with 91% of the variance in effect sizes attributable to between-study variance. The funnel plot was symmetric (see Appendix C Figure 4C) and trim-and-fill procedures did not suggest any missing studies.

Moderator Analyses

As both the Q statistic and I^2 index indicated significant heterogeneity of effect sizes for all four outcomes, exploratory analyses of potential moderators were conducted. These analyses assessed whether effect sizes differed based on the characteristics of the included groups and interventions.

Study, sample, and intervention characteristics. Two intervention characteristics were examined as potential categorical moderators: EI category and the country in which intervention took place. As a reminder, EI category was defined as follows: “Model” programs were intervention programs associated with universities and/or hospitals, “TAU” programs were specific treatment as usual program, treatment as usual from local school/agency, or standard educational provisions, and the “Variable EI” category included participants in a wait-list group, services as usual in the community where participants received an unclear variety of different services and some participants may not have gotten any services. Age of the sample at intake, intervention duration, approximate hours of intervention, and year of publication were examined as continuous variables. Year of publication was utilized as a proxy for the recency of the intervention, as less than half of the groups reported when the intervention occurred.

Cognitive. As seen in Table 5, all three EI categories had significant positive effect sizes. Children receiving intervention in Model EI programs made moderate gains (Hedges's $g = 0.51$), while children receiving treatment as usual and variable EI made small gains (Hedges's $g = 0.25$ and 0.24 respectively). The differences among the three EI categories reached a marginal level of significance ($p = 0.060$). Interventions conducted in the United States and United Kingdom had significant positive effects on cognitive scores (Hedges's $g = 0.48$ and 0.22 respectively) and the effects on cognitive scores for interventions conducted in Norway reached marginal significance ($p = 0.052$). Interventions conducted in Australia, Canada, Israel, and Italy did not have significant effects on cognitive outcomes. These differences in outcomes among different countries did not reach significance. The age of the sample at intake (based on the 18 groups that reported this information), intervention duration, approximate total hours of intervention (based on the 19 groups that reported this information), and year of publication were not significantly associated with the effect sizes for cognitive outcomes (all p values > 0.4).

Communication. As seen in Table 5, Model and treatment as usual programs had significant positive effects on communication outcomes (Hedges's $g = 0.41$ and 0.31

respectively), while children receiving variable EI did not make significant gains. However these differences among EI categories did not reach significance. Interventions conducted in Australia, Israel, Italy, and United States had significant positive effects on communication scores (Hedges's *g* range from 0.31 to 0.60), while those in Norway and the United Kingdom did not. These differences among outcomes from different countries reached marginal significance ($p = 0.052$). Based on meta-regression results, the age of the sample at intake (based on the 22 groups that reported this information) was not significantly associated with the effect sizes for communication outcomes ($p = 0.51$). Year of intervention was positively associated with the effect size of communication outcomes (*slope* = 0.04, $p = 0.01$). Intervention duration and approximate total intervention hours were both negatively associated with effect sizes for communication outcomes (*slope* = -0.01 and -0.0001 respectively, all *p*s < .05).

Social. As seen in Table 5, the effect sizes among the EI categories differed significantly ($p < .05$). Model programs and treatment as usual programs has significantly positive effects on social outcomes (Hedges's $g = 0.44$ and 0.22 respectively), while variable EI did not ($p = 0.9$). Although interventions conducted in Israel, Italy, and the United States had significantly positive effect sizes, and those conducted in Australia, Norway, and the United Kingdom did not, the differences between these outcomes did not reach significance. Based on meta-regression results, the age of the sample at baseline (based on the 21 groups that reported this information), year of publication, intervention duration, and total approximate intervention hours (based on the 19 studies that reported this information) were not significantly associated with the effect sizes of social outcomes.

Adaptive behavior. As seen in Table 5, the effect sizes among EI categories differed significantly. Model programs had a significantly positive average effect size (Hedges's $g = 0.44$), while treatment as usual programs and variable services had very small effect sizes that did not reach significance. The country that the intervention took place was also a significant moderator. Italy and the United States had significantly positive effect sizes (Hedges's $g = 0.57$ and 0.32, respectively), and Norway had a positive effect size with a marginal level of significance (Hedges's $g = 0.26$). The effect sizes for Australia, Israel, Sweden, and the United Kingdom were not significant. Based on meta-regression results, the age of the sample at intake (based on the 19 groups that reported this information) and publication year were not significantly associated with the effect sizes for adaptive behavior outcomes ($p > .79$). Intervention duration was significantly negatively associated with effect sizes for adaptive behavior outcomes (slope = -0.02, $p = .001$). Total approximate intervention hours (from the 18 groups that reported this information) was also negatively associated with effect sizes for social outcomes and reached marginal significance (slope = -0.0001, $p = 0.07$).

INSERT TABLE 5 HERE

Discussion

We found that the effect sizes associated with community-based intervention for children with autism were small, ranging from 0.21 for adaptive behavior to 0.31 for communication. These stand in stark contrast to those observed in university-based clinical trials, which find effect sizes of 0.4 to 1.2 for these same domains. It should be noted that the effect sizes from these trials represent the difference between the treatment and control groups, instead of the total effect size over time, which makes the difference even greater between university-based clinical trials and community-based interventions.

Despite the low average effect sizes, a number of programs (e.g., Children's Toddler School, Rutgers Autism Program) showed strong outcomes that approached those observed in clinical trials. As these programs were developed within the context of community settings they offer the potential to be replicable and sustainable community programs (Stahmer & Aarons, 2009).

Duration of intervention and total intervention hours were negatively associated with communication and adaptive behavior outcomes, suggesting that more intervention is not necessarily beneficial. These results highlight the potential importance of receiving shorter-term quality interventions over longer ones, and highlight the importance of ongoing monitoring of treatment response, so that intervention targets, strategies, or programs can be adapted or changed if benefit is not observed after a limited duration (National Autism Center, 2015).

Communication results improved over time, but not cognitive, social, or adaptive behavior. This finding suggests that evidence-based practices are not making their way into standard community care. However, this may be a result of restriction of range, as most studies published prior to 2004 did not meet our inclusion criteria. Year of publication may also have been a poor proxy for the year that data were collected.

That programs associated with universities and hospitals had significantly better outcomes than other types community programs suggests that expert or academic involvement may bolster the effectiveness of EI programs. We were limited in our ability to investigate this question further, as we excluded studies that were research-funded replication, dissemination, or implementation studies in community settings because they involved research support and did not reflect current standard care available in community settings. However, these types of studies reflect an important step in studying treatments in “progressively more genuine circumstances” (Chorpita, 2004; Southam-Gerow, Silverman, & Kendall, 2006; Weisz, 2004) and would more directly address the effectiveness of collaborations between academics and community practices and ideal models of training and ongoing support.

A number of other study limitations should be noted. Uncontrolled effect sizes should be interpreted with caution. We were limited in the characteristics of the intervention models and the participants that we could include in our analysis. Next steps include examining the role of other participant characteristics (e.g., baseline IQ, socio-economic status), intervention (e.g., inclusion of parent training, use of manualized intervention) and study characteristics (e.g., method of allocation to intervention). Parent reported outcomes (i.e., the Vineland Adaptive Behavior Scales) may be biased towards programs that include parent training/model programs. We also required standardized scores, which may have resulted in important studies being excluded. Next steps include exploring other outcome measures.

Despite these limitations, these findings hold important implications. Foremost are concerns regarding the difference in outcomes between children enrolled in research trials and children who receive community-based intervention. There are a number of possible reasons for this difference. First, smaller effects community-based interventions could be due to insufficient translation of research into practice. For example, community providers may lack opportunity for high quality training and supervision in the interventions used in research trials. Second, community sites have fewer resources and may be unable to implement complex, resource intensive programs. Third, the difference in outcomes could be due to differences in characteristics of children and families between community settings and research trials. Lord et al. (2005) point out that in treatment studies that report demographic characteristics of participants, the overwhelming majority are white and of relatively high socio-economic status. Families that learn about and enroll in studies may have more resources, fewer obstacles, and more motivation/skill. Finally, unlike research trials, community sites are often required to accept all children and do not have exclusion criteria. Thus, community sites may be more likely to work with more heterogeneous populations within the same program, including differences in functioning, at-home support, family resources, native language, and complex comorbidities.

More work is needed to improve outcomes for children with ASD receiving early intervention in community settings.

References

Anan, R. M., Warner, L. J., McGillivary, J. E., Chong, I. M., & Hines, S. J. (2008). Group Intensive Family Training (GIFT) for preschoolers with autism spectrum disorders. *Behavioral Interventions*, *23*(3), 165–180. https://doi.org/10.1002/bin.262

Baker-Ericzén, M. J., Stahmer, A. C., & Burns, A. (2007). Child Demographics Associated With Outcomes in a Community-Based Pivotal Response Training Program. *Journal of Positive Behavior Interventions*, *9*(1), 52–60. https://doi.org/10.1177/10983007070090010601

Beal, D. J., Corey, D. M., & Dunlap, W. P. (2002). On the bias of Huffcutt and Arthur's (1995) procedure for identifying outliers in the meta-analysis of correlations. *Journal of Applied Psychology*, *87*(3), 583–589. https://doi.org/10.1037//0021-9010.87.3.583

Ben-Itzchak, E., Watson, L. R., & Zachor, D. A. (2014). Cognitive Ability is Associated with Different Outcome Trajectories in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *44*(9), 2221–2229. https://doi.org/10.1007/s10803-014-2091-0

Borenstein, M., Hedges, L., Higgins, J. P. T., & Rothstein, H. (2005). *Comprehensive Meta-analysis Version 2.* Englewood, NJ: Biostat.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470743386.refs

Carter, A. S., Messinger, D. S., Stone, W. L., Celimli, S., Nahmias, A. S., & Yoder, P. (2011). A randomized controlled trial of Hanen's "More Than Words" in toddlers with early autism symptoms. *Journal of Child Psychology and Psychiatry*, *52*(7), 741–752. https://doi.org/10.1111/j.1469-7610.2011.02395.x

Charman, T., Howlin, P., Berry, B., & Prince, E. (2004). Measuring Developmental Progress of Children with Autism Spectrum Disorder on School Entry Using Parent Report. *Autism*, *8*(1), 89–100. https://doi.org/10.1177/1362361304040641

Chorpita, B. F. (2004). Four Considerations for Dissemination of Intervention Innovations. *Clinical Psychology: Science and Practice*, *11*(4), 364–367. https://doi.org/10.1093/clipsy/bph093

Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early Intensive Behavioral Treatment: Replication of the UCLA Model in a Community Setting. *Journal of Developmental & Behavioral Pediatrics*, *27*(2).

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., … Varley, J. (2010). Randomized, Controlled Trial of an Intervention for Toddlers With Autism: The Early Start Denver Model. *Pediatrics*, *125*(1), e17–e23. https://doi.org/10.1542/peds.2009-0958

Duval, S., & Tweedie, R. (2000). Trim and Fill: A Simple Funnel-Plot-Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2009). Meta-Analysis of Early Intensive Behavioral Intervention for Children With

Autism. *Journal of Clinical Child & Adolescent Psychology*, *38*(3), 439–450. https://doi.org/10.1080/15374410902851739

Eldevik, S., Hastings, R. P., Jahr, E., & Hughes, J. C. (2012). Outcomes of Behavioral Intervention for Children with Autism in Mainstream Pre-School Settings. *Journal of Autism and Developmental Disorders*, *42*(2), 210–220. https://doi.org/10.1007/s10803-011-1234-9

Fernell, E., Hedvall, Å., Westerlund, J., Höglund Carlsson, L., Eriksson, M., Barnevik Olsson, M., … Gillberg, C. (2011). Early intervention in 208 Swedish preschoolers with autism spectrum disorder. A prospective naturalistic study. *Research in Developmental Disabilities*, *32*(6), 2092–2101. https://doi.org/10.1016/j.ridd.2011.08.002

Grindle, C. F., Hastings, R. P., Saville, M., Hughes, J. C., Huxley, K., Kovshoff, H., … Remington, B. (2012). Outcomes of a Behavioral Education Model for Children With Autism in a Mainstream School Setting. *Behavior Modification*, *36*(3), 298–319. https://doi.org/10.1177/0145445512441199

Hedges, L. V. (1981). Distribution Theory for Glass's Estimator of Effect size and Related Estimators. *Journal of Educational and Behavioral Statistics*, *6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Hess, K. L., Morrier, M. J., Heflin, L. J., & Ivey, M. L. (2008). Autism Treatment Survey: Services Received by Children with Autism Spectrum Disorders in Public School Classrooms. *Journal of Autism and Developmental Disorders*, *38*(5), 961–971. https://doi.org/10.1007/s10803-007-0470-5

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*(11), 1539–1558. https://doi.org/10.1002/sim.1186

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ : British Medical Journal*, *327*(7414), 557–560.

Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities*, *26*(4), 359–383. https://doi.org/10.1016/j.ridd.2004.09.005

Huffcutt, A. I., & Arthur, W. (1995). Development of a new outlier statistic for meta-analytic data. *Journal of Applied Psychology*, *80*(2), 327–334. https://doi.org/10.1037/0021-9010.80.2.327

Klintwall, L., Macari, S., Eikeseth, S., & Chawarska, K. (2015). Interest level in 2-year-olds with autism spectrum disorder predicts rate of verbal, nonverbal, and adaptive skill acquisition. *Autism*, *19*(8), 925–933. https://doi.org/10.1177/1362361314555376

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., … Moher, D. (2009). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLoS Medicine*, *6*(7), e1000100. https://doi.org/10.1371/journal.pmed.1000100

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, Calif: Sage Publications.

Lord, C., Wagner, A., Rogers, S., Szatmari, P., Aman, M., Charman, T., … Yoder, P. (2005). Challenges in Evaluating Psychosocial Interventions for Autistic Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *35*(6), 695–708. https://doi.org/10.1007/s10803-005-0017-6

Magiati, I., Charman, T., & Howlin, P. (2007). A two-year prospective follow-up study of community-based early intensive behavioural intervention and specialist nursery provision for children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *48*(8), 803–812. https://doi.org/10.1111/j.1469-7610.2007.01756.x

Makrygianni, M. K., & Reed, P. (2010). A meta-analytic review of the effectiveness of behavioural early intervention programs for children with Autistic Spectrum Disorders. *Research in Autism Spectrum Disorders*, *4*(4), 577–593. https://doi.org/10.1016/j.rasd.2010.01.014

Mullen, E. M. (1995). *Mullen Scales of Early Learning manual*. Circle Pines, MN: American Guidance Service.

Muratori, F., & Narzisi, A. (2014). Exploratory study describing 6 month outcomes for young children with autism who receive treatment as usual in Italy. *Neuropsychiatric Disease and Treatment*, 577. https://doi.org/10.2147/NDT.S58308

Murza, K. A., Schwartz, J. B., Hahs-Vaughn, D. L., & Nye, C. (2016). Joint attention interventions for children with autism spectrum disorder: a systematic review and meta-analysis: Joint attention meta-analysis. *International Journal of Language & Communication Disorders*, *51*(3), 236–251. https://doi.org/10.1111/1460-6984.12212

National Autism Center. (2015). *Evidence-based practice and autism in the schools* (2nd ed.). Randolph, MA: Author.

Ospina, M. B., Krebs Seida, J., Clark, B., Karkhaneh, M., Hartling, L., Tjosvold, L., … Smith, V. (2008). Behavioural and Developmental Interventions for Autism Spectrum Disorder: A Clinical Systematic Review. *PLoS ONE*, *3*(11), e3755. https://doi.org/10.1371/journal.pone.0003755

Reed, P., Osborne, L. A., & Corness, M. (2010). Effectiveness of special nursery provision for children with autism spectrum disorders. *Autism*, *14*(1), 67–82. https://doi.org/10.1177/1362361309340030

Reichow, B. (2012). Overview of Meta-Analyses on Early Intensive Behavioral Intervention for Young Children with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *42*(4), 512–520. https://doi.org/10.1007/s10803-011-1218-9

Reichow, B., Barton, E. E., Boyd, B. A., & Hume, K. (2012). Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD). In The Cochrane Collaboration (Ed.), *Cochrane Database of Systematic Reviews*.

Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/14651858.CD009260.pub2

Remington, B., Hastings, R. P., Kovshoff, H., degli Espinosa, F., Jahr, E., Brown, T., … Ward, N. (2007). Early Intensive Behavioral Intervention: Outcomes for Children With Autism and Their Parents After Two Years. *American Journal on Mental Retardation*, *112*(6), 418–438. https://doi.org/10.1352/0895-8017(2007)112[418:EIBIOF]2.0.CO;2

Rickards, A. L., Walstab, J. E., Wright-Rossi, R. A., Simpson, J., & Reddihough, D. S. (2007). A Randomized, Controlled Trial of a Home-Based Intervention Program for Children with Autism and Developmental Delay. *Journal of Developmental & Behavioral Pediatrics*, *28*(4).

Rivard, M., Terroux, A., & Mercier, C. (2014). Effectiveness of early behavioral intervention in public and mainstream settings: The case of preschool-age children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *8*(9), 1031–1043. https://doi.org/10.1016/j.rasd.2014.05.010

Roberts, J., Williams, K., Carter, M., Evans, D., Parmenter, T., Silove, N., … Warren, A. (2011). A randomised controlled trial of two early intervention programs for young children with autism: Centre-based with parent program and home-based. *Research in Autism Spectrum Disorders*, *5*(4), 1553–1566. https://doi.org/10.1016/j.rasd.2011.03.001

Rogers, S. J., Estes, A., Lord, C., Vismara, L., Winter, J., Fitzpatrick, A., … Dawson, G. (2012). Effects of a Brief Early Start Denver Model (ESDM)–Based Parent

Intervention on Toddlers at Risk for Autism Spectrum Disorders: A Randomized Controlled Trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(10), 1052–1065. https://doi.org/10.1016/j.jaac.2012.08.003

Roid, G. H., & Miller, L. (1997). Leiter international performance scale-revised: Examiner's manual. In G. H. Roid & L. J. Miller (Eds.), *Leiter international performance scale-revised*. Wood Dale, IL: Stoelting.

Roid, G. H., & Sampers, J. L. (2004). *Merrill-Palmer-Revised : scales of development*. Wood Dale (IL): Stoelting.

Salt, J., Shemilt, J., Sellars, V., Boyd, S., Coulson, T., & Mc Cool, S. (2002). The Scottish Centre for Autism Preschool Treatment Programme: II: The Results of a Controlled Treatment Outcome Study. *Autism*, *6*(1), 33–46. https://doi.org/10.1177/1362361302006001004

Schertz, H. H., Odom, S. L., Baggett, K. M., & Sideris, J. H. (2013). Effects of Joint Attention Mediated Learning for toddlers with autism spectrum disorders: An initial randomized controlled study. *Early Childhood Research Quarterly*, *28*(2), 249–258. https://doi.org/10.1016/j.ecresq.2012.06.006

Schreibman, L., Dawson, G., Stahmer, A. C., Landa, R., Rogers, S. J., McGee, G. G., … Halladay, A. (2015). Naturalistic Developmental Behavioral Interventions: Empirically Validated Treatments for Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *45*(8), 2411–2428. https://doi.org/10.1007/s10803-015-2407-8

Shire, S. Y., Chang, Y.-C., Shih, W., Bracaglia, S., Kodjoe, M., & Kasari, C. (2017). Hybrid implementation model of community-partnered early intervention for toddlers with autism: a randomized trial. *Journal of Child Psychology and Psychiatry*, *58*(5), 612–622. https://doi.org/10.1111/jcpp.12672

Shire, S. Y., & Kasari, C. (2014). Train the Trainer Effectiveness Trials of Behavioral Intervention for Individuals With Autism: A Systematic Review. *American Journal on Intellectual and Developmental Disabilities*, *119*(5), 436–451. https://doi.org/10.1352/1944-7558-119.5.436

Southam-Gerow, M. A., Silverman, W. K., & Kendall, P. C. (2006). Client Similarities and Differences in Two Childhood Anxiety Disorders Research Clinics. *Journal of Clinical Child & Adolescent Psychology*, *35*(4), 528–538. https://doi.org/10.1207/s15374424jccp3504_4

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1985). *Vineland : Adaptive behavior scales : Interview edition : Survey form manual*. Circle Pines, Minnesota: American Guidance Service.

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland-II: Vineland Adaptive Behavior Scales, survey forms manual*. Circle Pines, MN: American Guidance Service.

Stahmer, A. C., & Aarons, G. A. (2009). Attitudes toward adoption of evidence-based practices: A comparison of autism early intervention providers and children's mental health providers. *Psychological Services*, *6*(3), 223–234. https://doi.org/10.1037/a0010738

Stahmer, A. C., Akshoomoff, N., & Cunningham, A. B. (2011). Inclusion for toddlers with autism spectrum disorders: The first ten years of a community program. *Autism*, *15*(5), 625–641. https://doi.org/10.1177/1362361310392253

Stahmer, A. C., Collings, N. M., & Palinkas, L. A. (2005). Early intervention practices for children with autism: Descriptions from community providers. *Focus on Autism and Other Developmental Disabilities*, *20*(2), 66–79.

Strauss, K., Mancini, F., & Fava, L. (2013). Parent inclusion in early intensive behavior interventions for young children with ASD: A synthesis of meta-analyses from 2009 to 2011. *Research in Developmental Disabilities*, *34*(9), 2967–2985. https://doi.org/10.1016/j.ridd.2013.06.007

Strauss, K., Vicari, S., Valeri, G., D'Elia, L., Arima, S., & Fava, L. (2012). Parent inclusion in Early Intensive Behavioral Intervention: The influence of parental stress, parent treatment fidelity and parent-mediated generalization of behavior targets on child outcomes. *Research in Developmental Disabilities*, *33*(2), 688–703. https://doi.org/10.1016/j.ridd.2011.11.008

Tonge, B., Brereton, A., Kiomall, M., Mackinnon, A., & Rinehart, N. J. (2014). A randomised group comparison controlled trial of "preschoolers with autism": A parent education and skills training intervention for young children with autistic disorder. *Autism*, *18*(2), 166–177. https://doi.org/10.1177/1362361312458186

Turner-Brown, L., Hume, K., Boyd, B. A., & Kainz, K. (2016). Preliminary Efficacy of Family Implemented TEACCH for Toddlers: Effects on Parents and Their

Toddlers with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-016-2812-7

Virués-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood: Meta-analysis, meta-regression and dose–response meta-analysis of multiple outcomes. *Clinical Psychology Review*, *30*(4), 387–399. https://doi.org/10.1016/j.cpr.2010.01.008

Vivanti, G., Paynter, J., Duncan, E., Fothergill, H., Dissanayake, C., Rogers, S. J., & the Victorian ASELCC Team. (2014). Effectiveness and Feasibility of the Early Start Denver Model Implemented in a Group-Based Community Childcare Setting. *Journal of Autism and Developmental Disorders*, *44*(12), 3140–3153. https://doi.org/10.1007/s10803-014-2168-9

Weiss, M. J. (1999). Differential rates of skill acquisition and outcomes of early intensive behavioral intervention for autism. *Behavioral Interventions*, *14*(1), 3–22. https://doi.org/10.1002/(SICI)1099-078X(199901/03)14:1<3::AID-BIN25>3.0.CO;2-F

Weisz, J. R. (2004). Treatment Dissemination and Evidence-Based Practice: Strengthening Intervention Through Clinician-Researcher Collaboration. *Clinical Psychology: Science and Practice*, *11*(3), 300–307. https://doi.org/10.1093/clipsy/bph085

Weitlauf, A. S., McPheeters, M. L., Peters, B., Sathe, N., Travis, R., Aiello, R., … Warren, Z. (2014). *Therapies for Children With Autism Spectrum Disorder*.

Agency for Healthcare Research and Quality (US). Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK241444/

What Works Clearinghouse. (2014, March). Procedures and Standards Handbook Version 3.0. Retrieved July 5, 2017, from https://ies.ed.gov/ncee/wwc/Handbooks

Zachor, D. A., & Ben-Itzchak, E. (2010). Treatment approach, autism severity and intervention outcomes in young children. *Research in Autism Spectrum Disorders*, *4*(3), 425–432. https://doi.org/10.1016/j.rasd.2009.10.013

Tables

Table 1

Sample Characteristics

Study Reference	Group	% Male	% Minority	*M* age in mo	*M* (*SD*) BL Cog	Analysis n			
						Cog	Com	Social	AB
(Anan et al., 2008)	GIFT	85	NR	44.0	51.7 (6.3)	72		72	72
(Baker-Ericzén et al., 2007)	3 year olds	FS: 83	FS: 73	NR	NR		90	90	86
(Baker-Ericzén et al., 2007)	4-5 year olds	FS: 83	FS: 73	NR	NR		55	55	52
(Ben-Itzchak et al., 2014)	ABA	85	NR	25.5	71.4 (20.2)	33			36
(Carter et al., 2011)	BAU	FS: 82	FS: 53	21.5	NR		24	24	
(Charman, Howlin, Berry, & Prince, 2004)	Cohort 1	NR	NR	56.6	NR				57
(Cohen et al., 2006)	Comparison	81	NR	NR	59.4 (14.7)	21	21	21	21
(Dawson et al., 2010)	A/M	FS: 78	FS: 27	23.1	59.4 (8.6)	24	24	24	
(Eldevik, Hastings, Jahr, & Hughes, 2012)	EIBI	81	NR	42.2	51.6 (16.9)	31	31	31	31
(Eldevik et al., 2012)	TAU	67	NR	46.2	51.7 (18.1)	12	12	12	12
(Fernell et al., 2011)	Intensive ABA	NR	NR	NR	NR				91
(Fernell et al., 2011)	Non-intensive ABA	NR	NR	NR	NR				101
(Grindle et al., 2012)	ABA class	82	NR	58.2	59.5 (13.2)		11	11	11
(Howard et al., 2005)	IBT	86	28	30.9	58.5 (18.2)	28			
(Howard et al., 2005)	AP	81	50	37.4	53.7 (13.5)	16			
(Howard et al., 2005)	GP	100	43	34.6	59.9 (14.9)	16			

Study Reference	Group	% Male	% Minority	*M* age in mo	*M* (*SD*) BL Cog	Analysis n Cog	Com	Social	AB
(Klintwall et al., 2015)	Community services	89	26	21.9	V: 46.7 (26.7), NV: 76.8 (18.7)				70
(Magiati et al., 2007)	EIBI	96	25	38.0	83.0 (27.9)		26	26	26
(Magiati et al., 2007)	Nursery	75	31	42.5	65.2 (26.9)		14	14	14
(Muratori & Narzisi, 2014)	TAU	81	NR	35.2	62.3 (13.4)	70	70	70	70
(Reed et al., 2010)	General special	92	NR	44.3	47.7 (22.3)	12			12
(Reed et al., 2010)	PACTS	92	NR	40.8	49.3 (13.2)	13			13
(Remington et al., 2007)	EIBI	NR	NR	35.7	61.4 (16.4)	23			
(Remington et al., 2007)	Comparison	NR	NR	38.4	62.3 (16.6)	21			
(Rickards et al., 2007)	Center-based	FS: 79	NR	FS: 43.1	55.7 (22.1)	21			
(Rivard, Terroux, & Mercier, 2014)	EBI	FS: 75	NR	FS: 46.0	60.1 (16.4)	85			
(Roberts et al., 2011)	CB	FS: 91	NR	43.7	66.5 (17.7)		29	29	
(Roberts et al., 2011)	HB	FS: 91	NR	41.5	57.0 (11.7)		28	28	
(Roberts et al., 2011)	Wait-list	FS: 91	NR	43.7	63.3 (15.5)		28	28	
(Rogers et al., 2012)	Community services	63	25	20.9	63.1 (15.9)	49	49	49	49
(Salt et al., 2002)	SCA	92	NR	42.4	39.4 (13.5)		12	12	12
(Schertz et al., 2013)	Control	NR	NR	27.5	NR		12		
(Stahmer et al., 2011)	CTS	85	38	28.1	63.9 (13.3)	100	98	98	98
(Strauss et al., 2012)	Eclectic	95	NR	41.9	74.3 (29.4)	15	20	20	20
(Tonge et al., 2014)	Control	91	NR	50.1	63.3 (28.5)	35	35	35	
(Turner-Brown et al., 2016)	Services as Usual	94	29	29.7	61.8 (17.8)	17			
(Vivanti et al., 2014)	Control	90	NR	42.0	49.0 (17.4)	30	30	30	30

Study Reference	Group	% Male	% Minority	*M* age in mo	*M* (*SD*) BL Cog	Analysis n			
						Cog	Com	Social	AB
(Weiss, 1999)	Rutgers Autism	NR	NR	41.5	NR				20
(Zachor & Ben-Itzchak, 2010)	ABA	FS: 91	NR	25.1	72.2 (19.2)		45	45	
(Zachor & Ben-Itzchak, 2010)	Eclectic	FS: 91	NR	26.0	73.3 (22.2)		33	33	

Note: AB= Adaptive Behavior outcome, ABA = Applied Behavior Analysis group , BL= Baseline, Cog= Cognitive outcome, Com= Communication outcome, FS= Full study sample, M= Mean, mo= months, NA= Not applicable, NR= Not reported, NV= Nonverbal Developmental Quotient, SD= Standard Deviation, TAU= Treatment as usual group V = Verbal Developmental Quotient

Table 2

Intervention Characteristics

Study Reference	Group	CO	Years of EI	EI duration (mo)	EI intensity (hr/wk)	EI category	EI description
(Anan et al., 2008)	GIFT	US	NR	3.0	15	Model	Group Intensive Family Training Program, parent training model
(Baker-Ericzén et al., 2007)	3 year olds	US	1999-2003	3.0	1	Model	PRT
(Baker-Ericzén et al., 2007)	4 -5 year olds	US	1999-2003	3.0	1	Model	PRT
(Ben-Itzchak et al., 2014)	ABA	IS	NR	12.0	20	TAU	Center-based ABA
(Carter et al., 2011)	BAU	US	NR	9	NR	VAR	Business as usual control group
(Charman et al., 2004)	Cohort 1	UK	2000-2001	11.1	NR	TAU	dedicated autism primary schools or specialist units
(Cohen et al., 2006)	Comparison	US	1995-2003	36.0	NR	TAU	services from local public schools
(Dawson et al., 2010)	A/M	US	NR	29.3	18.4	VAR	Assess and monitor group
(Eldevik et al., 2012)	EIBI	NO	2000-2011	25.1	preschool = 20, EIBI = 13.6	Model	EIBI
(Eldevik et al., 2012)	TAU	NO	2000-2011	24.6	preschool = 20, TAU ≥ 5	TAU	Eclectic
(Fernell et al., 2011)	Intensive ABA	SW	NR	20.9	15-40	TAU	ABA

Study Reference	Group	CO	Years of EI	EI duration (mo)	EI intensity (hr/wk)	EI category	EI description
(Fernell et al., 2011)	Non-intensive ABA	SW	NR	20.9	<15	TAU	ABA
(Grindle et al., 2012)	ABA class	UK	2005-2009	12.0	ABA= 15, max school= 30	Model	The Westwood ABA Class
(Howard et al., 2005)	IBT	US	1996-2003	14.2	< 3y: 25-30, > 3y 35-40	Model	Intensive behavior analytic treatment
(Howard et al., 2005)	AP	US	1996-2003	13.3	25-30	TAU	Autism educational programming, eclectic approaches
(Howard et al., 2005)	GP	US	1996-2003	14.8	15	TAU	Generic educational programming serving children with a variety of disabilities
(Klintwall et al., 2015)	Community	US	NR	16.3	13.9	VAR	Community based treatment, variety of interventions
(Magiati et al., 2007)	EIBI	UK	1998-2002	25.5	T1: 32.4, T2: 33.2	TAU	EIBI
(Magiati et al., 2007)	Nursery	UK	1998-2002	26.0	T1: 25.6, T2: 27.4	TAU	Nursey programs utilizing eclectic intervention practices
(Muratori & Narzisi, 2014)	TAU	IT	NR	6.0	11.2	TAU	centers with specific treatments performed by child neuropsychiatric services and school inclusion with individual support teacher

Study Reference	Group	CO	Years of EI	EI duration (mo)	EI intensity (hr/wk)	EI category	EI description
(Reed et al., 2010)	General special	UK	NR	9.0	11.5	TAU	nursey schools that catered for children with all special needs, including autism, interventions used were eclectic
(Reed et al., 2010)	PACTS	UK	NR	9.0	12.6	TAU	Parents of Autistic Children Training and Support (PACTS), developed by Bexley Local Education Authority, home-based + parent-training
(Remington et al., 2007)	EIBI	UK	NR	24.0	25.6	TAU	home-based EIBI
(Remington et al., 2007)	Comparison	UK	NR	24	NR	VAR	local education authorities' standard provision for young children with autism, variety of interventions
(Rickards et al., 2007)	Center-based	AU	2000-2003	13.0	5	TAU	center-based programs utilizing eclectic intervention practices
(Rivard et al., 2014)	EBI	CA	2009-2012	12.0	16-20	TAU	Early behavioral intervention at a rehabilitation center providing developmental services to persons with intellectual disabilities and ASD, ABA program
(Roberts et al., 2011)	Wait-list	AU	2006-2008	12	NR	VAR	Wait-list
(Roberts et al., 2011)	HB	AU	2006-2008	12.0	2-3/fortnight	TAU	Building Blocks home-based program

Study Reference	Group	CO	Years of EI	EI duration (mo)	EI intensity (hr/wk)	EI category	EI description
(Roberts et al., 2011)	CB	AU	2006-2008	12.0	2-3	TAU	Building Blocks center-based program
(Rogers et al., 2012)	Community	US	NR	3.0	3.68	VAR	Community-based services
(Salt et al., 2002)	SCA	UK	NR	10.0	SCA: 4, total 30.38h non-SCA	Model	Scottish Centre for Autism: comprehensive treatment program, includes 1:1 intensive treatment and parent training, designed to complement child's nursery placement
(Schertz et al., 2013)	Control	US	NR	7.0	Indiana: 12.82, Kansas: 21.35, NC: 6.25	VAR	services commonly available in the community
(Stahmer et al., 2011)	CTS	US	1998-2008	8.3	21	Model	Children's Toddler School
(Strauss et al., 2012)	Eclectic	IT	NR	6.0	12	TAU	eclectic intervention group: in-home developmental intervention and cognitive behavioral treatment
(Tonge et al., 2014)	Control	AU	NR	12.0	7.9	TAU	business as usual control group: weekly intervention comprised attendance at the local preschool, a child-focused early intervention therapy group, and individual speech and/or occupational therapy

Study Reference	Group	CO	Years of EI	EI duration (mo)	EI intensity (hr/wk)	EI category	EI description
(Turner-Brown et al., 2016)	Services as Usual	US	2012-2014	6.5	8.6	VAR	Services as usual: variety of community interventions, including speech, occupational, developmental, and behavioral therapy
(Vivanti et al., 2014)	Control	AU	NR	12.0	≥ 15	TAU	‘‘generic’’ intervention program for ASD: does not subscribe to a single method, philosophy, or theoretical approach, but instead aims to be comprehensive and offer a range of teaching strategies derived from best practice guidelines.
(Weiss, 1999)	Rutgers Autism Program	US	NR	24	NR	Model	Rutgers Autism Program at the Center for Applied Psychology, ABA
(Zachor & Ben-Itzchak, 2010)	Eclectic	IS	NR	12.0	19	TAU	eclectic approach: integrated developmental, DIR, TEACCH, speech, OT
(Zachor & Ben-Itzchak, 2010)	ABA	IS	NR	12.0	20	TAU	ABA

Note: ABA = Applied Behavior Analysis, AU = Australia, CA = Canada, CO = Country, EI: Early intervention, EIBI = Early Intensive Behavioral Intervention, UK = United Kingdom, IT = Italy, IS = Israel, NC = North Carolina, NO = Norway, NR = Not reported, PRT = Pivotal Response Training, SW = Sweden, TAU: Treatment as usual, US = United States, VAR = variable services.

Table 3

Random Weighted Uncontrolled Effect Sizes and SAMDs

Study	Group	Cognitive		Communication		Social		Adaptive Behavior	
		Hedges's *g*	SAMD	Hedges's *g*	SAMD	Hedges's *g*	SAMD	Hedges's *g*	SAMD
(Anan et al., 2008)	GIFT	0.44***	0.35			0.65***	1.80	0.56***	1.67
(Baker-Ericzén et al., 2007)	3 year olds			0.33***	0.08	0.41***	0.92	0.43***	1.23
(Baker-Ericzén et al., 2007)	4-5 year olds			0.31***	0.00	0.32***	0.38	0.38***	0.80
(Ben-Itzchak et al., 2014)	ABA	-0.03	-1.08					0.03	-0.39
(Carter et al., 2011)	BAU			0.70***	0.93	-0.25**	-1.09		
(Charman et al., 2004)	Cohort 1							0.06	-0.36
(Cohen et al., 2006)	Comparison	0.68***	0.73	0.10	-0.48	0.01	-0.45	-0.25	-0.89
(Dawson et al., 2010)	A/M	0.44**	0.22	-0.01	-0.77	-0.96***	**-2.78**	-1.25***	**-3.34**
(Eldevik et al., 2012)	EIBI	0.02***	-0.91	0.46***	0.40	0.49***	0.76	0.43***	0.75
(Eldevik et al., 2012)	TAU	0.54	0.31	0.00	-0.50	-0.24	-0.73	0.05	-0.17
(Fernell et al., 2011)	Intensive ABA							0.01	-0.65
(Fernell et al., 2011)	Non-intensive ABA							0.11***	-0.23
(Grindle et al., 2012)	ABA class			0.67**	0.54	0.00	-0.32	0.48	0.49
(Howard et al., 2005)	AP	0.38***	0.06						
(Howard et al., 2005)	GP	0.56***	0.40						
(Howard et al., 2005)	IBT	1.50***	**2.99**						
(Klintwall et al., 2015)	Community							0.10	-0.25
(Magiati et al., 2007)	EIBI			-0.26*	-1.41	0.14	-0.17	-0.21	-0.90
(Magiati et al., 2007)	Nursery			0.06	-0.44	0.01	-0.36	-0.58*	-1.29
(Muratori & Narzisi, 2014)	TAU	0.38***	0.12	0.62***	1.23	0.43***	0.88	0.54***	1.56
(Reed et al., 2010)	General special	0.26*	0.00					0.30	0.22

Study	Group	Cognitive		Communication		Social		Adaptive Behavior	
		Hedges's *g*	SAMD	Hedges's *g*	SAMD	Hedges's *g*	SAMD	Hedges's *g*	SAMD
(Reed et al., 2010)	PACTS	0.35+	0.00					-0.07	-0.38
(Remington et al., 2007)	EIBI	0.37***	0.04						
(Remington et al., 2007)	TAU	-0.07	-0.93						
(Rickards et al., 2007)	Center-based	-0.23*	-1.30						
(Rivard et al., 2014)	EBI	0.32***	-0.14						
(Roberts et al., 2011)	CB			0.45***	0.37	0.19*	-0.04		
(Roberts et al., 2011)	HB			0.24***	-0.19	-0.30***	-1.31		
(Roberts et al., 2011)	Wait-list			0.33***	0.05	0.21*	0.00		
(Rogers et al., 2012)	Community	0.28**	-0.26	0.42***	0.39	0.06	-0.50	0.16***	0.03
(Salt et al., 2002)	SCA			-0.24	-0.87	0.75***	0.86	-0.02	-0.29
(Schertz et al., 2013)	Control			-0.05	-0.58				
(Stahmer et al., 2011)	CTS	0.67***	1.55	0.54***	1.05	0.45***	1.13	0.35***	0.90
(Strauss et al., 2012)	Eclectic	0.06	-0.55	0.56***	0.53	0.58***	0.79	0.62***	0.99
(Tonge et al., 2014)	Control	0.15+	-0.59	0.26*	-0.15	0.54***	0.94		
(Turner-Brown et al., 2016)	SAU	0.37*	0.04						
(Vivanti et al., 2014)	Control	0.33**	-0.05	0.37***	0.15	0.26**	0.12	0.19***	0.09
(Weiss, 1999)	Rutgers Autism							0.95***	1.70
(Zachor & Ben-Itzchak, 2010)	ABA			0.39***	0.15	0.39	0.58		
(Zachor & Ben-Itzchak, 2010)	Eclectic			0.58***	0.25	0.58	1.05		

Note: Bolded text indicates outlier excluded from subsequent analyses.

+ $p < .08$, * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4

Random Effects Models

Outcome		*n*	Hedges's *g*	95% CI	*Q*(*df*)	I^2
Cognitive	Total (all studies)	22	0.35***	0.23-0.48	203.09(21)***	89.66
	Total (outlier excluded)	21	0.30***	0.20-0.40	120.37(20)***	83.38
	Total (trim-and-fill correction)		0.24	0.13-0.35	194.63	
Communication	Total (all studies)	23	0.31***	0.22-0.41	149.27(22)***	85.26
	Total (outlier excluded)	NA				
	Total (trim-and-fill correction)	NA				
Social	Total (all studies)	23	0.21**	0.08-0.34	235.23(22)***	90.65
	Total (outlier excluded)	22	0.26***	0.14-0.37	177.19(21)***	88.15
	Total (trim-and-fill correction)	NA				
Adaptive Behavior	Total (all studies)	24	0.16**	0.05 - 0.26	314.40(23)***	92.68
	Total (outlier excluded)	23	0.21***	0.11 - 0.30	249.42(22)***	91.18
	Total (trim-and-fill correction)	NA				

Note: CI = confidence interval; *n* = studies included, NA = Not applicable.

* $p < .05$, ** $p <.01$, *** $p < .001$.

Table 5

Analyses of Moderation for Uncontrolled Effect Sizes

Outcome	Moderator	k	Hedges's *g*	95% CI	*Q(df)*	*p*
Cognitive	EI category				5.62(2)	0.060+
	Model	4	0.51***	0.32 - 0.71		
	TAU	13	0.25***	0.13 - 0.36		
	Variable EI	4	0.24*	0.02 - 0.45		
	Country				11.17(6)	0.083
	AU	3	0.08	-0.17 - 0.33		
	CA	1	0.32	-0.07 - 0.71		
	IS	1	-0.03	-0.47 - 0.41		
	IT	2	0.23	-0.06 - 0.52		
	NO	2	0.30+	0.00 - 0.60		
	UK	4	0.22*	0.00 - 0.45		
	US	8	0.48***	0.33 - 0.64		
Communication	EI category				3.03(2)	0.219
	Model	7	0.41***	0.24 - 0.59		
	TAU	11	0.31***	0.17 - 0.45		
	Variable EI	5	0.18	-0.03 - 0.38		
	Country				10.98(5)	0.052+
	AU	5	0.33***	0.16 - 0.50		
	IS	2	0.48**	0.19 - 0.78		
	IT	2	0.60***	0.31 - 0.89		
	NO	2	0.26	-0.05 - 0.56		
	UK	4	0.01	-0.23 - 0.26		
	US	8	0.31***	0.17 - 0.46		
Social	EI category				7.37(2)	0.025*
	Model	7	0.44***	0.26 - 0.63		
	TAU	12	0.22**	0.08 - 0.36		
	Variable EI	3	0.01	-0.25 - 0.28		
	Country				3.21(5)	0.668
	AU	5	0.18	-0.08 - 0.43		

Outcome	Moderator	k	Hedges's *g*	95% CI	*Q(df)*	*p*
	IS	2	0.48*	0.07 - 0.89		
	IT	2	0.49*	0.08 - 0.90		
	NO	2	0.15	-0.27 - 0.57		
	UK	4	0.21	-0.11 - 0.53		
	US	7	0.24*	0.03 - 0.46		
Adaptive Behavior	EI category				12.70(2)	0.002**
	Model	8	0.44***	0.28 – 0.60		
	TAU	13	0.08	-0.04 – 0.20		
	Variable EI	2	0.13	-0.15 – 0.41		
	Country				17.71(6)	0.007**
	AU	1	0.19	0.08 – 0.32		
	IS	1	0.03	-0.07 - 0.11		
	IT	2	0.57***	0.45 – 0.60		
	NO	2	0.26+	-0.10 – 0.51		
	SW	2	0.06	0.01 – 0.11		
	UK	7	-0.02	-0.15 – 0.18		
	US	8	0.32***	0.14 – 0.38		

Note: AU = Australia, CA = Canada, CO = Country, EI: Early intervention, UK = United Kingdom, IT = Italy, IS = Israel, NO = Norway, SW = Sweden, TAU: Treatment as usual, US = United States.

+ p <.08, * p < .05, ** p < .01, *** p < .001.

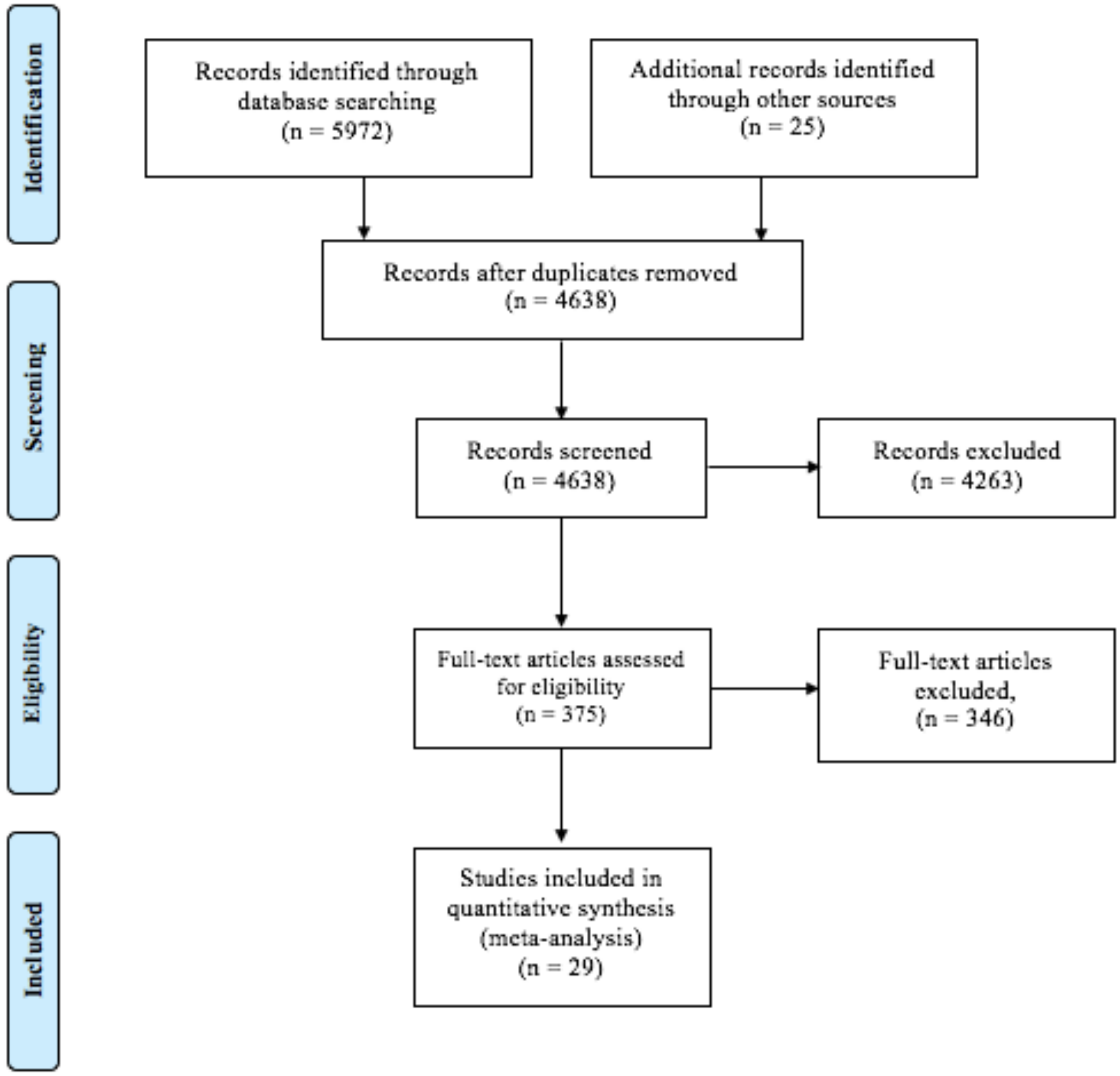

Figure 1. PRISMA Flow Diagram

Appendix A

PsycINFO Search Strategy

((TI,AB(infant OR infancy OR toddler OR toddlers OR "young children" OR "early intervention" OR preschool* OR pre-schooler) OR SU(Infancy OR Preschool OR "early childhood education" OR "early intervention" OR "young children" OR toddlers OR "autistic young children" OR "infants and children")) AND (SU("treatment" OR "behavior modification" OR "behavior therapy" OR "contingency management" OR "token economy programs" OR "classroom behavior management" OR "fading conditioning" OR "omission conditioning" OR "omission training" OR "overcorrection" OR "bibliotherapy" OR "milieu therapy" OR "mulitmodal treatment approach" OR "early intervention") OR (TI,AB(mediated OR implemented) NEAR/3 (TI,AB(parent* OR caregiver* OR maternal* OR paternal* OR mother* OR father*)) AND (TI,AB(intervention OR treatment OR training OR program OR therapy))) OR TI,AB("behavio*r modification" OR "behavio*r* analysis" OR reinforcement OR prompting OR "time delay" OR "functional communication" OR "picture exchange communication system" OR "PECS" OR extinction OR "task analysis" OR "work system" OR "structured teaching" OR "environment* modification*" OR "natural language paradigm" OR "visual supports" OR "response interruption" OR "redirection" OR "Denver Model" OR "TEACCH" OR "ABA" OR "DTT" OR "PRT" OR "SCERTS" OR "Social Communication Emotional Regulation Transactional Support" OR "verbal behavio*r" OR "CABAS" OR Hanen OR "More than words" OR "floortime" OR "floor-time" OR "RDI" OR "DIR" OR "developmental individual difference relationship-based") OR (TI,AB(intervention OR treatment OR program OR programme OR programs OR programmes OR training OR teaching OR therapy OR learning OR instruction) NEAR/3 (ti,ab(early OR individual OR intensive OR incidental OR reciprocal OR development* OR behavio*r* OR parent* OR caregiver* OR care-giver* OR mother* OR father* OR family OR families OR maternal* OR paternal* OR effectiveness OR efficacy OR milieu OR home OR clinic OR naturalistic OR antecedent OR "discrete trial" OR "pivotal response" OR "joint attention" OR "play" OR "communication" OR outcome)))) AND TI,AB(autis* OR "ASD" OR "ASDs" OR "PDD" OR "PDDs" OR "PDD-NOS" OR "pervasive development* disorder*"))

Appendix B

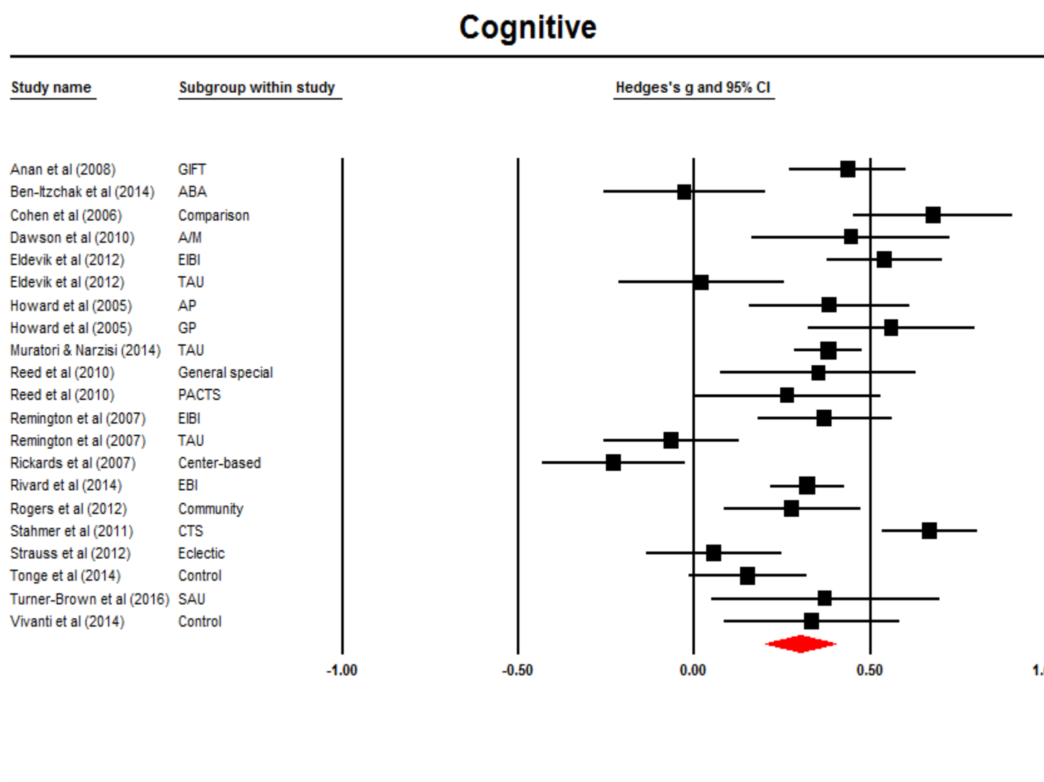

Figure 1B. Forest plot of uncontrolled random effects sizes and 95% confidence intervals for cognitive results. The red diamond indicates the overall effect size.

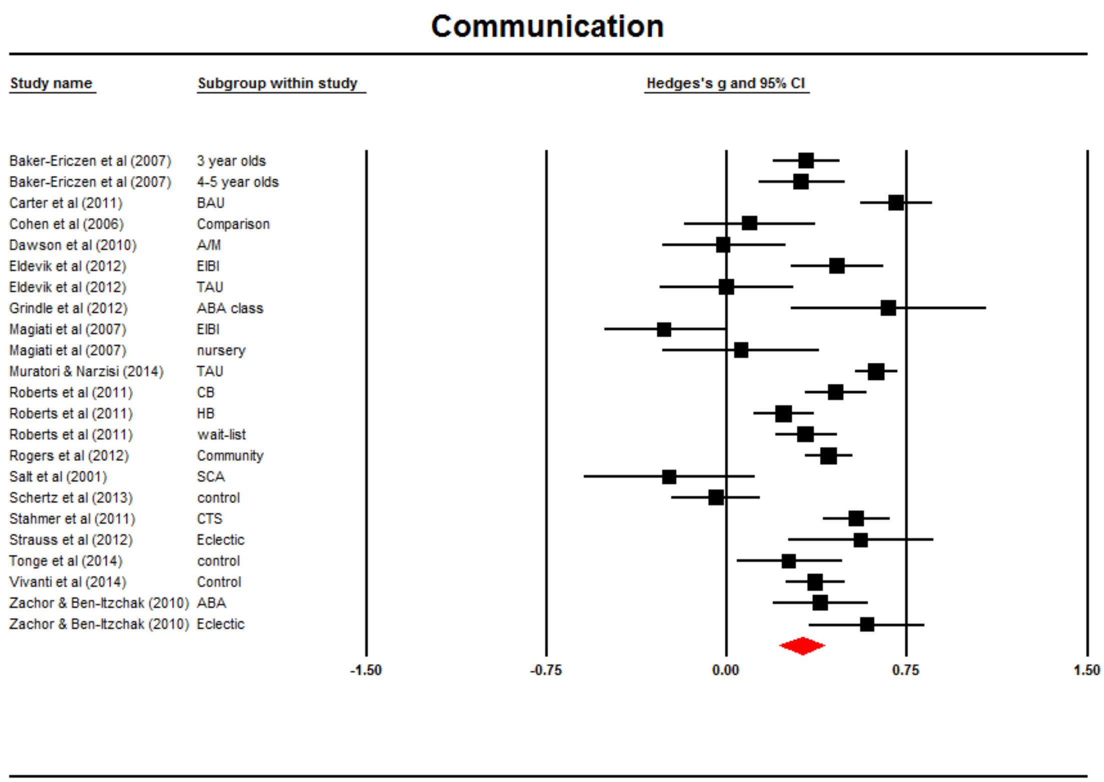

Figure 2B. Forest plot of uncontrolled random effects sizes and 95% confidence intervals for communication results. The red diamond indicates the overall effect size.

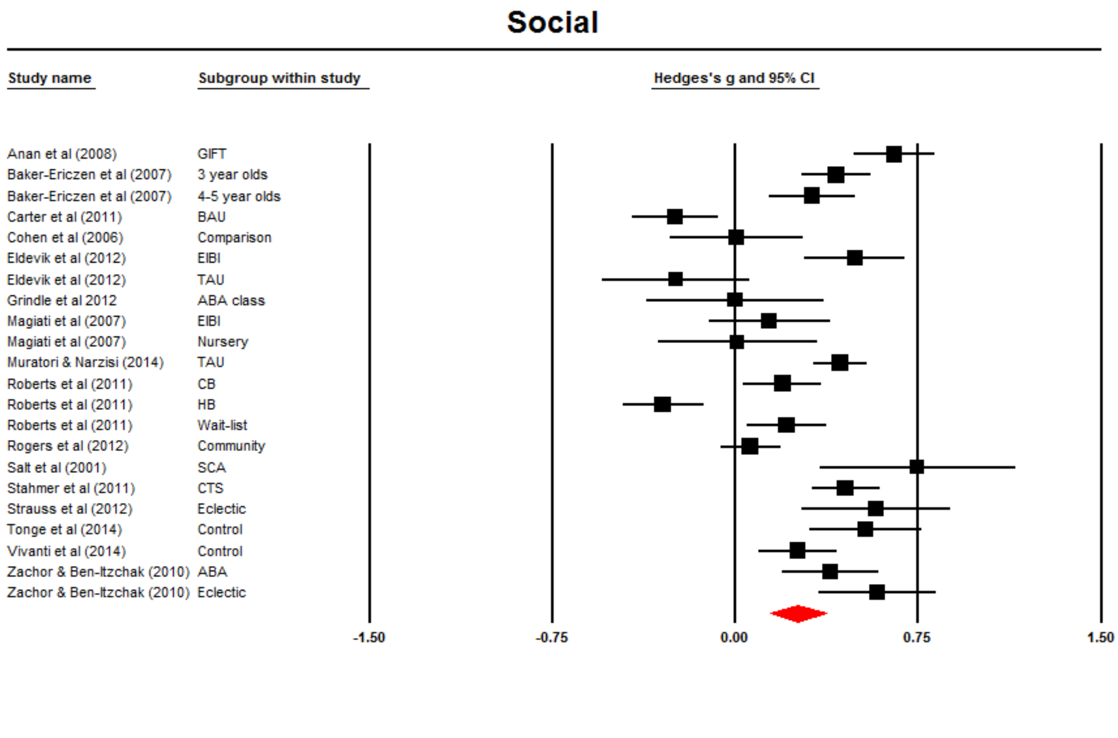

Figure 3B. Forest plot of uncontrolled random effects sizes and 95% confidence intervals for social results. The red diamond indicates the overall effect size.

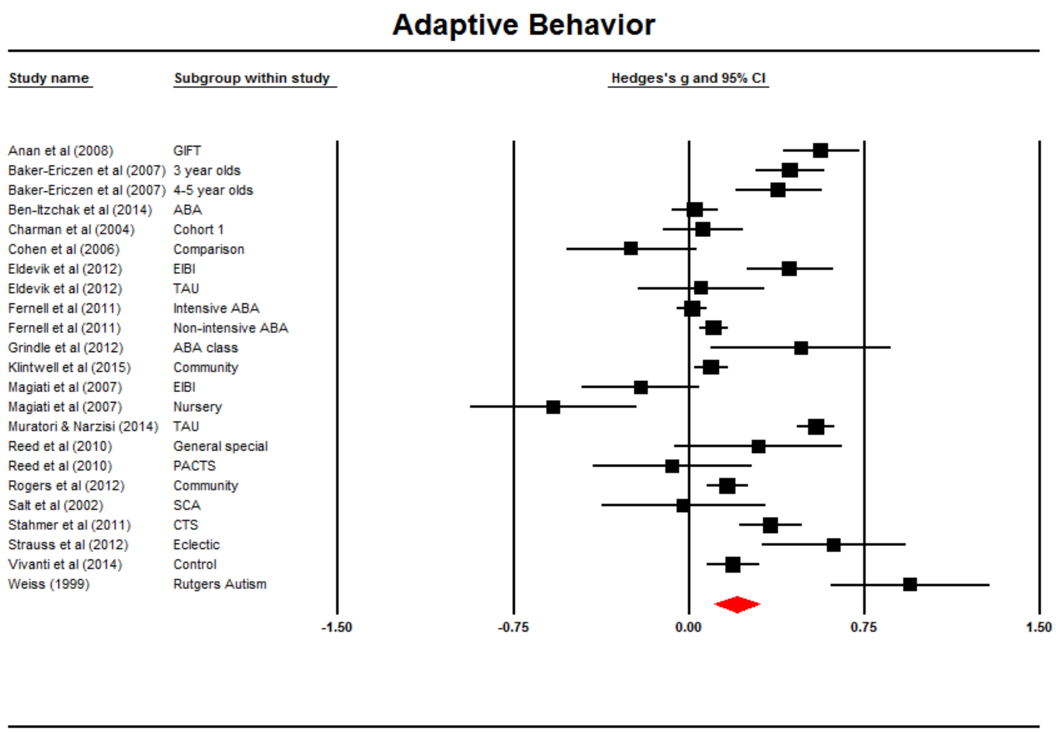

Figure 4B. Forest plot of uncontrolled random effects sizes and 95% confidence intervals for adaptive behavior results. The red diamond indicates the overall effect size.

Appendix C

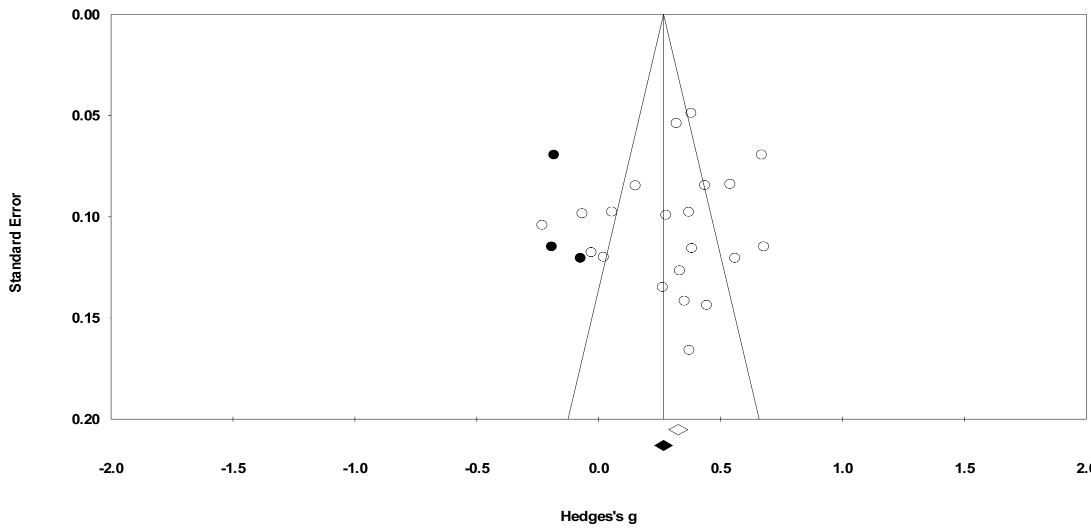

Figure 1C. Funnel plot of standard error by Hedges's *g* for cognitive results. White circles indicate observed groups, the white diamond indicates the overall random weight effect size for the observed groups. Black circles indicate missing studies suggested by trim-and-fill procedures, the black diamond indicates the corrected average effect size including these studies.

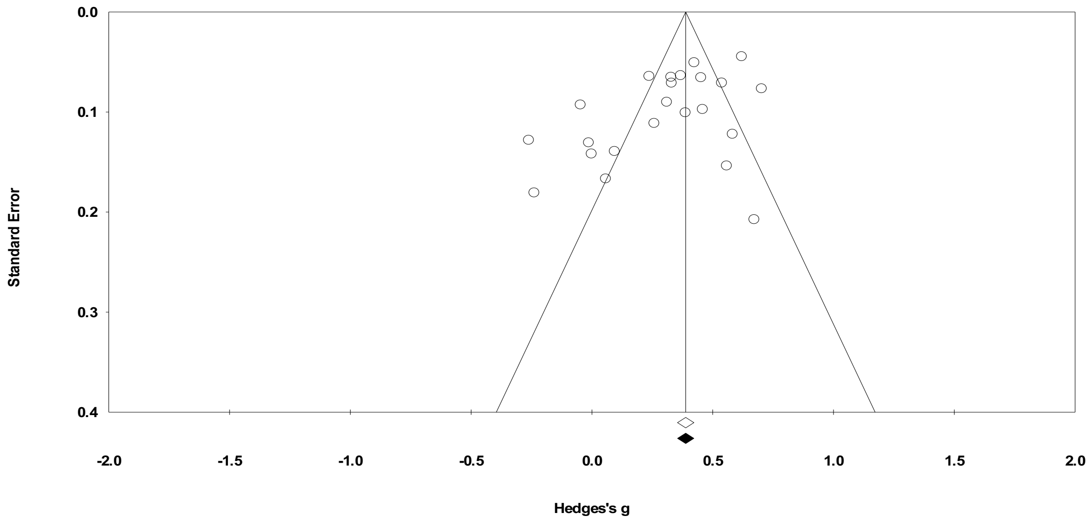

Figure 2C. Funnel plot of standard error by Hedges's *g* for communication results. White circles indicate observed groups, the white diamond indicates the overall random weighted effect size for the observed groups. The black diamond indicates the overall random weighted effect size adjusted for any missing studies.

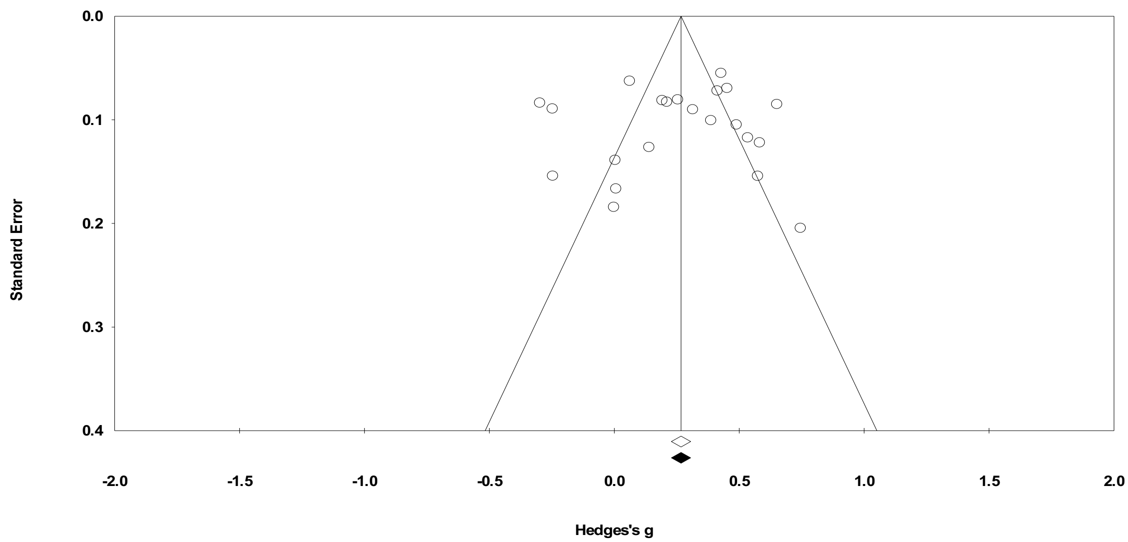

Figure 3C. Funnel plot of standard error by Hedges's *g* for social results. White circles indicate observed groups, the white diamond indicates the overall random weighted effect size for the observed groups. The black diamond indicates the overall random weighted effect size adjusted for any missing studies.

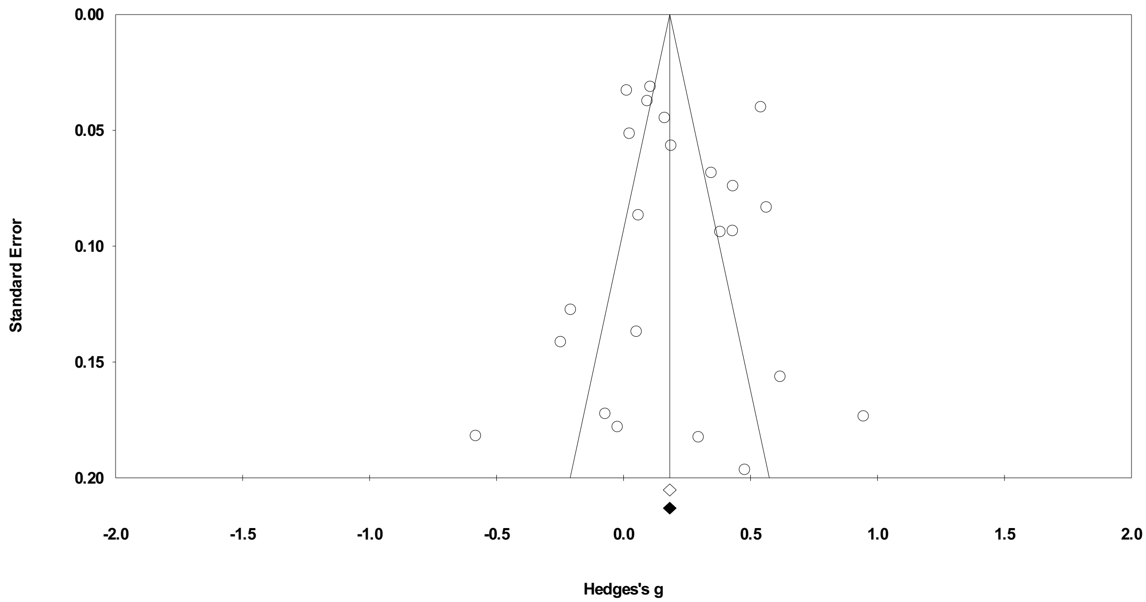

Figure 4C. Funnel plot of standard error by Hedges's *g* for adaptive behavior results. White circles indicate observed groups, the white diamond indicates the overall random weighted effect size for the observed groups. The black diamond indicates the overall random weighted effect size adjusted for any missing studies.

CHAPTER 2: The Effects of Preschool Characteristics on Outcomes for Children with Autism Spectrum Disorder Receiving Community-based Early Intervention

Abstract

Seventy-nine preschoolers with Autism Spectrum Disorder (ASD, 66 males, mean age = 44.8 (6.9) months) received community-based preschool early intervention (EI) in an urban school district and were followed for nine months. EI provider use of recommended intervention practices for young children with ASD was observed at each child's primary intervention setting (n per setting type: Autism-Only = 28, Mixed Disability = 25, Inclusion = 26). Autism-Only settings demonstrated the best implementation of classroom structure, classroom environment, and curriculum and instruction recommended practices, while Inclusion settings were better at supporting social and peer relationships. The implementation of practices to support social and peer relationships emerged as a unique predictor of cognitive gains after participation in community-based preschool EI for nine months, and was particularly beneficial for children with lower initial receptive language skills. The implications for research and practice are discussed.

Keywords: Autism Spectrum Disorders, preschool, early intervention, community settings

The Effects of Preschool Characteristics on Outcomes for Children with Autism Spectrum Disorder Receiving Community-based Early Intervention

The purpose of this study was to examine the outcomes of preschool children with Autism Spectrum Disorder (ASD) receiving community-based early intervention services and to determine what characteristics of those services and settings were associated with positive outcomes. Children receiving high-quality early intervention (EI) services can make large gains in cognitive, communication, and adaptive behaviors skills, with positive long term effects (McEachin, Smith, & Lovaas, 1993; Reichow, Barton, Boyd, & Hume, 2012; Rogers & Vismara, 2008; Weitlauf et al., 2014). Most studies demonstrating these gains were conducted in university-based research settings using expert clinicians and examined the outcomes of highly manualized interventions. These studies provide little insight into the extent to which outcomes from community practice mirror those found in university-based trials.

As Kasari and Smith (2013) note, the large majority of children served in public schools are not represented in studies typifying the "evidence base" because the large majority of studies are of small size, include homogeneous samples. In addition, these studies also often exclude children with ASD who test as lower functioning, nonverbal, and non-English speaking, and who have multiple disabilities. Emerging research is beginning to demonstrate that evidence-based interventions (EBIs) can be disseminated to, and implemented in, community settings (Kasari et al., 2014; Smith et al., 2010; Vivanti et al., 2014) and that community providers can be trained in EBIs (Lawton &

Kasari, 2012; Shire & Kasari, 2014) with heavy expert support. However, these results do not address current community practices.

Studying intervention as it is delivered in community settings can provide important insights into which practices have the potential to be most effective in the context of the resources available in these settings (Stahmer & Aarons, 2009). However, few studies have measured both the type and quality of intervention in community-based settings and the associated outcomes for participants, or have included diverse samples that represent the full range of backgrounds and levels of functioning of children with ASD. This small body of research suggests that evidence-based interventions for youth with ASD are rarely found in community settings where most youth with ASD receive services (Wood, McLeod, Klebanoff, & Brookman-Frazee, 2015). Previous studies have surveyed EI providers about which practices they use (Stahmer, Collings, & Palinkas, 2005), and have found that often their practices do not mirror the evidence-base. Interpretations of the extant literature on community-based EI for children with ASD is further limited in that many studies were retrospective (Flanagan, Perry, & Freeman, 2012; Perry et al., 2008), or lacked a comparison group, relied on outcomes only from one program, type of intervention, or intervention setting (Ben Itzchak & Zachor, 2009; Eapen, Črnčec, & Walter, 2013; Fernell et al., 2011; Magiati, Charman, & Howlin, 2007; Stahmer, Akshoomoff, & Cunningham, 2011).

EI settings vary considerably in the extent to which the intervention delivered mirrors the interventions studied in the evidence base. A second dimension on which they vary is the extent to which children with ASD interact with typically developing children.

Although consistent opportunities to interact with typically developing peers often are a recommended practice for young children with ASD (Koegel, Robinson, & Koegel, 2009; National Research Council, 2001; Strain, Wolery, & Izeman, 1998; Tsai, 1998), and the few studies of inclusive preschool programs for children with ASD suggest that preschoolers with ASD can make gains in cognitive, academic, language, functional and social skills when placed with their typically developing peers (Ferraioli & Harris, 2011; Odom et al., 2004; Schwartz, Sandall, McBride, & Boulware, 2004; Strain & Bovey, 2011), there remains debate about the appropriateness of inclusive settings for children with ASD (Barned, Knapp, & Neuharth-Pritchett, 2011; Lowenthal, 1999; Mesibov & Shea, 1996; Odom et al., 2006). Most research to date has evaluated interventions implemented in more segregated settings (such as individual services provided in homes or clinics, or in classrooms consisting only children with ASD) that do not routinely offer such opportunities (National Research Council, 2001; Odom, Collet-Klingenberg, Rogers, & Hatton, 2010), and has not compared inclusive to non-inclusive settings. The few comparative studies of these types of settings types are inconclusive (Boyd et al., 2014; Harris, Handleman, Kristoff, Bass, & Gordon, 1990).

To our knowledge, only one study has examined how different community-based EI setting characteristics are associated with child outcomes. Nahmias and colleagues (2014) found that children with ASD who received preschool early intervention in inclusive placements had higher cognitive scores when they started elementary school than children who received early intervention in more restrictive placements. This was particularly the case for children with baseline higher communication skills, lower social-emotional skills, and lower adaptive behavior skills. The authors hypothesized that inclusive placements provide more opportunities to interact with typically developing peers, which, in turn, may be associated with better outcomes, but they lacked the data to test this hypothesis. This study also was hampered by its retrospective design and limited characterization of both children and intervention.

Across intervention programs, children with ASD vary in their response to treatment, and although a priority for parents and providers alike, predicting which children will respond to which intervention remains a challenge. In addition to the moderators reported by Nahmias and colleagues (2014), language abilities (Gordon et al., 2011; Kasari, Paparella, Freeman, & Jahromi, 2008), social-communication skills (Kasari et al., 2008; Yoder & Stone, 2006b, 2006a), adaptive behavior (Eldevik et al., 2010; Flanagan et al., 2012; Remington et al., 2007), IQ (Eldevik et al., 2010; Harris & Handleman, 2000; Magiati et al., 2007; Perry et al., 2011; Remington et al., 2007), object exploration (Carter et al., 2011; Yoder & Stone, 2006b, 2006a), age (Flanagan et al., 2012; Harris & Handleman, 2000; Perry et al., 2011), and autism severity (Gordon et al., 2011; Remington et al., 2007) have emerged as potential moderators of outcomes for young children with ASD in some studies of various treatment programs. Their interaction with characteristics of preschool settings in predicting outcomes, however, has not been examined prospectively.

We built on the study by Nahmias and colleagues (2014) and other research by following preschoolers with ASD as they received publicly-funded preschool early intervention services provided across the full range of EI placements (i.e., inclusive, mixed-disability, autism-only) in one city. We carefully characterized children at baseline, measured characteristics of the intervention they received, and assessed outcomes at 9 months. This study takes advantage of the variation in intervention practices in community settings, and provides insight into characteristics of interventions most associated with positive outcomes. Findings from the present study will provide a benchmark against which to measure future progress in community-based interventions, and to the extent that certain intervention characteristics are associated with better outcomes, it can lead to experimental studies of what works best in community settings.

Methods

Participants

The sample consists of children with ASD receiving preschool early intervention in an urban school district. Participants were assigned to intervention services based on standard community practices in a naturalistic study design. Children were eligible to enroll in this study if they: 1) were between 36 and 59 months of age; 2) had a documented diagnosis of ASD; and 3) received services through the public preschool EI or behavioral health system. Exclusion criteria were: 1) caregivers do not speak English or 2) either the caregiver or EI provider for a given child does not consent to participate. The sample was recruited on a rolling basis through the preschool early intervention system from July 2014 to August 2016.

Procedure

Children were assessed at two time points: "T1" at entry to the study and "T2" after 9 months of preschool early intervention services. This time frame of 9 months was selected because it is the standard length of the academic year.

Measures

Mullen Scales of Early Learning (MSEL). Our primary outcome measure was the Mullen Scales of Early Learning (Mullen, 1995) a standardized, reliable, and valid measure of early cognitive development for children from birth to 68 months old. The Early Learning Composite (ELC) is based on 4 MSEL scales (visual reception (VR), fine motor (FM), expressive language (EL), and receptive language (RL)). The Mullen covers the full age range of our sample. Because more than half of the sample had T-scores below 20 on the Receptive and Expressive Language scales at baseline, developmental quotients were calculated by dividing the age equivalent by the child's chronological age in months and multiplying by 100 as previously done in the ASD literature (see Chawarska, Klin, Paul, Macari, & Volkmar, 2009; Eapen, Črnčec, & Walter, 2013; Kaale, Smith, & Sponheim, 2012). For the ELC, the developmental quotient was calculated by first averaging the age equivalences of the VR, FM, EL, and RL scales. The change on the MSEL ELC DQ between baseline and follow-up was used as the primary dependent variable. MSEL EL and RL scores were explored as predictors of treatment response.

Educational Program Review. Our primary independent variables were derived from the Educational Program Review (EPR), a measure of classroom characteristics that was developed specifically for characterizing settings in which children with autism receive early intervention. It is also known as the PDA Program Assessment (Professional Development in Autism Center, 2008), and has been validated (Hume et al., 2011) and used in other studies of preschool intervention (Boyd et al., 2014). The EPR consists of seven subscales describing recommended practices for children with ASD: Teaming, Classroom Structure, Classroom Environment, Curriculum and Instruction, Social/Peer Relationships, Management of Challenging Behaviors, and Building a Positive Instructional Climate. Items are rated on a 1-5 scale from 1 = "Minimal/no implementation," 3 = "Partial Implementation," to 5= "Full implementation" by a rater based on a 60-minute direct observation and a teacher/EI provider interview. Post-doctoral fellows, graduate students, and research assistants were trained to reliability by an expert coder. Two scales (Management of Challenging Behaviors and Teaming) were not included in analyses due to insufficient interrater reliability and missing data, respectively. The remaining five scales demonstrated excellent interrater reliability, as they were all above .75 (Fleiss, 1986). The intraclass correlations (ICC) Type 1,1 for each scale were as follows: Classroom Structure = .81, Classroom environment = .91, Curriculum and Instruction = .88, Social/Peer Relationships = .94, and Building a Positive Instructional Climate = .78. Setting type was also used as an independent variable and was coded as: home, Autism-only, Mixed-Disability, or Inclusion based on teacher report during the EPR. The EPR was conducted in the intervention setting that the participant spent the most time in at the approximate halfway point of the study.

Autism Diagnostic Observation Schedule 2nd Edition (ADOS). The Autism Diagnostic Observation Schedule-2nd Edition (Lord et al., 2012) is a semi-structured play-based assessment considered to be the gold-standard observational measure for assessing the presence of ASD. The ADOS was administered by a graduate student or post-doctoral fellow in psychology trained to research reliability and supervised by a licensed clinical psychologist. Calibrated Severity Scores were used in analyses.

Social Communication Questionnaire (SCQ). The Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003) is a brief parent questionnaire that evaluates the presence of ASD based on questions from the Autism Diagnostic Interview-Revised (Lord, Rutter, & Le Couteur, 1994). Raw scores were used as a measure of parent-reported ASD symptoms.

Adaptive Behavior Assessment System- 2nd Edition (ABAS). Children's functioning was measured using the Adaptive Behavior Assessment System- 2nd Edition (Harrison & Oakland, 2003), a parent-report questionnaire used to assess adaptive behavior in the home. Subscales include communication, social, and daily living skills, which can be combined into a global adaptive composite. Due to our previous work suggesting the potential role of adaptive behavior and social-emotional skills in moderating treatment effects (Nahmias et al., 2014), standard scores from the Social composite and Global Adaptive Composite were explored as moderators in analyses.

Developmental Play Assessment (DPA). The DPA was adapted from (Lifter, Gitlin-Weiner, Sandgrund, & Schafer, 2000) to measure children's interest in playing with objects and toys. This measure was administered as described in Carter et al. (2011). Briefly, an assessor presented two standard sets of toys within the child's reach for approximately 3.5 minutes each during a free-play session. The number of toys with which children used differentiated play at Time 1 was coded by raters trained to reliability. Undifferentiated actions (e.g., mouthing, banging, shaking, close inspection) were not coded. Inter-rater reliability of object interest (i.e., number of toys played with) was excellent (ICC (1,1) = .93). Object interest was explored as a moderator due to its association with treatment gains in other studies (Carter et al., 2011; Yoder & Stone, 2006b, 2006a).

Demographic questionnaire. We also collected demographic data from parents and teachers/EI providers to use as covariates in analyses.

Analytic strategy

First, we conducted One-way Analysis of Variance (ANOVA) for continuous variables and chi square tests for dichotomous variables by site type. Bonferroni corrected post-hoc comparisons were utilized to explore differences by setting type on continuous variables. Cohen's *d* was utilized as an effect size metric. Cohen's *d* can be interpreted as follows: 0.2 as small, 0.5 as medium, and 0.8 as large (Cohen, 1988). A paired t-test was used to assess change over time across all participants. Next, we used linear regression to examine the effects of child characteristics (e.g., baseline scores, demographic features) and setting characteristics (e.g., EPR scales, setting type) on changes in the MSEL ELC. We first examined unadjusted models to determine which putative moderators and covariates would be included in the final adjusted model. Variables with a bivariate association with the outcome significant at $p \leq .20$ were included in the multiple regression model. This screening criteria for initial variable selection is based on the recommendation of Hosmer, Lemeshow, and Sturdivant (2013) who have found that the use of a more traditional p-value (such as 0.05) often fails to identify variables known to be important and influential confounders, while the use of a higher p-value has the disadvantage of including variables of questionable importance. Standardized DFBETAs were examined to assess for influential outliers. We then tested for interactions between setting and EPR subscales with proposed child characteristic moderators of interest. Variables were mean-centered prior to the creation of the interaction term to facilitate interpretation and reduce collinearity. Significant interactions were then probed for regions of significance utilizing the method described by Preacher, Curran, and Bauer (2006). The region of significance defines the specific values of a moderator at which the regression of an outcome (i.e., MSEL ELC change) on a focal predictor (e.g., setting type, EPR scale) moves from non-significance to significance. If a region of significance contains no data, that region is considered uninterpretable.

Results

Sample description

Children. Eighty-six participants had sufficient data to be included in the analytic sample (i.e., completed the MSEL at T1 and T2 and the EPR). As only seven participants received EI primarily in a home setting, this group was excluded. Descriptive statistics for the analyzed sample of 79 children can be found in Table 1. At the time of the EPR, 28 participants received services in an Autism Only setting, 25 in a Mixed Disability setting, and 26 in an Inclusion setting. The participants were predominately male (84%), the plurality were black (44%) and most had a household income below $40,000 (58%). Seventy-nine percent met autism spectrum cut-off scores based on the ADOS. As seen in

Table 1 some baseline participant characteristics varied by setting, including sex and participation in birth-to-3 EI services (all *p*s < .05). Children in Inclusion settings had significantly higher receptive language, expressive language and cognitive skills based on the MSEL, and significantly higher adaptive behavior and social skills based on the ABAS, than did children in the other two settings (all *p*s < .05). Children in Autism Only settings had significantly higher clinician-rated ASD symptoms based on the ADOS and younger age of ASD diagnosis than children in the other two settings (all *p*s < .05).

INSERT TABLE 1 HERE.

Teacher/EI providers. Forty-two providers out of 66 participating providers completed the teacher/EI provider demographic questionnaire. As seen in Table 2, teachers and EI providers were largely female (79%), white (59%), and reported receiving regular autism training (45%). Most providers also identified their current role as a Special Education preschool teacher (64%).

INSERT TABLE 2 HERE.

Use of Recommended Practices by Setting

As seen in Figure 1, the use of recommended practices based on the EPR showed large differences among the setting types. Autism Only settings had significantly better implementation of classroom structure, classroom environment, and curriculum and instruction recommended practices than Inclusion settings (all *p*s < .05, *d* = 1.8, 0.9, 1.5, respectively). Mixed Disability settings had significantly better implementation of classroom structure ($p < .05$, $d = 1.1$) and curriculum and instruction ($p < .05$, $d = 0.9$) practices than Inclusion settings, but worse implementation of classroom structure than

Autism Only settings ($p < .05$, $d = -0.9$). Inclusion settings had significantly better implementation of recommended strategies to support social and peer relationships than the other settings (all *p*s < .05, $d = 1.7$ for Autism Only and 1.0 for Mixed Disability).

INSERT FIGURE 1 HERE.

Cognitive Outcomes

Full sample. Participants across all three groups improved in the Mullen ELC score between Time 1 (Mean = 54.12, SD = 19.59) and Time 2 (Mean = 56.64, SD = 20.15, $t(78) = -2.4$, $p = .02$), but this effect was small ($d = 0.3$)

Unadjusted models. Table 3 presents the results of the regression analyses in predicting change in overall cognitive ability (as measured by the MSEL ELC). In unadjusted analyses, no measures significantly predicted changes in Mullen ELC scores, including the child characteristics that the groups differed on at baseline.

Adjusted models. Table 3 presents the results of the adjusted regression analysis predicting change in MSEL ELC. In the adjusted analysis, in which only variables with a bivariate statistical significance of $p < 0.2$ were included, only the use of recommended practices supporting social and peer relationships significantly predicted children's cognitive outcome at $p < 0.05$. Each point increase on the EPR Social/Peer Relationships scale was associated with a 4.40 point average increase in MSEL ELC change score. There were no significant main effects for setting type.

INSERT TABLE 3 HERE.

Interactions. Of the initial putative moderators, only receptive language on the MSEL met criteria for inclusion in the adjusted model. As presented in Figure 2, receptive language significantly moderated the relationship between the implementation of recommended practices to support social and peer relationships (based on the EPR) and cognitive changes (*B* = -0.14, *SE* = 0.05, *p* = .005). Only the lower region of significance was interpretable, such that for children with baseline MSEL RL developmental quotients below 56.35, children with lower baseline receptive language scores made greater gains on the MSEL ELC in settings with higher EPR Social/Peer Relationship Scale scores. Although the interaction between baseline MSEL Receptive Language and Autism Only as compared to Inclusion settings was also significant (*B* = 0.21, *SE* = 0.10, $p < .05$), none of the regions of significance were interpretable (see Appendix A for figure).

INSERT FIGURE 2 HERE.

Discussion

The present study provides some of the most rigorous observational evidence to date of both expected outcomes across a variety of community-based early intervention programs and the association of characteristics of these programs with children's cognitive outcomes.

The main effects and moderation results by setting from our earlier work (Nahmias et al., 2014) were not replicated when controlling for implementation of recommended practices to support social/peer relationships. One possible explanation for this discrepancy in findings may be related to our use of different measures (i.e., MSEL and ABAS instead of DAYC (Voress & Maddox, 1998) and DAS (Elliott, 2007)) over shorter intervention period (i.e., 9.3 months vs. 2.1 years) than our previous study.

Future research should explore longer-term impacts of community-based early intervention.

Settings that served only children with ASD had better implementation of almost all recommended practices, but only implementation of strategies to support social/peer relationships (more common in inclusion settings) was significantly associated with overall cognitive gains. It may be that use of recommended classroom structure, curriculum and instruction, and classroom environment strategies are associated with gains in domains not assessed in this study (e.g., challenging behaviors, academic readiness). A previous study (Boyd et al., 2014) demonstrated that community programs with at least partial implementation based on the EPR (average scores above 3) had significant improvements in autism characteristics and severity, communication, and fine motor skills (cognitive skills were not directly explored as an outcome). Another possible explanation this pattern of results may be that community providers need more training/support to appropriately individualize and tailor their use of the variety the recommended practices they were implementing (Kasari & Smith, 2016).

The development of social relationships is an important part of typical development as well as for children with ASD (Frankel et al., 2010; Kasari, Locke, Ishijima, & Kretzmann, 2013). Research (McConnell, 2002; National Autism Center, 2015; Odom et al., 2010) has demonstrated that interventions targeting social skills can lead to gains in those skills as well as language. Less research has explored the effect that targeting social and peer relationships has on cognitive outcomes. One exception is a RCT comparing training in a comprehensive treatment model (LEAP) that includes peer-mediated interventions as a core component to a manual-only condition, which demonstrated significant cognitive gains (in addition to improved language, reduced autism symptom severity, improved social behavior and reduced problem behavior) compared with a comparison condition (Strain & Bovey, 2011). It may be that supporting social and peer relationships leads to improved attention when interacting with others, resulting in improved performance during the Mullen. Previous research with elementary school children without ASD (McClelland, Morrison, & Holmes, 2000) found associations between learning-related social skills and later academic performance. In elementary school-aged children with ASD (Pellecchia et al., 2016), baseline parent reported social phobia symptoms was associated with cognitive gains. As social-communication differences are one of the core diagnostic features of ASD (American Psychiatric Association, 2013), it may be that by targeting the core social deficit leads to increased learning opportunities and generalization of gains by enabling children with ASD to take more advantage of interactions with peers that model age-appropriate behavior, social, play, and language skills.

We found that better implementation of recommended practices to support social and peer relationships was associated with greater cognitive gains for all children with ASD in our sample, and especially those lower baseline receptive language skills. This suggests that a common educational approach of placing most children with ASD in segregated education settings (Strain, 2017), may be missing an opportunity to maximize cognitive gains in early intervention. As Pellecchia and colleagues (2016) suggested in regards to school-based interventions elementary-age children with ASD, programs may need to adapt their educational approach, as addressing social impairment and supporting social interactions may be a necessary precursor or adjunct to improving response to intervention. Programs in which EI provider supported social and peer relationships are less common (e.g., more restrictive placements) may want to consider incorporating supported instruction with peers for all children with ASD, regardless of baseline functioning level. If opportunities for social and peer relationships are not available as part of a child's publicly funded EI program, families may want to consider pursuing additional programming.

As some of the very early items on the Receptive Language scale of the Mullen Scales of Early Learning could also be conceptualized as relating to attention (e.g., reacts reflexively to loud noise, alerts to sound, attends to words and movement), a possible alternate explanation is that very low baseline receptive language scores were indications of poor attention skills or behavioral challenges during testing. In a sensitivity analysis, the percentage of time the child was observed to be on task during a 10-minute observation of the child in his or her early intervention placement was included as a covariate in the adjusted model for the 70 participants that this measure was available for. The EPR Social/Peer Relationships scale and the interaction between the EPR Social/Peer Relationships scale and MSEL Receptive Language scores remained significant, with similar coefficients as in the model that did not include on task behavior (see Appendix B). The addition of observed on task behavior accounted for an additional 7.7% of the variance in MSEL ELC change scores, and was a significant predictor of large cognitive changes ($B = 11.94$, $SE = 4.5$, $p = .01$). This suggests that children's abilities to sustain attention and participate in the learning opportunities in their EI placement (or how well their EI placement facilitates this for them) is an important predictor of cognitive gains across setting types. In addition, even controlling for observed on-task behavior during intervention programming, supporting social and peer relationships remains an important contributor to cognitive gains, especially for children low receptive language skills.

This study has several limitations that warrant mentioning. First, as this was a naturalistic study, children were not randomly assigned to placements, so unmeasured contributors cannot be ruled out. Due to the challenges inherent with working with low-resource families (e.g., phone disconnection), we were unable to collect all measures from all participants. Our measurement of the intervention received was limited to a snapshot of the child's intervention programming, so we may not have fully captured the full extent of the intervention they received across the 9 months.

Despite these limitations, there are promising future directions for this research. The use of recommended practices to support social and peer relationships emerged as a particularly important set of practices associated with cognitive gains in community-based preschool EI. Further examination of specific practices to test utilizing better controlled studies (e.g., randomized controlled trials) is warranted. In addition, strategies that support children's on-task behavior and ability to access the EI curriculum may warrant further research.

References

American Psychiatric Association (Ed.). (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (5th ed). Washington, D.C: American Psychiatric Association.

Barned, N. E., Knapp, N. F., & Neuharth-Pritchett, S. (2011). Knowledge and Attitudes of Early Childhood Preservice Teachers Regarding the Inclusion of Children With Autism Spectrum Disorder. *Journal of Early Childhood Teacher Education*, *32*(4), 302–321. https://doi.org/10.1080/10901027.2011.622235

Ben Itzchak, E., & Zachor, D. A. (2009). Change in autism classification with early intervention: Predictors and outcomes. *Research in Autism Spectrum Disorders*, *3*(4), 967–976. https://doi.org/10.1016/j.rasd.2009.05.001

Boyd, B. A., Hume, K., McBee, M. T., Alessandri, M., Gutierrez, A., Johnson, L., … Odom, S. L. (2014). Comparative Efficacy of LEAP, TEACCH and Non-Model-Specific Special Education Programs for Preschoolers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *44*(2), 366–380. https://doi.org/10.1007/s10803-013-1877-9

Carter, A. S., Messinger, D. S., Stone, W. L., Celimli, S., Nahmias, A. S., & Yoder, P. (2011). A randomized controlled trial of Hanen's "More Than Words" in toddlers with early autism symptoms. *Journal of Child Psychology and Psychiatry*, *52*(7), 741–752. https://doi.org/10.1111/j.1469-7610.2011.02395.x

Chawarska, K., Klin, A., Paul, R., Macari, S., & Volkmar, F. (2009). A prospective study of toddlers with ASD: short-term diagnostic and cognitive outcomes. *Journal of*

Child Psychology and Psychiatry, *50*(10), 1235–1245. https://doi.org/10.1111/j.1469-7610.2009.02101.x

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Eapen, V., Črnčec, R., & Walter, A. (2013). Clinical outcomes of an early intervention program for preschool children with Autism Spectrum Disorder in a community group setting. *BMC Pediatrics*, *13*(1). https://doi.org/10.1186/1471-2431-13-3

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2010). Using Participant Data to Extend the Evidence Base for Intensive Behavioral Intervention for Children With Autism. *American Journal on Intellectual and Developmental Disabilities*, *115*(5), 381–405. https://doi.org/10.1352/1944-7558-115.5.381

Elliott, C. (2007). *Differential Abilities Scales* (2nd ed.). San Antonio, TX: Harcourt Assessment, Inc.

Fernell, E., Hedvall, Å., Westerlund, J., Höglund Carlsson, L., Eriksson, M., Barnevik Olsson, M., … Gillberg, C. (2011). Early intervention in 208 Swedish preschoolers with autism spectrum disorder. A prospective naturalistic study. *Research in Developmental Disabilities*, *32*(6), 2092–2101. https://doi.org/10.1016/j.ridd.2011.08.002

Ferraioli, S. J., & Harris, S. L. (2011). Effective Educational Inclusion of Students on the Autism Spectrum. *Journal of Contemporary Psychotherapy*, *41*(1), 19–28. https://doi.org/10.1007/s10879-010-9156-y

Flanagan, H. E., Perry, A., & Freeman, N. L. (2012). Effectiveness of large-scale community-based Intensive Behavioral Intervention: A waitlist comparison study exploring outcomes and predictors. *Research in Autism Spectrum Disorders*, *6*(2), 673–682. https://doi.org/10.1016/j.rasd.2011.09.011

Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York: John Wiley.

Frankel, F., Myatt, R., Sugar, C., Whitham, C., Gorospe, C. M., & Laugeson, E. (2010). A Randomized Controlled Study of Parent-assisted Children's Friendship Training with Children having Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *40*(7), 827–842. https://doi.org/10.1007/s10803-009-0932-z

Gordon, K., Pasco, G., McElduff, F., Wade, A., Howlin, P., & Charman, T. (2011). A communication-based intervention for nonverbal children with autism: What changes? Who benefits? *Journal of Consulting and Clinical Psychology*, *79*(4), 447–457. https://doi.org/10.1037/a0024379

Harris, S. L., & Handleman, J. S. (2000). Age and IQ at Intake as Predictors of Placement for Young Children with Autism: A Four- to Six-Year Follow-Up. *Journal of Autism and Developmental Disorders*, *30*(2), 137–142. https://doi.org/10.1023/A:1005459606120

Harris, S. L., Handleman, J. S., Kristoff, B., Bass, L., & Gordon, R. (1990). Changes in language development among autistic and peer children in segregated and integrated preschool settings. *Journal of Autism and Developmental Disorders*, *20*(1), 23–31.

Harrison, P., & Oakland, T. (2003). *Adaptive Behavior Assessment System* (2nd ed.). San Antonio, TX: The Psychological Corporation.

Hosmer, D., Lemeshow, S., & Sturdivant, R. (2013). *Applied Logistic Regression* (3rd ed.). Hoboken, NJ: John Wiley.

Hume, K., Boyd, B., McBee, M., Coman, D., Gutierrez, A., Shaw, E., … Odom, S. (2011). Assessing implementation of comprehensive treatment models for young children with ASD: Reliability and validity of two measures. *Research in Autism Spectrum Disorders*, *5*(4), 1430–1440. https://doi.org/10.1016/j.rasd.2011.02.002

Kaale, A., Smith, L., & Sponheim, E. (2012). A randomized controlled trial of preschool-based joint attention intervention for children with autism: Preschool-based JA-intervention for children with autism. *Journal of Child Psychology and Psychiatry*, *53*(1), 97–105. https://doi.org/10.1111/j.1469-7610.2011.02450.x

Kasari, C., Lawton, K., Shih, W., Barker, T. V., Landa, R., Lord, C., … Senturk, D. (2014). Caregiver-mediated intervention for low-resourced preschoolers with autism: an RCT. *Pediatrics*, *134*(1), e72-79. https://doi.org/10.1542/peds.2013-3229

Kasari, C., Locke, J., Ishijima, E., & Kretzmann, M. (2013). Peer acceptance, social engagement and friendship: critical goals for children with autism spectrum disorders. *Social Skills and Adaptive Behavior in Learners with Autism Spectrum Disorders: Current Status and Future Directions. Brooks Publishing*.

Kasari, C., Paparella, T., Freeman, S., & Jahromi, L. B. (2008). Language outcome in autism: Randomized comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology*, *76*(1), 125–137. https://doi.org/10.1037/0022-006X.76.1.125

Kasari, C., & Smith, T. (2013). Interventions in schools for children with autism spectrum disorder: Methods and recommendations. *Autism*, *17*(3), 254–267. https://doi.org/10.1177/1362361312470496

Kasari, C., & Smith, T. (2016). Forest for the Trees: Evidence-Based Practices in ASD. *Clinical Psychology: Science and Practice*, *23*(3), 260–264. https://doi.org/10.1111/cpsp.12161

Koegel, L. K., Robinson, S., & Koegel, R. L. (2009). Empirically Supported Intervention Practices for Autism Spectrum Disorders in School and Community Settings: Issues and Practices. In W. Sailor, G. Dunlap, G. Sugai, & R. Horner (Eds.), *Handbook of Positive Behavior Support* (pp. 149–176). Springer US. https://doi.org/10.1007/978-0-387-09632-2_7

Lawton, K., & Kasari, C. (2012). Teacher-implemented joint attention intervention: Pilot randomized controlled study for preschoolers with autism. *Journal of Consulting and Clinical Psychology*, *80*(4), 687–693. https://doi.org/10.1037/a0028506

Lifter, K., Gitlin-Weiner, K., Sandgrund, A., & Schafer, C. (2000). Linking assessment to intervention for children with developmental disabilities or at-risk for developmental delay: The developmental play assessment (DPA) instrument. *Play Diagnosis and Assessment*, 228–261.

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). Autism diagnostic observation schedule–2nd edition (ADOS-2). *Los Angeles, CA: Western Psychological Corporation*.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659–685. https://doi.org/10.1007/BF02172145

Lowenthal, B. (1999). Early Childhood Inclusion in the United States. *Early Child Development and Care*, *150*(1), 17–32. https://doi.org/10.1080/0300443991500102

Magiati, I., Charman, T., & Howlin, P. (2007). A two-year prospective follow-up study of community-based early intensive behavioural intervention and specialist nursery provision for children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry*, *48*(8), 803–812. https://doi.org/10.1111/j.1469-7610.2007.01756.x

McClelland, M. M., Morrison, F. J., & Holmes, D. L. (2000). Children at risk for early academic problems: the role of learning-related social skills. *Early Childhood Research Quarterly*, *15*(3), 307–329. https://doi.org/10.1016/S0885-2006(00)00069-7

McConnell, S. R. (2002). Interventions to Facilitate Social Interaction for Young Children with Autism: Review of Available Research and Recommendations for Educational Intervention and Future Research. *Journal of Autism and*

Developmental Disorders, *32*(5), 351–372. https://doi.org/10.1023/A:1020537805154

McEachin, J. J., Smith, T., & Lovaas, O. I. (1993). Long-term outcome for children with autism who received early intensive behavioral treatment. *American Journal of Mental Retardation: AJMR*, *97*(4), 359-372-391.

Mesibov, G. B., & Shea, V. (1996). Full inclusion and students with autism. *Journal of Autism and Developmental Disorders*, *26*(3), 337–346. https://doi.org/10.1007/BF02172478

Mullen, E. M. (1995). *Mullen Scales of Early Learning manual*. Circle Pines, MN: American Guidance Service.

Nahmias, A. S., Kase, C., & Mandell, D. S. (2014). Comparing cognitive outcomes among children with autism spectrum disorders receiving community-based early intervention in one of three placements. *Autism*, *18*(3), 311–320. https://doi.org/10.1177/1362361312467865

National Autism Center. (2015). *Findings and conclusions: National standards project, phase 2*. Randolph, MA.

National Research Council. (2001). *Educating Children with Autism*. National Academies Press.

Odom, S. L., Collet-Klingenberg, L., Rogers, S. J., & Hatton, D. D. (2010). Evidence-Based Practices in Interventions for Children and Youth with Autism Spectrum Disorders. *Preventing School Failure: Alternative Education for Children and Youth*, *54*(4), 275–282. https://doi.org/10.1080/10459881003785506

Odom, S. L., Vitztum, J., Wolery, R., Lieber, J., Sandall, S., Hanson, M. J., … Horn, E. (2004). Preschool inclusion in the United States: a review of research from an ecological systems perspective. *Journal of Research in Special Educational Needs*, *4*(1), 17–49. https://doi.org/10.1111/J.1471-3802.2004.00016.x

Odom, S. L., Zercher, C., Li, S., Marquart, J. M., Sandall, S., & Brown, W. H. (2006). Social Acceptance and Rejection of Preschool Children With Disabilities: A Mixed-method Analysis. *Journal of Educational Psychology*, *98*(4), 807–823.

Pellecchia, M., Connell, J. E., Kerns, C. M., Xie, M., Marcus, S. C., & Mandell, D. S. (2016). Child characteristics associated with outcome for children with autism in a school-based behavioral intervention. *Autism*, *20*(3), 321–329. https://doi.org/10.1177/1362361315577518

Perry, A., Cummings, A., Dunn Geier, J., Freeman, N. L., Hughes, S., LaRose, L., … Williams, J. (2008). Effectiveness of Intensive Behavioral Intervention in a large, community-based program. *Research in Autism Spectrum Disorders*, *2*(4), 621–642. https://doi.org/10.1016/j.rasd.2008.01.002

Perry, A., Cummings, A., Geier, J. D., Freeman, N. L., Hughes, S., Managhan, T., … Williams, J. (2011). Predictors of outcome for children receiving intensive behavioral intervention in a large, community-based program. *Research in Autism Spectrum Disorders*, *5*(1), 592–603. https://doi.org/10.1016/j.rasd.2010.07.003

Preacher, K. J., Curran, P. J., & Bauer, D. J. (2006). Computational Tools for Probing Interactions in Multiple Linear Regression, Multilevel Modeling, and Latent

Curve Analysis. *Journal of Educational and Behavioral Statistics*, *31*(4), 437–448. https://doi.org/10.3102/10769986031004437

Professional Development in Autism Center. (2008). *PDA Program Assessment*. Seattle, WA: PDA Center at University of Washington.

Reichow, B., Barton, E. E., Boyd, B. A., & Hume, K. (2012). Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD). In The Cochrane Collaboration (Ed.), *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/14651858.CD009260.pub2

Remington, B., Hastings, R. P., Kovshoff, H., degli Espinosa, F., Jahr, E., Brown, T., … Ward, N. (2007). Early Intensive Behavioral Intervention: Outcomes for Children With Autism and Their Parents After Two Years. *American Journal on Mental Retardation*, *112*(6), 418–438. https://doi.org/10.1352/0895-8017(2007)112[418:EIBIOF]2.0.CO;2

Rogers, S. J., & Vismara, L. A. (2008). Evidence-Based Comprehensive Treatments for Early Autism. *Journal of Clinical Child & Adolescent Psychology*, *37*(1), 8–38. https://doi.org/10.1080/15374410701817808

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*. Western Psychological Services.

Schwartz, I. S., Sandall, S. R., McBride, B. J., & Boulware, G.-L. (2004). Project DATA (Developmentally Appropriate Treatment for Autism): An Inclusive School-Based Approach to Educating Young Children with Autism. *Topics in Early Childhood*

Special Education, *24*(3), 156–168. https://doi.org/10.1177/02711214040240030301

Shire, S. Y., & Kasari, C. (2014). Train the Trainer Effectiveness Trials of Behavioral Intervention for Individuals With Autism: A Systematic Review. *American Journal on Intellectual and Developmental Disabilities*, *119*(5), 436–451. https://doi.org/10.1352/1944-7558-119.5.436

Smith, I. M., Koegel, R. L., Koegel, L. K., Openden, D. A., Fossum, K. L., & Bryson, S. E. (2010). Effectiveness of a Novel Community-Based Early Intervention Model for Children With Autistic Spectrum Disorder. *American Journal on Intellectual and Developmental Disabilities*, *115*(6), 504–523. https://doi.org/10.1352/1944-7558-115.6.504

Stahmer, A. C., & Aarons, G. A. (2009). Attitudes toward adoption of evidence-based practices: A comparison of autism early intervention providers and children's mental health providers. *Psychological Services*, *6*(3), 223–234. https://doi.org/10.1037/a0010738

Stahmer, A. C., Akshoomoff, N., & Cunningham, A. B. (2011). Inclusion for toddlers with autism spectrum disorders: The first ten years of a community program. *Autism*, *15*(5), 625–641. https://doi.org/10.1177/1362361310392253

Stahmer, A. C., Collings, N. M., & Palinkas, L. A. (2005). Early intervention practices for children with autism: Descriptions from community providers. *Focus on Autism and Other Developmental Disabilities*, *20*(2), 66–79.

Strain, P. S. (2017). Four-Year Follow-Up of Children in the LEAP Randomized Trial: Some Planned and Accidental Findings. *Topics in Early Childhood Special Education*, 271121417711531. https://doi.org/10.1177/0271121417711531

Strain, P. S., & Bovey, E. H. (2011). Randomized, Controlled Trial of the LEAP Model of Early Intervention for Young Children With Autism Spectrum Disorders. *Topics in Early Childhood Special Education*, *31*(3), 133–154. https://doi.org/10.1177/0271121411408740

Strain, P. S., Wolery, M., & Izeman, S. (1998). Considerations for Administrators in the Design of Service Options for Young Children With Autism and Their Families. *Young Exceptional Children*, *1*(2), 8–16. https://doi.org/10.1177/109625069800100202

Tsai, L. Y. (1998). Pervasive Developmental Disorders. NICHCY Briefing Paper FS20. Retrieved from https://eric.ed.gov/?id=ED416637

Vivanti, G., Paynter, J., Duncan, E., Fothergill, H., Dissanayake, C., Rogers, S. J., & the Victorian ASELCC Team. (2014). Effectiveness and Feasibility of the Early Start Denver Model Implemented in a Group-Based Community Childcare Setting. *Journal of Autism and Developmental Disorders*, *44*(12), 3140–3153. https://doi.org/10.1007/s10803-014-2168-9

Voress, J., & Maddox, T. (1998). *Developmental Assessment of Young Children (DAYC)*. Austin, TN: PRO-ED.

Weitlauf, A. S., McPheeters, M. L., Peters, B., Sathe, N., Travis, R., Aiello, R., … Warren, Z. (2014). *Therapies for Children With Autism Spectrum Disorder*.

Agency for Healthcare Research and Quality (US). Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK241444/

Wood, J. J., McLeod, B. D., Klebanoff, S., & Brookman-Frazee, L. (2015). Toward the Implementation of Evidence-Based Interventions for Youth With Autism Spectrum Disorders in Schools and Community Agencies. *Behavior Therapy*, *46*(1), 83–95. https://doi.org/10.1016/j.beth.2014.07.003

Yoder, P., & Stone, W. L. (2006a). A Randomized Comparison of the Effect of Two Prelinguistic Communication Interventions on the Acquisition of Spoken Communication in Preschoolers With ASD. *Journal of Speech, Language, and Hearing Research*, *49*(4), 698–711. https://doi.org/10.1044/1092-4388(2006/051)

Yoder, P., & Stone, W. L. (2006b). Randomized comparison of two communication interventions for preschoolers with autism spectrum disorders. *Journal of Consulting and Clinical Psychology*, *74*(3), 426–435. https://doi.org/10.1037/0022-006X.74.3.426

Tables

Table 1

Child Participant Characteristics

	Autism Only		Mixed Disability		Inclusion		Total		N
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	
Time 1 Age (months)	46.29	6.93	43.16	6.08	44.65	7.54	44.76	6.92	79
Time 1 MSEL Receptive Language DQ**	36.18^{I}	23.08	41.13^{I}	23.08	$59.00^{A,M}$	27.19	45.26	26.14	79
Time 1 MSEL Expressive Language DQ**	40.51^{I}	24.67	42.98^{I}	21.95	$63.52^{A,M}$	23.90	48.87	25.49	79
Time 1 MSEL Early Learning Composite DQ***	46.92^{I}	18.76	50.39^{I}	15.69	$65.44^{A,M}$	19.39	54.12	19.59	79
ABAS General Adaptive Composite**	55.46^{I}	8.37	57.83^{I}	9.87	$67.25^{A,M}$	15.03	60.21	12.43	71
ABAS Social **	58.63^{I}	9.45	60.39^{I}	9.12	$70.67^{A,M}$	15.77	63.27	12.89	71
ADOS Severity Score**	$7.86^{M,I}$	1.51	6.04^{A}	1.51	6.46^{A}	2.32	6.82	1.96	79
Social Communication Questionnaire	20.19	5.84	19.33	5.44	17.44	5.90	19.01	5.78	76
DPA object interest	5.13	1.96	5.56	1.62	5.41	1.59	5.34	1.73	64
Age of Diagnosis (years)***	$2.23^{M,I}$	0.79	2.97^{A}	0.60	3.05^{A}	0.74	2.72	0.80	75
	n	%	*n*	%	*n*	%	*n*	%	
Male**	24	85.71	25	100.00	17	65.38	66	83.54	79
Birth-3 EI*	26	92.86	17	70.83	23	92.00	66	85.71	77
Race/Ethnicity									79
White	8	28.57	6	24.00	4	15.38	18	22.78	
Black	11	39.29	12	48.00	12	46.15	35	44.30	
Hispanic	-	-	-	-	-	-	11	13.92	
Mixed	-	-	-	-	-	-	12	15.19	
Maternal Education									75
Did not complete High School	1	3.70	1	4.55	0	0.00	2	2.67	

	Autism Only		Mixed Disability		Inclusion		Total		N
	n	%	*n*	%	*n*	%	*n*	%	
High School or GED	4	14.81	6	27.27	6	23.08	16	21.33	
2 year college	5	18.52	3	13.64	2	7.69	10	13.33	
Some college	7	25.93	5	22.73	6	23.08	18	24.00	
College degree	4	14.81	5	22.73	8	30.77	17	22.67	
Graduate degree	6	22.22	2	9.09	4	15.38	12	16.00	
Paternal education									70
Did not complete High School	0	0.00	2	9.09	2	8.33	4	5.71	
High School or GED	6	25.00	10	45.45	10	41.67	26	37.14	
2 year college	3	12.50	2	9.09	1	4.17	6	8.57	
Some college	5	20.83	7	31.82	8	33.33	20	28.57	
College degree	7	29.17	1	4.55	2	8.33	10	14.29	
Graduate degree	3	12.50	0	0.00	1	4.17	4	5.71	
Household Income									76
Under $20,000	7	25.93	10	41.67	6	24.00	23	30.26	
$20,000-$40,000	7	25.93	9	37.50	5	20.00	21	27.63	
$40,000-$60,000	2	7.41	1	4.17	4	16.00	7	9.21	
Over $60,000	11	40.74	4	16.67	10	40.00	25	32.89	

Note: - = not reported in accordance with FERPA guidelines ^{A} = significantly different than Autism Only setting, ^{M} = significantly different than Mixed Disability setting, ^{I} = significantly different than Inclusion setting, ABAS = Adaptive Behavior Assessment System-2nd Edition, ADOS = Autism Diagnostic Observation Schedule- 2nd Edition, EI = early intervention, DPA = Developmental Play Assessment, MSEL = Mullen Scales of Early Learning,

* = group difference $p < .05$, ** = group difference $p < .01$, *** = group difference $p < .001$.

Table 2

Teacher/EI provider Characteristics

		n	%
Gender	Female	33	78.6
	Male	1	2.4
	Not reported	8	19.0
Race/Ethnicity	White	25	59.5
	Black	9	21.4
	Other	4	9.5
	Not reported	4	9.5
Current role	Special education preschool teacher	27	64.3
	General education preschool teacher	6	14.3
	Special education preschool teacher	4	9.5
	One-to-one therapist	2	4.8
	General education teacher assistant/aide	1	2.4
	Not reported/Other	2	4.8
Setting (past year)^	Autism support classroom	14	33.3
	Mixed-disability classroom	16	38.1
	Inclusion classroom	4	9.5
	Reverse mainstream classroom	5	11.9
	Daycare	1	2.4
	Early intervention	1	2.4
	Therapy-based program	1	2.4
	Highly structured preschool program- blended	1	2.4
	Office	1	2.4
Education	Some college or Vocational/Associates degree	6	14.3
	College	16	38.1
	Graduate/Professional	17	40.5
	Other	2	4.8
	Not reported	1	2.4
Certifications^^	General Education	18	42.9
	Special Education	20	47.6
	Early Childhood Education	16	38.1
	Early Childhood Special Education	9	21.4
	Speech/Language Therapy	1	2.4

		n	%
	School Psychology	1	2.4
	Elementary Education	2	4.8
	Child Development Associate	1	2.4
	Autism Endorsement	1	2.4
Autism training^^^	None	5	11.9
	A couple workshops or courses	5	11.9
	Some workshops or courses	7	16.7
	Regular Autism training	19	45.2
	Certification or degree related to autism training	7	16.7
		M	*SD*
Years in current position		4.6	7.0
Age		34.2	11.2

Note: ^Percentages do not add up to 100% because some providers selected multiple settings, ^^Percentages do not add up to 100% because some providers held multiple certifications,

^^^Percentages do not add up to 100% because some providers selected multiple levels of autism training.

Table 3

Unadjusted and Adjusted Regressions on MSEL ELC Change Score

Variable	Unadjusted models		Adjusted model	
	Estimate	*p*	Estimate	*p*
Setting Characteristics				
Setting (Inclusion as reference group)				
Mixed Disability	-2.20	.39	-1.32	.65
Autism Only	-2.00	.45	-0.88	.78
EPR Social/Peer Relationships scale	0.17	.14	4.40	.01
EPR Classroom Structure scale	-0.09	.45		
EPR Classroom Environment scale	0.09	.43		
EPR Curriculum and Instruction scale	-0.03	.77		
EPR Positive Instructional Climate scale	-0.01	.92		
Complete EPR teacher interview	0.05	.64		
Change setting	0.10	.40		
Intervention duration	-0.13	.24		
Child and Family Characteristics				
MSEL ELC Time 1	-0.18	.12	-0.11	.43
MSEL Receptive Language Time 1	-0.16	.15	0.002	.99
MSEL Expressive Language Time 1	-0.06	.60		
ABAS General Adaptive Composite Time 1	0.06	.63		
ABAS Social Composite Time 1	-0.002	.99		
SCQ	0.07	.57		
ADOS Severity Score	0.004	.97		
DPA Object Interest	0.03	.80		
Female	0.07	.54		
Time 1 age	0.02	.88		
Age at ASD diagnosis	-0.03	.79		
Participate in Birth – 3 EI services	0.06	.64		
Maternal education	0.08	.51		
Paternal education	-0.11	.36		
Household income	0.34	.77		
Race/Ethnicity (Black as reference group)				
Caucasian/White	-0.80	.77	0.18	.95
Hispanic	-1.03	.75	-2.71	.41
Mixed	-3.75	.23	-2.80	.39
Other	7.19	.20	6.36	.26

Note. ABAS = Adaptive Behavior Assessment System-2nd Edition, ADOS = Autism Diagnostic Observation Schedule- 2nd Edition, DPA = Developmental Play Assessment, EI = early intervention, ELC = Early Learning Composite, EPR = Educational Program Review, MSEL = Mullen Scales of Early Learning SCQ = Social Communication Questionnaire.

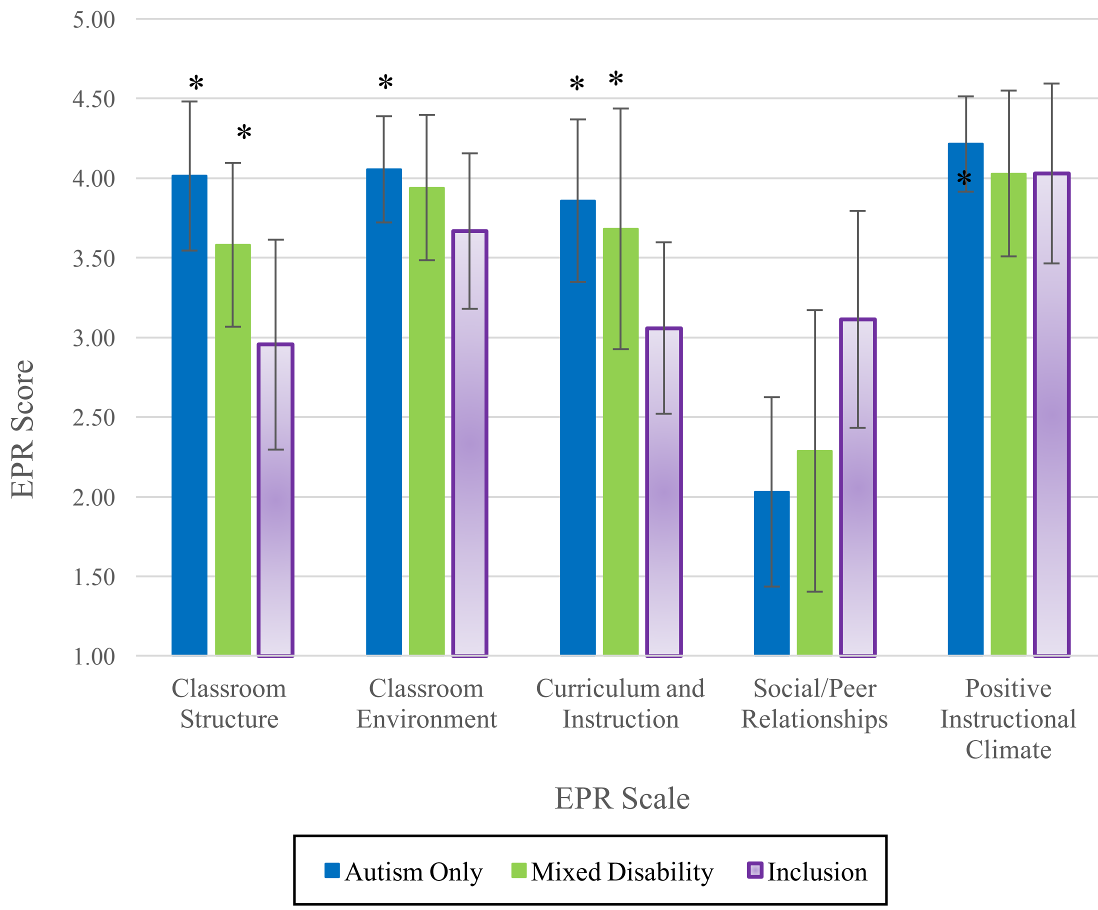

Figure 1. Use of recommended practices by setting type. EPR = Educational Program Review.

* = $p < .05$

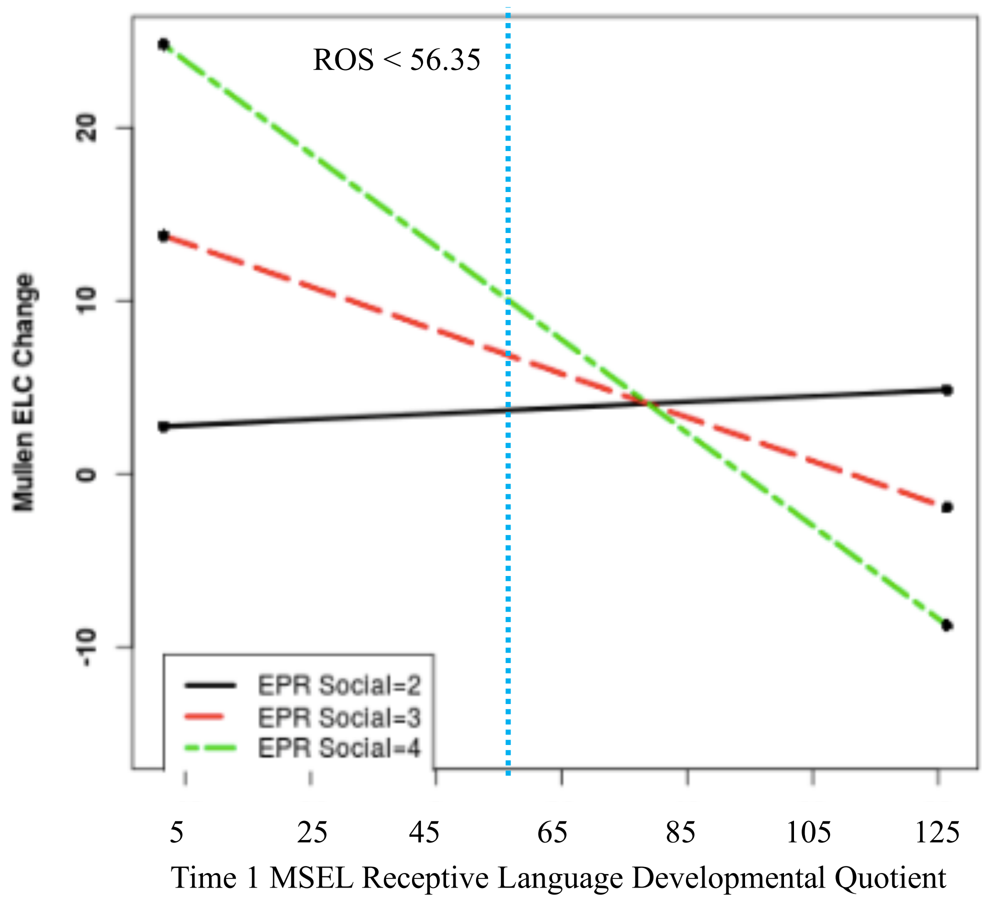

Figure 2. Baseline receptive language moderates the association between the support of social and peer relationships and cognitive changes. Only the Region of Significance below a MSEL Receptive Language Developmental Quotient value of 56.35 is interpretable. ELC = Early Learning Composite, EPR Social = Educational Program Review Social/Peer Relationships scale (range: 1 = "Minimal/no implementation," 3 = "Partial Implementation," to 5= "Full implementation"), MSEL = Mullen Scales of Early Learning, ROS = Region of significance.

Appendix A

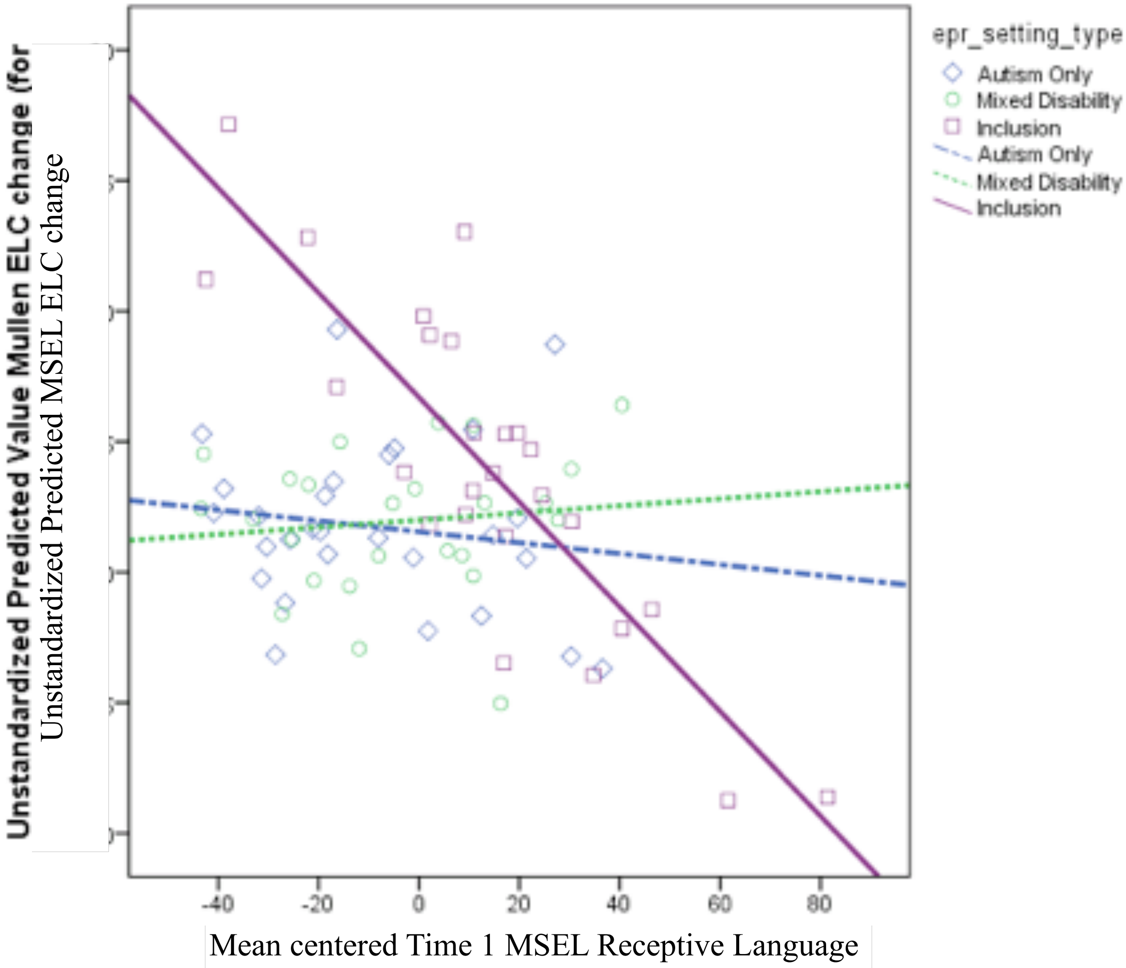

Figure 1A. Moderation of the association between setting type and cognitive Gains by baseline receptive language. Regions of significance indicate that simple slopes are significant at mean centered values of MSEL Receptive Language below -55.9 and above 388.9, which as seen above, are outside the data range. Therefore this interaction is not able to be interpreted.

Note: EPR = Educational Program Review, MSEL = Mullen Scales of Early Learning

Appendix B

Table 1B

Adjusted Regression (Including On-task Behavior Covariate) on MSEL ELC Change Score

Variable	Adjusted model	
	Estimate	*p*
Setting Characteristics		
Setting (Inclusion as reference group)		
Mixed Disability	-1.81	.53
Autism Only	-2.28	.44
EPR Social/Peer Relationships scale	4.83	<.01
Child and Family Characteristics		
MSEL Early Learning Composite Time 1	-0.15	.29
MSEL Receptive Language Time 1	-0.03	.77
Race/Ethnicity (Black as reference group)		
Caucasian/White	1.40	.60
Hispanic	2.12	.53
Mixed	-4.00	.20
Other	11.39	.03
Interaction		
EPR Social/Peer Relationships X Time 1 MSEL Receptive Language	-0.13	.02
Percentage of on-task behavior during EI observation	11.94	.01

Note. EI = Early Intervention ELC = Early Learning Composite, EPR = Educational Program Review, MSEL = Mullen Scales of Early Learning

CHAPTER 3: Preliminary Validation of an Executive Functioning Battery with Low Language Demands for Preschoolers with Autism Spectrum Disorder

Abstract

This study validates the use of portable non-verbal direct assessment measures of executive functioning (EF) skills in preschoolers with Autism Spectrum Disorder (ASD). Sixty-seven preschoolers with ASD (mean age 50.7 months) and low language abilities (mean verbal age equivalent 22.2 months) completed six EF tasks that assess core EF domains (i.e., updating, set-shifting, inhibition) that did not require verbal responses. Feasibility, test-retest reliability, and validity were examined for each task. The Spatial Reversal and Leiter-3 Forward Memory tasks, assessing set-shifting and updating, respectively, demonstrated the most promising validation results, with evidence of adequate feasibility, reliability, and convergent and divergent validity. One inhibition task, Tongue Task, demonstrated excellent reliability, while the other, Balance Beam, did not. Two common standardized measures of EF (i.e., NEPSY-II Statue task and Leiter-3 Reverse Memory) were not valid in this sample, with over 60 percent of participants unable to complete or achieve a non-zero score on the tasks. Implications for research and practice are discussed.

Keywords: Autism Spectrum Disorder, preschool, executive functioning

Preliminary Validation of an Executive Functioning Battery with Low Language Demands for Preschoolers with Autism Spectrum Disorder

The purpose of this study was to validate an executive functioning (EF) battery that does not require verbal responses in a sample of young children with autism spectrum disorder (ASD).

EF refers to a set of cognitive processes that regulate thoughts and emotions into socially appropriate goal-directed behavior. Core processes that comprise EF include updating (constant monitoring and rapid addition/deletion of working memory contents), inhibition (deliberate overriding of dominant or prepotent responses), and set-shifting/cognitive flexibility (switching flexibly between tasks or mental sets) (Miyake et al., 2000; Miyake & Friedman, 2012). EF can be broken down into 'cool' (emotion-independent; e.g., working memory, and set-shifting) and 'hot' (emotion laden; e.g., emotion regulation) domains (Hongwanishkul, Happaney, Lee, & Zelazo, 2005; Roiser & Sahakian, 2013; Zelazo & Carlson, 2012).

Previous research has shown that EF develops in infancy (Diamond, 1990) and is an important predictor of social skills, school success, and later life outcomes (Ayduk et al., 2000; Best, Miller, & Naglieri, 2011; Mischel, Shoda, & Rodriguez, 1989; Mischel et al., 2011; Moffitt et al., 2011; Riggs, Jahromi, Razza, Dillworth-Bart, & Mueller, 2006; Shoda, Mischel, & Peake, 1990). For example, EF has been shown to predict social skills like joint attention and theory of mind cross-culturally in typically developing children (Carlson, Moses, & Breton, 2002; Carlson, Moses, & Claxton, 2004; Carlson, Moses, & Hix, 1998; Sabbagh, Moses, & Shiverick, 2006; Sabbagh, Xu, Carlson, Moses, & Lee, 2006). Further, interventions targeting EF in preschoolers can lead to improvements in pre-academic skills. For example, EF training with a play-based curriculum (vs. a control literacy curriculum) for one year led to higher scores on both traditional EF testing and academic readiness scores (Diamond, Barnett, Thomas, & Munro, 2007). A large study of a Head Start-based program targeting self-regulation (Chicago School Readiness Project) compared to traditional Head Start programming, also yielded significant improvement in self-regulation and pre-academic vocabulary, letter naming, and math skills in low-income preschoolers (Raver et al., 2011). The gains in pre-academic skills across the academic year were mediated by children's EF and attention/impulsivity skills, highlighting the important role that EF can play in academic success for preschoolers.

Children with ASD show a wide range of EF skills during the preschool years (Griffith, Pennington, Wehner, & Rogers, 1999; Pellicano, 2012; Yerys, Hepburn, Pennington, & Rogers, 2007). Similar to typical development, EF has been shown to predict social skills like joint attention and theory of mind in young children with ASD (Griffith et al., 1999; Pellicano, 2007, 2012). EF has also been shown to be related to adaptive behavior (Gilotty, Kenworthy, Sirian, Black, & Wagner, 2002) and repetitive behavior (D'Cruz et al., 2013; Lopez, Lincoln, Ozonoff, & Lai, 2005; Reed, Watts, & Truzoli, 2011; South, Ozonoff, & McMahon, 2007; Yerys et al., 2009) in children with ASD. In addition, a recent study of preschoolers with ASD that had average cognitive and language abilities demonstrated associations between pre-academic skills and EF skills, especially inhibitory control and updating/working memory (Pellicano et al., 2017). Given EF's role in the developmental of social skills and academic success in typical development, researchers have called for its consideration as a potential moderator or mediator of gains in intervention for children with ASD (Pellicano, 2012).

However, there are number of challenges in studying EF among young children with ASD. Edgin and colleagues (2010) identified several important challenges in the assessment of persons with Down Syndrome and other intellectual disabilities that are also applicable to the study of young children with ASD. We aim to begin to address with this study: 1) floor effects, 2) language ability, 3) reproducibility, lack of validation of measures in populations with developmental disabilities, 4) sensitivity of the measures to detect effects, 5) flexibility of use, and 6) assessment variability due to behavior and cooperation.

Floor effects and language ability are two of the major measurement challenges in studying EF among young children with ASD, especially among the substantial percentage of young children with ASD who are minimally verbal (Kasari, Brady, Lord, & Tager-Flusberg, 2013). The most common, well-validated instruments of EF in this age range (e.g., NEPSY-II (Korkman, Krik, & Kemp, 2007), NIH toolbox (Zelazo et al., 2013), and Shape School (Espy, 1997)) often have strong receptive and sometimes expressive language demands. Measuring EF with these measures confounds EF with verbal ability, making it difficult to estimate EF's independent role in predicting outcomes. To our knowledge, only one study, published after the initiation of this project, has attempted to explore the potential relationship between verbal ability and EF, and suggests that language ability may moderate the relationship between EF and later play skills for preschoolers with ASD (Faja et al., 2016). In addition, as large percentage of children with ASD identified in the preschool age range are likely to have language delays, they therefore may not have been represented in prior work. Thus, it is important to validate EF measures that remove expressive language demands and minimize receptive language demands as much as possible to capture the full range of functioning of preschoolers with ASD. Edgin and colleagues (2010) recommends utilizing measures that are primarily nonverbal, with nonverbal responses, and with a low floor.

Although EF has been studied in ASD with a variety of measures (Kenworthy, Yerys, Anthony, & Wallace, 2008; Wallace et al., 2016), there have been few measures validated with preschoolers with ASD. Most research thus far has focused on the ways in which EF in children with ASD compares to other populations (e.g., typically developing peers). However floor and ceiling effects on performance-based EF measures likely contribute to the masking of potential group differences (Hill, 2004). These effects also limit the sensitivity of measures to detect effects both within samples of children with ASD and between children with ASD and other populations. It's possible that some measures subject to floor effects may be sensitive for children above some minimum developmental level, but this has not been explicitly tested in young children with ASD. Therefore, it is important to not only examine absolute floor effects for the group, but to explore if children at floor share common features (e.g., very low receptive language or nonverbal reasoning) that might suggest a baseline of developmental functioning. There also have not been validations of the reliability and temporal stability of performance-based EF measures in young children with ASD, which could contribute to inconsistencies in results. Edgin and colleagues (2010) recommends collecting test-retest reliability estimates that are sample-specific, as well as using measures that demonstrate concurrent validity and have been documented to show differences between populations or be impaired in past literature.

A third limitation of previous work in children with ASD is that it has been all completed in university-based laboratory settings, which is also often accompanied with a convenience sample approach. In order to complete large studies to better approximate the population, measures that can be used flexibility in a variety of settings (e.g., home, school, and clinic) is warranted into order to reduce barriers to participation. Edgin and colleagues (2010) highlighted the need for test batteries for persons with developmental disabilities to be adaptable across a variety of contexts and cultures in order to be utilized for large genetic studies. In addition, it is important to know how EF measurement translates and is reliable when captured in the field, so that EF measurement can be incorporated into large-scale field-based treatment studies. In addition, given EF's associations with academic success, measures that are reliable and valid in educational settings could be beneficial as part of children's educational planning.

Another limitation of most studies of EF in young children with ASD is that they have relied on a single method (direct/performance-based measure or parent report). Single method assessment is limited in that it can be subject to significant bias. For example, direct measure is subject to child's current 'state' bias and situation variation, and parent and teacher report may be influenced by functioning in non-assessed domains and differing expectations and experiences (Burchinal & Cryer, 2003; Smith-Donald, Raver, Hayes, & Richardson, 2007). Kenworthy et al. (2008) also highlight the importance of utilizing ecologically valid measures of executive functioning that relate to real-world applications of EF skills.

To address these limitations, the present study assessed the reliability and validity of brief, easy-to-use field measures of EF among young children with ASD, utilizing the guidelines for test battery validation developed by Edgin and colleagues (2010). We used a variety of EF measures purported to assess different domains of EF and previously used in other research that did not require verbal responses from participants. We also assessed the application of EF in education settings via teacher report and direct observation, and attempted to capture any effects of a "state" bias through independent observations.

Findings from the present study will improve measurement of EF in young children with ASD by establishing valid measures and a minimum developmental level for the measures. This validated battery could then allow for further study of the developmental trajectory of EF from early childhood into school age, as well as any role EF may play in treatment outcomes for young children with ASD.

Methods

Participants

Participants were enrolled consecutively from a larger study of preschool early intervention services for children with ASD in a large urban district and includes measures not reported here. Children were eligible to enroll in the larger study if they: 1) were between 36 and 59 months of age; 2) had a documented diagnosis of ASD; and 3) were receiving services through the public preschool EI or behavioral health system.

Exclusion criteria were: 1) caregivers did not speak English or 2) caregiver or EI provider did not consent to participate. The sample in the larger study was recruited on a rolling basis through the preschool early intervention system. All procedures were approved by the Institutional Review Boards at the University of Pennsylvania and the Children's Hospital of Philadelphia.

Measures

Performance-based non-verbal executive function (EF) battery. The EF battery consists of six tasks to assess the core hot and cool EF processes (i.e., set-shifting/cognitive flexibility, working memory, inhibition, and emotion regulation) that do not require a verbal response from participants and have minimal receptive language demands. Tasks were also selected to have a short duration, to reduce the burden on the child and disruption in the educational settings, such that the entire battery could be completed in under 20 minutes. Tasks were also selected that were inexpensive, portable, and were expected to be feasible to administer in a variety of settings (e.g., preschools, homes). Tasks were selected that have been previously utilized with young children with ASD in lab settings (i.e., Spatial Reversal), utilized in previous field-based preschool intervention trials in children without ASD (i.e., Balance Beam and Tongue Task), and were from well-regarded standardized neuropsychological assessments of executive functioning with norms for preschoolers (i.e., NEPSY-II Statue task and Leiter-3 Forward and Reverse Memory). The tasks also tap all three cool EF domains (i.e., set shifting, updating/working memory, and 'cool' inhibition) and the hot EF domain (i.e., 'hot' inhibition). The order of task administration was counterbalanced across participants.

Tasks were administered by research assistants, a post-doctoral fellow, and a PhD-level psychologist with extensive experience with young children with ASD.

Spatial Reversal. The Spatial Reversal task (Kaufmann, Leckman, & Ort, 1989) assesses cognitive flexibility and perseveration when learning a new rule. This task was administered as described in previous research with young children with ASD (Griffith et al., 1999; McEvoy, Rogers, & Pennington, 1993; Yerys et al., 2007). This task challenges children to maintain the previous location of a reward and to flexibility shift reward association when their response does not yield the reward. In this task, the assessor sits first allows the child to pick a toy that will serve as the reward. Then the assessor places a screen in between themselves and the child, and hides the reward behind the screen in one of two containers placed to the right and left of the child's midline. Then the screen is removed and the child sees the two containers (the reward is under both cups on the first trial). After the child finds the reward in on one side, the assessor continues to hide the reward at that location until the child achieved a set (i.e., four consecutive correct searches). Then the reward location is switched without a cue. The child's responses were coded as correct (i.e., finding the reward) or incorrect (i.e., not finding the reward). After the first switch, the child's responses were coded as correct (i.e., adjusting to the change and choosing the correct container after feedback that the reward was not under the selected container), failure to maintain set errors (i.e., the child switched locations before completing a set of four) or perseverative (i.e., the child searched the same location after receiving feedback on the previous trial that the location of the reward had changed). Each child received a total of 23 trials, and therefore had the opportunity to make four switches. Raw scores of correct searches, perseverative responses, sets achieved, and failures to maintain set were utilized in analyses.

Statue. The Statue task the only subtest in of the Attention and Executive Functioning domain of the NEPSY-II (Korkman et al., 2007) valid for preschool-aged children. The NEPSY-II is a comprehensive standardized neuropsychological battery designed for assessing neurocognitive abilities, including executive functioning, in preschoolers, children, and adolescents. The Statute task assesses motor persistence and inhibition. The child is asked to maintain a body position with eyes closed during a 75-second period and to inhibit the impulse to respond to sound distractors (e.g., dropping pencil, coughing, knocking on table). Observations are made every 5 seconds for the presence of body movement, eye opening, and vocalizations. A score of 2 is recorded for each 5-second interval in which there is no movement, eye-opening, or talking, and a score of 1 is recorded for each interval in which there is one type of error. It demonstrates excellent reliability a normative sample of three-to-six year olds (internal consistency r = .82 -.88, test-retest r = .82). The total raw score (maximum raw score = 30) was utilized in analyses, higher scores indicate better inhibition skills.

Forward Memory and Reverse Memory. The Forward Memory (FM) and Reverse Memory (RM) tasks are taken from the Leiter International Performance Scale – 3 (Roid, Miller, Pomplum, & Koch, 2014). They assess visual spatial updating skills by asking children to remember and identify pictures in a sequential order. It is similar to the Corsi tapping task, but is normed for preschoolers, unlike the Corsi (Farrell Pagulayan, Busch, Medina, Bartok, & Krikorian, 2006). On the Forward Memory task, a child’s immediate retention memory span is measured by the child pointing to pictures of common objects in the order in which they were pointed to by the assessor. On the Reverse Memory task, the child is asked to point to pictures of common objects in the reverse order that they were pointed to by the examiner, and therefore requires use of more working memory than the forward condition. The Forward Memory scale demonstrates excellent reliability in a normative sample (internal consistency Cronbach’s alpha = .84 for two-to-six-year-old children, test-retest r = .83). The Reverse Memory scale also demonstrates excellent internal consistency in a normative sample of two-to-six-year-old children (Cronbach’s alpha range from .77 to .85); however, test-retest reliability was not examined in the standardization sample. Both scales report raw scores (maximum FM raw score = 28, maximum RM raw score = 23) and scaled scores (*Mean* = 10, *SD* = 3). Higher scores indicate better updating skills.

Balance Beam. The Balance Beam task (Murray & Kochanska, 2002) assesses ‘cool’ inhibition and effortful control, as the child has to suppress a dominant response in order to initiate a subdominant response (i.e., slowing down motor activity) in an emotionally neutral situation. The task was administered as described for a previous field-based preschool intervention study (Raver et al., 2011; Smith-Donald et al., 2007). In this task, the time it takes a child to walk on a masking tape line is recorded. First the child is told to walk on the “balance beam,” then the child is told to walk as slowly as he or she can on the same line for two trials. The average difference between the slow and regular trials in seconds was utilized in analyses.

Tongue Task. The Tongue Task (Murray & Kochanska, 2002) assesses 'hot' inhibition and effortful control, as the child has to delay gratification. This task was also administered as described for a previous field-based preschool intervention study (Raver et al., 2011; Smith-Donald et al., 2007). In this task, the child selects a candy (i.e., M&M or goldfish cracker), places it on their tongue, and withhold from eating it until cued by the assessor. The time in seconds until the child eats the candy, up to 40 seconds, was utilized in analyses.

Classroom-based executive functioning measures.

Classroom observation. Live coding of children's on and off task behavior for 10 minutes during their typical early intervention/classroom services was conducted by assessors trained to 80% agreement with an expert coder. Interrater reliability for the live classroom observations of on-task behavior based on 10 participants (including participants from the larger study) was excellent (ICC (1,1) = 0.99) (Fleiss, 1986). The percentage of intervals the child was observed to be demonstrating on-task behavior used in analyses.

BRIEF-P. The Behavior Rating Inventory of Executive Function-Preschool (BRIEF-P) assesses the teacher's perception of a child's broad EF in real world settings (Gioia, Espy, & Isquith, 1996) and has been previously used with preschoolers with ASD (Isquith, Gioia, & Espy, 2004; Smithson et al., 2013). It has the five following subscales that capture the three core EF domains: Inhibit, Shift, Emotional Control (EC), Working Memory, and Plan/Organize. The Inhibit and Emotional Control scales comprise the Inhibitory Self-Control Index (ISCI), the Shift and Emotional Control scales comprise the

Flexibility Index (FI), and the Emergent Metacognition Index (EMI) is comprised of the Working Memory and Plan/Organize scales. The BRIEF-P is a standardized measure with norms based on age and gender, with t-scores a mean of 50 and a standard deviation of 10. Higher scores indicate greater executive functioning impairment. It demonstrates appropriate temporal stability for teacher ratings (Pearson correlations range from .65 to .94 with a mean of 4.2 weeks between ratings). The internal consistency for all scales, indices, and composite went excellent for both the standardization sample (Cronbach's alpha ranging from .90 - .97) and the current sample (Cronbach's alpha ranging from .84 - .97).

Child demographic and clinical characteristics. Parents completed a questionnaire reporting socio-demographic information about their child and their household.

Mullen Scales of Early Learning (MSEL). The Mullen Scales of Early Learning (Mullen, 1995) is a standardized, reliable, and valid measure of early cognitive development for children from birth to 68 months old. The Early Learning Composite (ELC) is based on 4 MSEL scales (visual reception (VR), fine motor (FM), expressive language (EL), and receptive language (RL)). The VR scale assess early non-verbal cognitive skills such as pattern recognition, matching, sorting, and memory. The EL scale assesses language production and the RL scale assess language comprehension. The Mullen covers the full age range of our sample. Due to over half of the sample in the larger study having t-scores below 20 on the Receptive and Expressive Language scales at baseline, developmental quotients were calculated by dividing the age equivalent by the child's chronological age in months and multiplying by 100, as has been previously done in other ASD research (e.g., Chawarska, Klin, Paul, Macari, & Volkmar, 2009; Eapen, Črnčec, & Walter, 2013; Kaale, Smith, & Sponheim, 2012). For the ELC the developmental quotient was calculated by averaging the age equivalences of the VR, FM, EL, and RL scales.

Autism Diagnostic Observation Schedule- 2nd Edition (ADOS). The Autism Diagnostic Observation Schedule-2nd Edition (Lord et al., 2012) is a semi-structured play-based assessment considered the gold-standard observational measure for assessing the presence of ASD. The ADOS was administered by a graduate student or post-doctoral fellow in psychology trained to research reliability and supervised by a licensed clinical psychologist. The Calibrated Severity Score was utilized as an indicator of clinician-observed ASD symptoms.

Social Communication Questionnaire (SCQ). The Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003) is a brief parent questionnaire that evaluates the presence of ASD based questions from the Autism Diagnostic Interview-Revised (Lord, Rutter, & Le Couteur, 1994). Total raw score was utilized as an indicator of parent-reported ASD symptoms.

Objectives and Analytic Plan.

Objective 1: Measure variability, acceptability and feasibility. The distribution of children's responses on the proposed EF battery was explored for significant ceiling effects (i.e., more than 5% of the participants received the highest possible score on the measure), non-completion/floor effects (i.e., more than 20% of the children are unable to complete the task or score a zero). Cognitive, language, ASD symptom, age, observed on-task behavior, and teacher-reported emotional control differences between above-floor performers and non-completers/floor performers were compared utilizing t-tests. In addition, to further assess acceptability and feasibility of the EF measure, examiner errors and attrition effects (e.g., teacher fails to complete the BRIEF-P) was explored.

Objective 2: Measure reliability. To measure the temporal stability of the performance-based non-verbal EF measures, one-third of participants were selected to have re-testing within one month of their initial test date. Due to the low completion rates on the Statue task (see results below), it was not included in the re-testing session. Test-retest reliability was measured utilizing Kappa coefficients for categorical variables and intraclass correlations (ICC) for continuous variables. As recommended by Weir (2005), to determine which type of ICC to utilize, a repeated-measures ANOVA was performed and the *F* ratio for the trials effect was examined for significant systemic error. If the systematic error is deemed unimportant, then ICC Type 3,1 (Shrout & Fleiss, 1979) will be used, which has been used in other studies assessing test-retest reliability (e.g., Moessnang et al. (2016)). Kappa and ICC values above .75 were taken to represent excellent reliability and values between 0.4 and .75 were taken to represent fair to good reliability (Fleiss, 1986; Landis & Koch, 1977).

Objective 3: Measure validity. Validity was assessed for tasks that demonstrate adequate variability, feasibility, and reliability in objectives 1 and 2. Convergent validity (i.e., whether measures are associated with expected measures) of the EF battery was first assessed by examining correlations among the performance-based EF measures. Then the associations among the performance-based and classroom-based EF measures were explored. The teacher-report BRIEF-P ratings were expected to reflect trait EF abilities in everyday academic settings, while classroom observations were expected to reflect the child's attentional state on the day of testing. As some of the performance-based EF tasks were drawn from non-verbal cognitive batteries (e.g., Leiter-3), the tasks were also expected to correlate more strongly with the Visual Reception scale of the MSEL than the Fine Motor, Receptive Language, or Expressive Language scales. Predictive validity was explored in a subsample of participants (n = 45) that received follow-up testing with the MSEL a mean of 7.1 months after their EF testing, by examining the associations among EF variables and changes on the MSEL Early Learning Composite. For all correlations, Spearman's Rho were utilized for non-normally distributed measures and Pearson's r were be utilized for normally distributed measures.

Results

Participants

Table 1 presents descriptive information about the 67 preschoolers (mean (*SD*) age in months = 50.7 (7.7)) that participated in this study. Participants were predominately male (79%), black (47%), and had a household income below $40,000 (55%), which is representative of the urban school district the children were recruited from. Participants had very low cognitive and language abilities based on the Mullen Scales of Early Learning (mean (*SD*) age equivalencies for visual reception = 30.6 (10.9), receptive language = 23.0 (11.9), and expressive language = 23.4 (11.6)).

INSERT TABLE 1 HERE

Distribution

Table 2 presents the distribution, including the skewness and kurtosis for each EF scale and measure. The BRIEF-P t-scores and Leiter-3 Forward Memory standard scores were consistent with a normal distribution, however, most of the other measures demonstrated high skewness and/or kurtosis (i.e., skewness > $|0.8|$, kurtosis > $|3.0|$). A sensitivity analysis revealed a similar pattern of results when transformed versions of the non-normally distributed scores were used as when non-parametric statistics were used, so the non-transformed versions are presented for ease of interpretation.

INSERT TABLE 2 HERE

Completion and Floor Effects. Table 3 presents the percentage of children unable to complete each task and the percentage of children who demonstrated floor level of performance. The values for non-completion range from 8.9% on the Balance Beam task to 46.3% on the Leiter-3 Reverse Memory task. These values included children who did not start or fully complete the task due to severe behavioral difficulties, refusal, or failure to complete or receipt of a floor score on a prerequisite task (i.e., participants that failed to complete or acquired a floor score on the Leiter-3 Forward Memory task were not administered the Leiter-3 Reverse Memory Task). Two participants were not able to complete any of the direct assessment EF tasks.

Next floor effects were examined. For each measure the floor was equivalent to: 1) achieving zero sets on the Spatial Reversal task, 2) attaining a raw score of zero on the NEPSY-II Statue task, 3) eating candy in less than one second on the Tongue Task, 4) attaining a raw score of zero on the Leiter-3 Forward Memory Task, 5) attaining a raw score of zero on the Leiter-3 Reverse Memory task, and 6) walking more than seven seconds more slowing on the first Balance Beam trial than on the average of the second and third trials (because this was the lowest score in the initial Balance Beam validation study (Smith-Donald et al., 2007)). Spatial Reversal, Leiter-3 Forward Memory, and Balance Beam had the lowest floor effects, yielding floor performance in less than 5% of the sample that was able to complete each task. Floor performance on the Tongue Task was at 6.3%. The highest rates occurred on the Leiter-3 Reverse Memory (19.4%) and Statue Task (34.7%). Due to this poor performance on Statue task and Reverse Memory, administration of these tasks was discontinued after 17 and 30 participants, respectively, successfully completed the task.

INSERT TABLE 3 HERE

T-tests were completed to explore child characteristics associated with non-completion and floor effects (Table 4). Participants that were unable to complete or demonstrated floor performance on Spatial Reversal had significantly worse emotion regulation skills (based on the BRIEF-P Emotional Control Scale, $t(44) = 2.40$, $p = .02$, $d = 0.99$)) and higher ADOS autism symptom severity scores ($t(64) = 2.92$, $p = .005$, $d = 0.93$) than participants that achieved scores above the floor. Non-completers and floor performers on the Statue task had marginally significantly lower expressive language (based on MSEL Expressive Language Developmental Quotients, $t(21) = -1.94$, $p = .065$, $d = -0.83$) and higher ADOS autism symptom severity scores ($t(21) = 1.89$, $p = .072$, $d = 0.81$). On the Leiter-3 Forward Memory task, non-completers and floor performers were significantly younger ($t(59) = -3.62$, $p = .0006$, $d = -1.02$), had lower MSEL Visual

Reception ($t(59) = -2.85$, $p = .006$, $d = -0.80$), Receptive Language ($t(59) = -4.33$, $p = .00006$, $d = -1.21$), and Expressive Language scores ($t(59) = -2.92$, $p = .005$, $d = -0.82$), and higher ADOS autism symptom severity scores ($t(59) = 2.76$, $p = .008$, $d = 0.77$). On the Leiter-3 Reverse Memory task, non-completers and floor performers had significantly lower MSEL Visual Reception ($t(52) = -4.72$, $p = .00002$, $d = -1.38$), Receptive Language ($t(43.05) = -5.79$, $p < .00001$, $d = -1.5$), and Expressive Language scores ($t(52) = -3.21$, $p = .004$, $d = -0.88$). There was a marginally significant difference on MSEL Expressive Language ($t(65) = -1.85$, $p = .069$, $d = -0.74$) between participants who scored above floor and those that did not on the Balance Beam task. On the Tongue Task, non-completers and floor performers had significantly lower MSEL Receptive ($t(36.89) = 2.93$, $p = .006$, $d = -0.83$) and Expressive Language scores ($t(61) = -2.12$, $p = .04$, $d = -0.55$), and higher observed ASD symptoms ($t(61) = 2.75$, $p = .008$, $d = 0.71$). There was also a marginally significant difference in parent-reported ASD symptoms ($t(56)=1.90$, $p = .063$, $d = 0.51$) for these children. Completers did not differ from non-completers and floor performers in their on-task behavior during classroom observations. In a sensitivity analysis, these results were similar when participants that had classroom observations completed on alternate days were included ($n = 64$).

INSERT TABLE 4 HERE

To provide preliminary guidance regarding for whom these measures would be appropriate, the minimum nonverbal cognitive and comprehension abilities (based on the MSEL Visual Reception and Receptive Language scales), at which 80% of the sample was able to complete and demonstrate above floor performance, was explored. As seen in Table 3, over 80% of the full sample was able to complete and perform above floor on the Spatial Reversal (81.8%) and Balance Beam (89.6%) tasks. This suggests that these tasks are valid down to the lowest nonverbal cognitive (Visual Reception age equivalent of 10 months) and comprehension (Receptive Language age equivalent of 1 month) abilities in our sample. For the Leiter-3 Forward Memory Task, the 80% completion/above floor criterion was reached for children with nonverbal cognitive abilities similar to a 24 month old child (82.6%, n = 46) or receptive language abilities similar to an 11 month old child (83.0%, n = 47). No minimum Visual Reception or Receptive Language scores were identified at which 80% of the sample could complete and perform above floor for the other tasks, without excluding over half of the sample.

Ceiling effects. Table 3 presents the percentage of participants at the ceiling of each measure. The ceiling for each measure was defined as follows: 1) achieving a perfect score on the spatial reversal task, 2) achieving the highest possible raw score on the NEPSY-II Statue Task, 3) waiting for at least 40 seconds to eat the candy on the Tongue Task, 4) achieving the highest possible scaled score on the Leiter-3 Forward Memory Task, 5) achieving the highest possible scaled score on the Leiter-3 Reverse Memory Task, 6) completing the first trial more than 13.5 seconds faster than the average of the other two trials on the Balance Beam task, as this was the highest score achieved in the original validation study (Smith-Donald et al., 2007). Most of the tasks did not demonstrate ceiling effects (i.e., less 2% of participants were at ceiling on the Spatial Reversal, Balance Beam, Leiter-3 Forward Memory, Leiter-3 Reverse Memory, and

NEPSY-II Statute Task). The ceiling effects on the Tongue Task were just above our 5 percent criteria with 7 percent of participants performing at the ceiling level.

Acceptability and Feasibility

Acceptability and feasibility of administering the EF measures in home and school-based setting was also explored. Spatial Reversal, Tongue Task, and Leiter-3 Reverse Memory were not administered to one participant each due to examiner error. Tongue Task was also not able to be administered to one participant due to a food allergy. Classroom observations were completed on a different day than the other EF measures due to scheduling constraints for 18 participants (26.9%), and were not completed for three participants (4.5%) due to examiner and teacher availability. The BRIEF-P was not returned by 15 teachers (22.4%), was not given to 5 teachers (7.5%), and one teacher (1.5%) did complete enough items on the Working Memory scale for it to be scored.

Reliability

Table 5 presents the test-retest reliability for the Spatial Reversal, Leiter-3 Forward and Reverse Memory, Balance Beam, and Tongue tasks for 16 participants (24%). Task completion was reliable at excellent reliability levels (Kappa ≥ .75) for all tasks except for Balance Beam which had fair to good reliability (Kappa = 0.64) (Fleiss, 1986). As seen in Table 5, based on repeated measures ANOVAs, no trials effects were observed, therefore ICC (3,1) was utilized for all test-retest reliability analyses. Spatial Reversal Perseverations, Leiter-3 Forward Memory raw score and scaled score, and Tongue Task time to eat candy demonstrated excellent test-retest reliability (ICC ≥ .75).

Spatial Reversal total correct and Balance Beam average time between trials demonstrated fair to good reliability (ICC = 0.71 and 0.74 respectively).

INSERT TABLE 5 HERE

Validity

Preliminary convergent, divergent, and predictive validity was explored for the scales that demonstrated adequate reliability above .7 (i.e., Spatial Reversal Perseverations, Spatial Reversal Total Correct, and Leiter-3 Forward Memory). The scaled score for Leiter-3 Forward Memory was chosen over the raw score as it adjusted for potential age effects. Results are presented in Table 6.

Convergent validity. First correlations among the direct EF tasks were compared. As seen in Table 6, as predicted, the Spatial Reversal tasks and Leiter-3 Forward Memory were significantly ($p < .05$), and marginally significantly (for Forward Memory and Spatial Reversal Total Correct scales, $\rho = .28$, $p < .07$), associated with each other. However, Tongue Task was not significantly associated with any of the direct assessment EF tasks (all *p*s > .09).

Next preliminary associations between teacher-reported real-world executive functioning and the direct EF tasks were explored. Only the Leiter-3 Forward Memory scale demonstrated the predicted association with the BRIEF-P. Leiter-3 Forward Memory was significantly correlated with the BRIEF-P Emergent Metacognition Index ($r = -.35$, $p < .05$), which indicates that children performed better on the Forward Memory scale also had better teacher-reported memory, planning, and organization skills (i.e., lower scores on the BRIEF-P EMI scores). Leiter-3 Forward Memory was not, however, significantly correlated with the BRIEF-P Working Memory scale (*p* = .11). As the Spatial Reversal task assesses cognitive flexibility and set-shifting, the Spatial Reversal scores were expected to correlate with BRIEF-P Shift and Flexibility scales, however they did not (all *p*s > .6). As Tongue Task measures "hot inhibition" skills, it was expected to correlate with the BRIEF-P Emotional Control and Inhibitory Self-Control scales, however it did not (all *p*s > .8).

Then, in an exploratory analysis, the associations between the classroom observations of on-task behavior and the direct EF measures were tested. This was an attempt to capture potential effects of the child's attentional state on the day of testing, as well as the relationship between direct EF assessment and its application in an educational setting. As seen in Table 5, only Tongue Task was marginally significantly correlated with on-task behavior during the classroom observation (ρ (27) = .38, *p* = .053). In a sensitivity analysis, results were similar, but the magnitude of the correlations were smaller, when participants with classroom observations completed on different day were included in the analysis (for Tongue Task the magnitude of the correlation was much smaller and non-significant (ρ (41) = .03, *p* = .9).

Divergent validity. As seen in Table 6, as expected, no measures were significantly correlated with the MSEL Fine Motor scale (all *p*s > .3). In addition, children's performance on the direct assessment non-verbal EF tasks were not significantly correlated with language as measured by the MSEL (all *p*s > .1).

Predictive validity. Although on-task performance during the classroom observation was significantly associated with overall cognitive changes (based on MSEL

Early Learning Composite scores) over 9 months (ρ (58) = 0.30, p = .02), as seen in Table 6, none of the direct EF measures were (all *p*s > .2).

INSERT TABLE 6 HERE

Discussion

This study contributes to the growing field of research on executive functioning in preschoolers with ASD. Children with ASD with low cognitive and language levels were able to reliably complete performance-based EF measures in field-based settings. These positive feasibility findings may allow for assessment of populations (e.g., low socio-economic status, minimally verbal) often missing from ASD research. When choosing a measure of EF to assess preschoolers with ASD, this study points to several factors for researchers and practitioners to consider.

The Spatial Reversal and Leiter-3 Forward Memory tasks, measures of set-shifting/cognitive flexibility and updating, respectively, demonstrated the most promising validation results. Spatial reversal was feasible and valid to administer with the entire range of functioning present in this sample (age 36 to 67 months, nonverbal cognitive age equivalent 10 to 60 months, receptive language age equivalent 1 to 55 months, expressive language age equivalent 3 to 46 months). Non-completion and floor performance were associated with higher teacher-reported emotion control challenges, which suggests that emotion regulation difficulties may have contributed to children's difficulty completing the Spatial Reversal task. Clinician observed ASD symptoms were also higher for children that were unable to complete or performed at floor on the Spatial Reversal Task than for children who achieved above floor performance. Although the Spatial Reversal task does not include any overt social demands, the social attention skills assessed as part of the ADOS-2 may also include skills needed to succeed on the Spatial Reversal task. This interpretation is supported by previous research that performance on the Spatial Reversal task is associated with joint attention skills in young children with ASD (Griffith et al., 1999; McEvoy et al., 1993) and that adolescents with ASD demonstrated improved performance on a more advanced set-shifting/cognitive flexibility task when the social component was reduced and the task was presented on a computer screen (Ozonoff, 1995). The number of perseverations and total correct response scores demonstrated adequate temporal stability, while the number of sets achieved and failures to maintain set did not. Perseverations and total correct scores demonstrated medium to large correlations with Leiter-3 Forward Memory scaled scores, indicative of convergent validity. As expected, children who demonstrated better updating skills on the Forward Memory task committed fewer perseverative errors on the Spatial Reversal task. Preliminary evidence for divergent validity for the Spatial Reversal task was obtained by the lack of association between the total correct and number of perseverative error scores and the MSEL Fine Motor scale.

The Leiter-3 Forward Memory was valid for children with baseline nonverbal cognitive abilities above 24 months and receptive language abilities above 11 months. Children that achieved above floor performance had better language skills than those that did not, which is consistent the Forward Memory task being considered verbally mediated by the Leiter developers (Roid & Miller, 1997). Similar to the Spatial Reversal task, clinician-rated ASD symptom severity was associated with completion status. As the Forward Memory task requires participants to attend to and imitate which pictures the assessor is pointing to, children with lower social attention and imitation skills (as suggested by higher ADOS-2 Severity Scores) may have had more difficulty completing this task. The Forward Memory task also demonstrated medium convergent associations with the BRIEF-P Emergent Metacognition Index and MSEL Visual Reception scale, both of which include items related to updating/working memory, and the expected a lack of association with the MSEL Fine Motor scale. These results were consistent in a sensitivity analysis when raw scores were used instead of scaled scores.

The Tongue Task, which assesses 'hot' inhibition via a delay of gratification task, demonstrated excellent reliability. However, there was limited variability in performance: 32 percent of the sample was unable to complete the task, 6 percent performed at floor (i.e., eating the candy in less than one second) and 7 percent of the sample performed at ceiling (i.e., waiting at least 40 seconds to each the candy). Adaptations to the task, such as allowances for longer trials to reduce potential ceiling effects, additional teaching trials and visual supports to improve compliance, and more snack options to address children with allergies/food selectivity may be warranted. Although receptive language was lower was for non-completers/floor performers than above floor performers, no minimum receptive language (or visual reception) age equivalent was found at which 80% of the sample completed and performed above floor on the Tongue Task. Children that were able to complete and achieve above floor scores on the Tongue Task also had lower levels of ASD symptoms, which suggests that social attention and imitation skills may have also impacted children's abilities to complete task.

Performance on the Tongue Task was correlated with observed on-task behavior in educational settings at a marginally significant level. This finding can be interpreted one of two ways. It may suggest that the emotion regulation skills applied in waiting to eat a candy are related to the skills required to pay attention during early intervention services. On the other hand, it could also suggest that performance on the Tongue Task may be more sensitive to 'state' effects of the child's attention and behavior on the day of testing. Further research is need to differentiate these possible explanations, although that the results from the sensitivity analysis not limited to classroom observations from the same day had a smaller, non-significant association lends some support to the 'state' explanation.

Although most participants completed the Balance Beam task (91.1%), which assesses 'cool' inhibition, the test-retest reliability of completion was inadequate (Kappa = .64). In addition, the Balance Beam task had one of the highest language comprehension demands, as children were required understand the meaning of the word "slow." Based on clinical observation, participants may not have understood the directions, as 56.7 percent of the children walked faster on average on the "slow" trials than the "regular" trial. Participants in this study only slow down an average of 0.4 seconds, while participants in the validation study (Smith-Donald et al., 2007) slowed down significantly more (mean 2.0 seconds, $p < .01$, $d = -0.4$). Adaptations to this task may be warranted when considering its use with children with ASD, such as additions of teaching trials, modeling, and comprehension checks participants understanding of the concept "slow."

This study indicates that the two other standardized EF measures, NEPSY-II Statute task and Leiter-3 Reverse Memory, are not valid for preschoolers with ASD. The majority (over 60%) of participants were unable to complete or performed at floor on these tasks. In addition, the Reverse Memory scale demonstrated poor test-retest reliability (ICC = .2). Administration of the Statute task was stopped prior to the initiation of reliability testing due to participants' poor performance, so information about the test-reliability of this task was not assessed. Researchers and clinicians should be cautious about utilizing these measures with preschoolers with ASD, especially with those with low cognitive and language skills.

Although for many of the tasks children that achieved non-zero scores had better language skills than non-completers and floor performers, performance on the Spatial Reversal, Forward Memory, and Tongue Task were not significantly associated with language ability as measured by the MSEL. These results suggest that if the child successfully completes the task, then performance on these tasks is largely independent of language abilities.

The performance-based EF measures evaluated in this study were generally not significantly associated with teacher-reported EF on the BRIEF-P. Mahone and Hoffman (2007) found similar results in preschoolers diagnosed with Attention Deficit Hyperactivity Disorder: parent ratings on the BRIEF-P has consistently low correlations with performance-based measures of EF. Similar findings when researchers have used other versions of the BRIEF led Mahone and Hoffman (2007), as well as others (e.g., Anderson, Anderson, Northam, Jacobs, & Mikiewicz, 2002; Toplak, West, & Stanovich, 2013; Vriezen & Pigott, 2002), to argue that the BRIEF may be capturing different aspects of executive functioning than performance-based EF measures. These results highlight the importance of multi-method and multi-informant assessment to fully capture the range of children's EF skills. Investigations into latent EF constructs among performance-based and reported EF measures in young children with ASD and their relationship to real-world and longitudinal outcomes are warranted.

Only the classroom observation of on task behavior, but none of the performance-based EF measures, predicted cognitive gains in community-based preschool early intervention. One possible explanation for this finding is that the classroom observation measure may capture both child and intervention factors important for learning in preschool educational environments (e.g., both the child's ability to attend and the teacher's ability to capture and sustain the child's attention), while the EF tasks only capture child's ability. In addition, other outcome measures may be more sensitive to baseline executive functioning levels. Future directions include exploring the relationship among the most promising EF task and other concurrent and longitudinal measures that have previously been associated with EF skills in children with ASD (e.g., academic readiness skills, joint attention, play skills).

This study had several limitations that warrant mention. This study would have benefited from concurrent observer ratings of self-regulation and noncompliance during the performance-based EF tasks to better capture the child's emotional 'state' on the day of testing that may have contributed to task completion and performance. The BRIEF-P emotional control scale and the classroom observation measure were likely poor proxies for this information. As this study was added on to an existing longitudinal study, due to participant's timing in the larger study we were unable to collect follow-up data for all participants, which led to a greatly reduced sample for the predictive validity analyses. The low return of BRIEF-P by teachers is an additional limitation. In addition, all families elected to have their assessments completed in homes, schools, or daycares, so we were unable to compare task completion and performance in clinic vs. field-based settings.

Despite these limitations, these results provide preliminary support for nonverbal measures of executive functioning that can be used reliability and validly with preschoolers with ASD, including those with low language levels, in field-based research.

References

Anderson, V. A., Anderson, P., Northam, E., Jacobs, R., & Mikiewicz, O. (2002). Relationships Between Cognitive and Behavioral Measures of Executive Function in Children With Brain Disease. *Child Neuropsychology*, *8*(4), 231–240. https://doi.org/10.1076/chin.8.4.231.13509

Ayduk, O., Mendoza-Denton, R., Mischel, W., Downey, G., Peake, P. K., & Rodriguez, M. (2000). Regulating the interpersonal self: Strategic self-regulation for coping with rejection sensitivity. *Journal of Personality and Social Psychology*, *79*(5), 776–792. https://doi.org/10.1037//0022-3514.79.5.776

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, *21*(4), 327–336. https://doi.org/10.1016/j.lindif.2011.01.007

Burchinal, M. R., & Cryer, D. (2003). Diversity, child care quality, and developmental outcomes. *Early Childhood Research Quarterly*, *18*(4), 401–426. https://doi.org/10.1016/j.ecresq.2003.09.003

Carlson, S. M., Moses, L. J., & Breton, C. (2002). How specific is the relation between executive function and theory of mind? Contributions of inhibitory control and working memory. *Infant and Child Development*, *11*(2), 73–92. https://doi.org/10.1002/icd.298

Carlson, S. M., Moses, L. J., & Claxton, L. J. (2004). Individual differences in executive functioning and theory of mind: An investigation of inhibitory control and planning ability. *Journal of Experimental Child Psychology*, *87*(4), 299–319. https://doi.org/10.1016/j.jecp.2004.01.002

Carlson, S. M., Moses, L. J., & Hix, H. R. (1998). The Role of Inhibitory Processes in Young Children's Difficulties with Deception and False Belief. *Child Development*, *69*(3), 672–691. https://doi.org/10.1111/j.1467-8624.1998.tb06236.x

Chawarska, K., Klin, A., Paul, R., Macari, S., & Volkmar, F. (2009). A prospective study of toddlers with ASD: short-term diagnostic and cognitive outcomes. *Journal of Child Psychology and Psychiatry*, *50*(10), 1235–1245. https://doi.org/10.1111/j.1469-7610.2009.02101.x

D'Cruz, A.-M., Ragozzino, M. E., Mosconi, M. W., Shrestha, S., Cook, E. H., & Sweeney, J. A. (2013). Reduced behavioral flexibility in autism spectrum disorders. *Neuropsychology*, *27*(2), 152–160. https://doi.org/10.1037/a0031721

Diamond, A. (1990). Developmental Time Course in Human Infants and Infant Monkeys, and the Neural Bases of, Inhibitory Control in Reachinga. *Annals of the New York Academy of Sciences*, *608*(1), 637–676. https://doi.org/10.1111/j.1749-6632.1990.tb48913.x

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool Program Improves Cognitive Control. *Science (New York, N.Y.)*, *318*(5855), 1387–1388. https://doi.org/10.1126/science.1151148

Eapen, V., Črnčec, R., & Walter, A. (2013). Clinical outcomes of an early intervention program for preschool children with Autism Spectrum Disorder in a community group setting. *BMC Pediatrics*, *13*(1). https://doi.org/10.1186/1471-2431-13-3

Edgin, J. O., Mason, G. M., Allman, M. J., Capone, G. T., DeLeon, I., Maslen, C., … Nadel, L. (2010). Development and validation of the Arizona Cognitive Test Battery for Down syndrome. *Journal of Neurodevelopmental Disorders*, *2*(3), 149–164. https://doi.org/10.1007/s11689-010-9054-3

Espy, K. A. (1997). The shape school: Assessing executive function in preschool children. *Developmental Neuropsychology*, *13*(4), 495–499. https://doi.org/10.1080/87565649709540690

Faja, S., Dawson, G., Sullivan, K., Meltzoff, A. N., Estes, A., & Bernier, R. (2016). Executive function predicts the development of play skills for verbal preschoolers with autism spectrum disorders. *Autism Research*, *9*(12), 1274–1284. https://doi.org/10.1002/aur.1608

Farrell Pagulayan, K., Busch, R. M., Medina, K. L., Bartok, J. A., & Krikorian, R. (2006). Developmental normative data for the Corsi Block-tapping task. *Journal of Clinical and Experimental Neuropsychology*, *28*(6), 1043–1052. https://doi.org/10.1080/13803390500350977

Fleiss, J. L. (1986). *The design and analysis of clinical experiments*. New York: John Wiley.

Gilotty, L., Kenworthy, L., Sirian, L., Black, D. O., & Wagner, A. E. (2002). Adaptive skills and executive function in autism spectrum disorders. *Child Neuropsychology*, *8*(4), 241–8.

Gioia, G. A., Espy, K. A., & Isquith, P. K. (1996). *Behavior Rating Inventory of Executive Function - Preschool Version Professional Manual*. Lutz, FL: Psychological Assessment Resources.

Griffith, E. M., Pennington, B. F., Wehner, E. A., & Rogers, S. J. (1999). Executive Functions in Young Children with Autism. *Child Development*, *70*(4), 817–832. https://doi.org/10.1111/1467-8624.00059

Hill, E. L. (2004). Executive dysfunction in autism. *Trends in Cognitive Sciences*, *8*(1), 26–32. https://doi.org/10.1016/j.tics.2003.11.003

Hongwanishkul, D., Happaney, K. R., Lee, W. S. C., & Zelazo, P. D. (2005). Assessment of hot and cool executive function in young children: age-related changes and individual differences. *Developmental Neuropsychology*, *28*(2), 617–644. https://doi.org/10.1207/s15326942dn2802_4

Isquith, P. K., Gioia, G. A., & Espy, K. A. (2004). Executive function in preschool children: Examination through everyday behavior. *Developmental Neuropsychology*, *26*, 403–422.

Kaale, A., Smith, L., & Sponheim, E. (2012). A randomized controlled trial of preschool-based joint attention intervention for children with autism: Preschool-based JA-intervention for children with autism. *Journal of Child Psychology and Psychiatry*, *53*(1), 97–105. https://doi.org/10.1111/j.1469-7610.2011.02450.x

Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the Minimally Verbal School-Aged Child With Autism Spectrum Disorder. *Autism Research*, *6*(6), 479–493. https://doi.org/10.1002/aur.1334

Kaufmann, P., Leckman, J., & Ort, S. (1989). Delayed-response performance in males with fragile-x syndrome. *Journal of Clinical and Experimental Neuropsychology*, *12*(1), 69–69.

Kenworthy, L., Yerys, B. E., Anthony, L. G., & Wallace, G. L. (2008). Understanding Executive Control in Autism Spectrum Disorders in the Lab and in the Real World. *Neuropsychology Review*, *18*(4), 320–338. https://doi.org/10.1007/s11065-008-9077-7

Korkman, M., Krik, U., & Kemp, S. (2007). *NEPSY-II: A developmental neuropsychological assessment*. San Antonio, TX: The Psychological Corporation.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

Lopez, B. R., Lincoln, A. J., Ozonoff, S., & Lai, Z. (2005). Examining the relationship between executive functions and restricted, repetitive symptoms of Autistic Disorder. *Journal of Autism and Developmental Disorders*, *35*(4), 445–60. https://doi.org/10.1007/s10803-005-5035-x

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). Autism diagnostic observation schedule–2nd edition (ADOS-2). *Los Angeles, CA: Western Psychological Corporation*.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659–685. https://doi.org/10.1007/BF02172145

Mahone, E. M., & Hoffman, J. (2007). Behavior ratings of executive function among preschoolers with ADHD. *The Clinical Neuropsychologist*, *21*(4), 569–586. https://doi.org/10.1080/13854040600762724

McEvoy, R. E., Rogers, S. J., & Pennington, B. F. (1993). Executive Function and Social Communication Deficits in Young Autistic Children. *Journal of Child Psychology and Psychiatry*, *34*(4), 563–578. https://doi.org/10.1111/j.1469-7610.1993.tb01036.x

Mischel, W., Ayduk, O., Berman, M. G., Casey, B. J., Gotlib, I. H., Jonides, J., … Shoda, Y. (2011). “Willpower” over the life span: decomposing self-regulation. *Social Cognitive and Affective Neuroscience*, *6*(2), 252–256. https://doi.org/10.1093/scan/nsq081

Mischel, W., Shoda, Y., & Rodriguez, M. I. (1989). Delay of gratification in children. *Science*, *244*(4907), 933–938. https://doi.org/10.1126/science.2658056

Miyake, A., & Friedman, N. P. (2012). The Nature and Organization of Individual Differences in Executive Functions: Four General Conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Moessnang, C., Schäfer, A., Bilek, E., Roux, P., Otto, K., Baumeister, S., … Tost, H. (2016). Specificity, reliability and sensitivity of social brain responses during spontaneous mentalizing. *Social Cognitive and Affective Neuroscience*, *11*(11), 1687–1697. https://doi.org/10.1093/scan/nsw098

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., … Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, *108*(7), 2693–2698. https://doi.org/10.1073/pnas.1010076108

Mullen, E. M. (1995). *Mullen Scales of Early Learning manual*. Circle Pines, MN: American Guidance Service.

Murray, K. T., & Kochanska, G. (2002). Effortful control: Factor structure and relation to externalizing and internalizing behaviors. *Journal of Abnormal Child Psychology*, *30*(5), 503–514. https://doi.org/10.1023/A:1019821031523

Ozonoff, S. (1995). Reliability and validity of the Wisconsin Card Sorting Test in studies of autism. *Neuropsychology*, *9*(4), 491–500. https://doi.org/10.1037/0894-4105.9.4.491

Pellicano, E. (2007). Links between theory of mind and executive function in young children with autism: Clues to developmental primacy. *Developmental Psychology*, *43*(4), 974–990. https://doi.org/10.1037/0012-1649.43.4.974

Pellicano, E. (2012). The Development of Executive Function in Autism. *Autism Research and Treatment*, *2012*, 1–8. https://doi.org/10.1155/2012/146132

Pellicano, E., Kenny, L., Brede, J., Klaric, E., Lichwa, H., & McMillin, R. (2017). Executive function predicts school readiness in autistic and typical preschool children. *Cognitive Development*, *43*, 1–13. https://doi.org/10.1016/j.cogdev.2017.02.003

Raver, C. C., Jones, S. M., Li-Grining, C., Zhai, F., Bub, K., & Pressler, E. (2011). CSRP's Impact on Low-Income Preschoolers' Preacademic Skills: Self-Regulation as a Mediating Mechanism: CSRP's Impact on Low-Income Preschoolers' Preacademic Skills. *Child Development*, *82*(1), 362–378. https://doi.org/10.1111/j.1467-8624.2010.01561.x

Reed, P., Watts, H., & Truzoli, R. (2011). Flexibility in young people with autism spectrum disorders on a card sort task. *Autism*. https://doi.org/10.1177/1362361311409599

Riggs, N. R., Jahromi, L. B., Razza, R. P., Dillworth-Bart, J. E., & Mueller, U. (2006). Executive function and the promotion of social?emotional competence. *Journal of Applied Developmental Psychology*, *27*(4), 300–309. https://doi.org/10.1016/j.appdev.2006.04.002

Roid, G. H., & Miller, L. (1997). Leiter international performance scale-revised: Examiner's manual. In G. H. Roid & L. J. Miller (Eds.), *Leiter international performance scale-revised*. Wood Dale, IL: Stoelting.

Roid, G. H., Miller, L. J., Pomplum, M., & Koch, C. (2014). *Leiter International Performance Scale - 3*. North Tonawanda, NY: Multi-Health Systems Inc.

Roiser, J. P., & Sahakian, B. J. (2013). Hot and cold cognition in depression. *CNS Spectrums*, *18*(3), 139–149. https://doi.org/10.1017/S1092852913000072

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire: Manual*. Western Psychological Services.

Sabbagh, M. A., Moses, L. J., & Shiverick, S. (2006). Executive Functioning and Preschoolers' Understanding of False Beliefs, False Photographs, and False Signs. *Child Development*, *77*(4), 1034–1049. https://doi.org/10.1111/j.1467-8624.2006.00917.x

Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J., & Lee, K. (2006). The Development of Executive Functioning and Theory of Mind: A Comparison of Chinese and U.S. Preschoolers. *Psychological Science*, *17*(1), 74–81. https://doi.org/10.1111/j.1467-9280.2005.01667.x

Shoda, Y., Mischel, W., & Peake, P. K. (1990). Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions. *Developmental Psychology*, *26*(6), 978–986. https://doi.org/10.1037/0012-1649.26.6.978

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Smith-Donald, R., Raver, C. C., Hayes, T., & Richardson, B. (2007). Preliminary construct and concurrent validity of the Preschool Self-regulation Assessment (PSRA) for field-based research. *Early Childhood Research Quarterly*, *22*(2), 173–187. https://doi.org/10.1016/j.ecresq.2007.01.002

Smithson, P. E., Kenworthy, L., Wills, M. C., Jarrett, M., Atmore, K., & Yerys, B. E. (2013). Real World Executive Control Impairments in Preschoolers with Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *43*(8), 1967–1975. https://doi.org/10.1007/s10803-012-1747-x

South, M., Ozonoff, S., & McMahon, W. M. (2007). The relationship between executive functioning, central coherence, and repetitive behaviors in the high-functioning autism spectrum. *Autism*, *11*(5), 437–51. https://doi.org/10.1177/1362361307079606

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner Review: Do performance-based measures and ratings of executive function assess the same construct? *Journal of Child Psychology and Psychiatry*, *54*(2), 131–143. https://doi.org/10.1111/jcpp.12001

Vriezen, E. R., & Pigott, S. E. (2002). The Relationship Between Parental Report on the BRIEF and Performance-Based Measures of Executive Function in Children with Moderate to Severe Traumatic Brain Injury. *Child Neuropsychology*

(Neuropsychology, Development and Cognition: Section C), *8*(4), 296–303. https://doi.org/10.1076/chin.8.4.296.13505

Wallace, G. L., Yerys, B. E., Peng, C., Dlugi, E., Anthony, L. G., & Kenworthy, L. (2016). Assessment and Treatment of Executive Function Impairments in Autism Spectrum Disorder. *International Review of Research in Developmental Disabilities*, *51*, 85–122. https://doi.org/10.1016/bs.irrdd.2016.07.004

Weir, J. P. (2005). Quantifying Test-Retest Reliability Using the Intraclass Correlation Coefficient and the SEM. *The Journal of Strength and Conditioning Research*, *19*(1), 231. https://doi.org/10.1519/15184.1

Yerys, B. E., Hepburn, S. L., Pennington, B. F., & Rogers, S. J. (2007). Executive Function in Preschoolers with Autism: Evidence Consistent with a Secondary Deficit. *Journal of Autism and Developmental Disorders*, *37*(6), 1068–1079. https://doi.org/10.1007/s10803-006-0250-7

Yerys, B. E., Wallace, G. L., Harrison, B., Celano, M. J., Giedd, J. N., & Kenworthy, L. (2009). Set-shifting in children with autism spectrum disorders: reversal shifting deficits on the Intradimensional/Extradimensional Shift Test correlate with repetitive behaviors. *Autism*, *13*(5), 523–538. https://doi.org/10.1177/1362361309335716

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., & Weintraub, S. (2013). NIH Toolbox Cognition Battery (CB): Measuring Executive Function and Attention. *Monographs of the Society for Research in Child Development*, *78*(4), 16–33. https://doi.org/10.1111/mono.12032

Zelazo, P. D., & Carlson, S. M. (2012). Hot and Cool Executive Function in Childhood and Adolescence: Development and Plasticity. *Child Development Perspectives*, *6*(4), 354–360. https://doi.org/10.1111/j.1750-8606.2012.00246.x

Tables

Table 1

Participant Characteristics

	n	%	Mean	SD	Min	Max
Gender	67					
Male	53	79.1				
Race/Ethnicity	64					
Black	30	46.9				
White	15	23.4				
Mixed race/ethnicity	10	15.6				
Hispanic	8	12.5				
Asian/Pacific Islander	1	1.6				
Household Income	63					
Under $20,000	19	30.2				
$20,000-$40,000	16	25.4				
$40,000-$60,000	5	7.9				
Over $60,000	23	36.5				
Maternal Education	62					
Did not complete High School	2	3.2				
High School or GED	14	22.6				
Vocational/technical school	9	14.5				
Some college	12	19.4				
College degree	16	25.8				
Advanced degree	9	14.5				
Age in months	67		50.7	7.7	36.6	67.5
Mullen Scales of Early Learning						
Visual Reception DQ	67		62.4	19.1	26.7	115.6
Fine Motor DQ	67		60.6	13.8	25.0	100.0
Receptive Language DQ	67		46.8	23.3	1.9	106.8
Expressive Language DQ	67		48.0	23.2	6.4	97.7
Early Learning Composite DQ	67		54.5	17.0	18.8	102.3
ADOS Severity Score	67		6.8	1.9	3.0	10.0
Social Communication Questionnaire	62		18.8	5.8	5.0	32.0

Note: ADOS = Autism Diagnostic Observation Schedule-2, DQ = developmental quotient, MSEL = Mullen Scales of Early Learning

Table 2

Descriptives of Executive Functioning Measures

Measure: Score/Scale	n	Mean	SD	Min	Max	Skewness	Kurtosis
Spatial Reversal	57						
Total Correct		14.7	2.9	4.0	18.0	-1.7	3.8
Perseverations		2.0	2.9	0.0	14.0	2.7	8.7
Sets Achieved		2.5	1.1	0.0	4.0	-0.5	-0.6
Failure to Maintain Set		3.3	1.8	0.0	8.0	0.6	0.1
NEPSY-II Statue	17						
Raw score		3.3	4.4	0.0	12.0	1.0	-0.8
Leiter-3 Forward Memory	45						
Raw Score		4.0	2.6	0.0	12.0	1.4	2.3
Scaled Score		5.8	2.5	0.0	11.0	0.0	0.0
Leiter-3 Reverse Memory	30						
Raw Score		1.0	1.0	0.0	3.0	0.3	-1.4
Scaled Score		4.7	2.4	1.0	9.0	0.1	-1.5
Balance Beam	61						
Average time difference		0.4	4.0	-11.7	22.9	2.9	18.4
Tongue Task	43						
Time to eat candy		10.5	12.5	0.0	40.0	1.5	1.1
Classroom observation	46						
% on-task		0.7	0.2	0.0	1.0	-1.0	1.1
BRIEF-P							
Shift T score	47	62.0	13.1	40.0	91.0	0.2	-0.4
Emotional Control T score	47	64.0	14.1	41.0	89.0	0.0	-1.0
Working Memory T score	46	71.3	11.9	46.0	98.0	-0.1	-0.1
ISCI T score	47	65.1	13.3	44.0	91.0	0.2	-1.0
Flexibility Index T score	47	64.5	13.5	39.0	90.0	0.1	-0.8
EMI T score	46	70.9	12.1	44.0	100.0	0.1	0.0

Note: BRIEF-P = Behavior Rating Inventory of Executive Function-Preschool, EMI = Emergent Metacognition Index, ISCI = Inhibitory Self-Control Index.

Table 3

Completion, Floor, and Ceiling Rates for Performance-based EF Measures

Measure	n	% not completed	% floor	% Ceiling	Min VR AE	Min RL AE
Spatial Reversal	66	13.4	4.5	0.0	10	1
Statute	23	26.1	34.7	0.0	-	-
Forward Memory	61	26.2	3.2	0.0	24	11
Reverse Memory	54	46.3	19.4	0.0	-	-
Balance Beam	67	8.9	1.6	1.6	10	1
Tongue Task	63	31.7	6.3	7.0	-	-

Note: n varies based on introduction of the measure, examiner availability, and administration of Statue and Reverse Memory tasks were discontinued after 17 and 30 participants were able to complete the task respectively due to poor performance; Min VR AE = Minimum Mullen Visual Reception Scale age equivalent for at least 80% of the sample to complete and score above floor on the measure; Min RL AE = Minimum Mullen Receptive Language age equivalent for at least 80% of the sample to complete and score above floor on the measure.

Table 4

Differences Between Above Floor Performers and Non-completers/Floor Performers on Performance-based EF Measures

Measure	Spatial Reversal M/SD		Statue M/SD		Forward Memory M/SD		Reverse Memory M/SD		Balance Beam M/SD		Tongue Task M/SD	
	C	NC	C	NC	C	NC	C	NC	C	NC	C	NC
BRIEF-P EC	62.0/ 14.0	75.3/ 10.2*	64.3/ 18.1	68.1/ 10.7	63.0/ 15.1	65.6/ 13.1	65.6/ 14.0	63.2/ 13.8	63.1/ 14.0	73.5/ 12.8	60.8/ 15.0	67.1/ 13.2
Classroom observation: % on task	0.7/ 0.2	0.6/ 0.3	0.8/ 0.2	0.9/ 0.1	0.7/ 0.2	0.7/ 0.2	0.7/ 0.2	0.7/ 0.2	0.7/ 0.2	0.7/ 0.3	0.7/ 0.2	0.7/ 0.3
Age	47.6/ 6.8	51.4/ 7.8	49.4/ 5.9	49.6/ 8.5	52.9/ 7.4	45.5/ 6.7***	53.2/ 8.0	49.6/ 8.0	50.9/ 48.5	6.9/ 7.7	51.2/ 7.9	49.0/ 7.4
MSEL VR	61.8/ 16.7	60.7/ 15.8	68.5/ 20.4	63.1/ 27.3	65.4/ 17.5	52.3/ 13.4**	74.7/ 17.6	54.4/ 13.3***	62.6/ 19.7	61.2/ 14.8	65.5/ 15.8	58.2/ 23.3
MSEL RL	47.5/ 22.3	41.7/ 27.1	64.9/ 21.6	50.7/ 22.0	52.3/ 19.8	28.4/ 19.5***	61.8/ 13.9	34.6/ 19.8***	48.5/ 23.5	32.8/ 16.2	53.4/ 18.4	35.4/ 26.4**
MSEL EL	48.8/ 22.1	41.6/ 26.9	69.2/ 19.3	49.9/ 25.4+	52.0/ 20.5	34.6/ 22.8**	58.6/ 18.6	39.9/ 22.4**	49.7/ 23.3	32.9/ 17.7+	53.3/ 21.6	40.9/ 23.9*
ADOS SS	6.4/ 1.8	8.1/ 1.7**	5.4/ 2.0	7.1/ 2.0+	6.4/ 1.9	7.7/ 1.5**	6.2/ 1.8	6.9/ 1.8	6.7/ 1.9	7.1/ 1.8	6.3/ 1.9	7.5/ 1.6**
SCQ	18.7/ 6.1	19.6/ 4.3	19.0/ 3.6	20.1/ 6.9	18.7/ 5.8	18.9/ 6.3	17.9/ 5.8	19.5/ 5.6	18.9/ 5.8	17.2/ 6.6	17.6/ 6.1	20.6/ 5.2+

Note: ADOS SS = Autism Diagnostic Observation Schedule-2 Severity Score, BRIEF-P = Behavior Rating Inventory of Executive Function- Preschool, C= Complete and above floor on measure, EC = Emotional Control Scale, EL = Expressive Language

Developmental Quotient, MSEL = Mullen Scales of Early Learning, NC = did not complete or below floor on measure, RL = Receptive Language Developmental Quotient, SCQ = Social Communication Questionnaire, VR = Visual Reception Developmental Quotient.

*** $p < .001$, ** $p < .01$, * $p < .05$, + $p < .08$.

Table 5

Test-retest Reliability of Performance-based EF Measures

Measure	Scale	Trials effect	Reliability type	Reliability value
Spatial Reversal	Completed		Kappa	1.00
	Total correct	F (1,13) = 0.13, p =.72	ICC (3,1)	0.71
	Total sets achieved	F (1,13) = 0.06, p =.82	ICC (3,1)	0.37
	Perseverations	F (1,13) = 0.53, p = .48	ICC (3,1)	0.76
	Total failures to maintain set	F (1,13) = 0.00, p = 1.00	ICC (3,1)	0.55
Tongue Task	Completed		Kappa	0.88
	Time to eat candy	F (1,6) = 1.16, p = .32	ICC (3,1)	0.96
Forward Memory	Completed		Kappa	0.82
	Raw Score	F (1,11) = 1.21, p = .30	ICC (3,1)	0.92
	Scaled Score	F (1,11) = 0.45, p = .52	ICC (3,1)	0.91
Reverse Memory	Completed		Kappa	0.75
	Raw score	F (1, 8) = 2.7, p = .14	ICC (3,1)	0.25
	Scaled score	F (1, 8) = 3.12, p = .12	ICC (3,1)	0.26
Balance beam	Completed		Kappa	0.64
	Average time difference	F (1, 13) = 0.84, p = .38	ICC (3,1)	0.74

Note: ICC = Intraclass correlation.

Table 6

Convergent, Divergent, and Predictive Validity Correlations

	SR: Total Correct	SR: Perseverations	Forward Memory	Tongue Task Time
SR: Total Correct	-	-	-	-
SR: Perseverations	-0.84***	-	-	-
Forward Memory	0.28+	-0.39*	-	-
Tongue Task Time	-0.23	0.28	-0.11	-
BRIEF Shift T score	0.04	0.03	-0.04	-0.13
BRIEF Emotional Control T score	-0.12	0.09	-0.33	-0.04
BRIEF Working Memory T score	0.07	0.11	-0.29	0.03
BRIEF ISCI	-0.11	0.14	-0.33+	-0.04
BRIEF FI	-0.05	0.01	-0.23	-0.09
BRIEF EMI	0.05	0.13	-0.35*	-0.001
Classroom observation: % on task	0.01	0.07	-0.14	0.38+
MSEL VR DQ	-0.02	-0.11	0.29+	0.29
MSEL FM DQ	-0.01	-0.13	0.14	0.17
MSEL RL DQ	0.06	-0.15	0.14	0.18
MSEL EL DQ	0.13	-0.20	0.24	0.18
MSEL ELC DQ change	0.15[a]	-0.06[a]	0.27[b]	0.24[c]

Note: [a] = n = 38, [b] = n = 27, [c] = n = 26, Spearman's Rho utilized for non-normally distributed measures, Pearson's r utilized for normally distributed measures, DQ = Developmental Quotient, EL = Expressive Language, ELC = Early Learning Composite, EMI = Emergent Metacognition Index, FI = Flexibility Index, FM = Fine Motor, ISCI = Inhibitory Self Control Index, MSEL = Mullen Scales of Early Learning, SR= Spatial Reversal, VR = Visual Reception. *** $p < .001$, * $p < .05$, + $p < .08$.