

Psychological Bulletin

Manuscript version of

Project AIM: Autism Intervention Meta-Analysis for Studies of Young Children

Micheal Sandbank, Kristen Bottema-Beutel, Shannon Crowley, Margaret Cassidy, Kacie Dunham, Jacob I. Feldman, Jenna Crank, Susanne A. Albarran, Sweeya Raj, Prachy Mahbub, Tiffany G. Woynaroski

Funded by:

- National Institute on Deafness and Other Communication Disorders
- National Institutes of Health, Eunice Kennedy Shriver National Institute of Child Health and Human Development
- National Institutes of Health, Training Program in Fundamental Neuroscience

© 2019, American Psychological Association. This manuscript is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final version of record is available via its DOI: <https://dx.doi.org/10.1037/bul0000215>

This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Project AIM: Autism Intervention Meta-Analysis for Studies of Young Children

Micheal Sandbank

The University of Texas at Austin, Department of Special Education

Kristen Bottema-Beutel

Boston College, Lynch School of Education and Human Development

Shannon Crowley

Boston College, Lynch School of Education and Human Development

Margaret Cassidy

Vanderbilt University, College of Arts and Sciences

Kacie Dunham

Vanderbilt University, Vanderbilt Brain Institute

Jacob I. Feldman

Vanderbilt University, Department of Hearing and Speech Sciences

Jenna Crank

The University of Texas at Austin, Department of Special Education

Susanne A. Albarran

The University of Texas at Austin, Department of Special Education

Sweeya Raj

Vanderbilt University, College of Arts and Sciences

Prachy Mahbub

Mount Holyoke College, Department of Neuroscience and Behavior

Tiffany G. Woynaroski

Vanderbilt University Medical Center, Vanderbilt Brain Institute, Vanderbilt Kennedy Center

Author Note

Micheal Sandbank, PhD, is an Assistant Professor in the Department of Special Education at The University of Texas at Austin. Kristen Bottema-Beutel, PhD, is an Associate Professor in the Lynch School of Education and Human Development at Boston College. Shannon Crowley is a doctoral student in the Lynch School of Education and Human Development at Boston College. Margaret Cassidy and Sweeya Raj are undergraduate students in the College of Arts and Sciences at Vanderbilt University. Kacie Dunham is a doctoral student in the Vanderbilt Brain Institute at Vanderbilt University. Jacob I. Feldman, MS, CCC-SLP, is a doctoral student in the Department of Hearing and Speech Sciences at Vanderbilt University. Jenna Crank is a doctoral student in the Department of Special Education at The University of Texas at Austin and a paid employee of a private early intervention clinic that provides NDBI services. Susanne A. Albarran is a doctoral student in the Department of Special Education at the University of Texas at Austin and a Board Certified Behavior Analyst who receives payment for providing ABA therapy. Prachy Mahbub is an undergraduate student at Mount Holyoke College. Tiffany Woynaroski, PhD, CCC-SLP, is an Assistant Professor in the Department of Hearing and Speech Sciences at Vanderbilt University Medical Center, Vanderbilt Brain Institute, and Vanderbilt Kennedy Center. This work was funded in part by Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) of the National Institutes of Health (U54HD083211; PI: Neul), the Training Program in Fundamental Neuroscience of the National Institutes of Health (T32 MH064913; Winder) and the National Institute on Deafness and Other Communication Disorders (1R21DC016144; Woynaroski). The content is solely the responsibility of the authors and does not necessarily represent the official views of the funding agencies. Correspondence concerning this article should be addressed to Micheal Sandbank via

email at msandbank@austin.utexas.edu, or via phone at 512-232-3589. We would like to acknowledge Matt Sandbank for his assistance in typing.

Abstract

In this comprehensive systematic review and meta-analysis of group design studies of nonpharmacological early interventions designed for young children with autism spectrum disorder (ASD), we report summary effects across seven early intervention types (behavioral, developmental, naturalistic developmental behavioral intervention [NDBI], TEACCH, sensory-based, animal-assisted, and technology-based), and 15 outcome categories indexing core and related ASD symptoms. A total of 1,615 effect sizes were gathered from 130 independent participant samples. A total of 6,240 participants, who ranged in age from 0-8 years, are represented across the studies. We synthesized effects within intervention and outcome type using a robust variance estimation approach to account for the nesting of effect sizes within studies. We also tracked study quality indicators, and report an additional set of summary effect sizes that restrict included studies to those meeting pre-specified quality indicators. Finally, we conducted moderator analyses to evaluate whether summary effects across intervention types were larger for proximal as compared to distal effects, and for context-bound as compared to generalized effects. We found that when study quality indicators were not taken into account, significant positive effects were found for behavioral, developmental, and NDBI intervention types. When effect size estimation was limited to studies with randomized controlled trial (RCT) designs, evidence of positive summary effects existed only for developmental and NDBI intervention types. This was also the case when outcomes measured by parent report were excluded. Finally, when effect estimation was limited to RCT designs and to outcomes for which there was no risk of detection bias, no intervention types showed significant effects on any outcome.

Keywords: autism spectrum disorder, intervention, treatment, meta-analysis, robust variance estimation, study quality

Public Significance Statement

This comprehensive meta-analysis of interventions for young children with autism spectrum disorder (ASD) suggests that naturalistic developmental behavioral interventions and developmental intervention approaches have amassed enough quality evidence to be considered promising for supporting children with ASD in achieving a range of developmental outcomes. Behavioral intervention approaches also show evidence of effectiveness, but methodological rigor remains a pressing concern in this area of research. There is little evidence to support the effectiveness of TEACCH, sensory-based interventions, animal-assisted interventions, and interventions mediated solely through technology at this time.

Project AIM: Autism Intervention Meta-Analysis for Studies of Young Children

Autism Spectrum Disorder

Autism spectrum disorder (ASD) is a relatively common neurodevelopmental disorder with a varied impact. Current prevalence estimates suggest that 1 in 59 meet the criteria for ASD, though this prevalence varies by sex, with males having a higher (approximately four times greater) likelihood of being affected (Baio et al., 2018). The diagnosis is primarily associated with core challenges in social communication, as well as restricted interests and repetitive behaviors and differences in sensory function (American Psychological Association [APA], 2013). Individuals with ASD, however, may also exhibit difficulty in a number of related areas, such as language, adaptive behavior, and academic achievement.

A substantial portion of autistic¹ individuals report drawing a sense of identity and empowerment from the diagnosis, and advocate for a neurodiversity conceptualization of ASD as a natural form of human difference (Houting, 2019). Researchers have recently articulated a view of early intervention that is consistent with a neurodiversity framework (e.g., Fletcher-Watson, 2018). Specifically, early intervention services provided throughout childhood may support children with ASD in developing competencies that will allow them to navigate into adulthood in ways they see fit. At present, long-term life outcomes of autistic individuals vary widely. Though a number of individuals that receive early diagnoses go on to develop adaptive and communicative skills within the average range, most require at least some support, and many require substantial support into adulthood (Renty & Roeyers, 2006). Importantly, quality of life

¹Though researchers and clinicians often feel more comfortable with and advocate for using person-first language such as “individuals with autism,” some autistic individuals and their parents have endorsed identity-first language that incorporates autism as a component of their identity over person-first language (Gernsbacher, 2017; Kenny et al., 2016). In this manuscript, we flexibly use identity-first and person-first language to acknowledge the diversity of opinions on this issue within the broader autism community (see Robison, 2019).

among autistic adults also varies between individuals (Howlin & Magiati, 2017). Improving the quality of intervention provided in early childhood may be one way to increase the likelihood that long term life-satisfaction is attainable for all autistic people.

Research on Interventions in Early Childhood

Common intervention recommendations. Recommendations abound regarding the nature and amount of intervention that should be provided to support development in children with ASD. Scholars and professionals have routinely asserted that intervention should be provided as early as possible, beginning at or even before diagnosis in toddlerhood or infancy; that intervention should be intensive (i.e., provided for 25-40 hours per week for over a year or longer); and that it should be comprehensive (i.e., targeting broader development rather than specific skills; Boyd, Odom, Humphreys, & Sam, 2010; Lovaas, 1987; McEachin, Smith, & Lovaas, 1993; Lord et al., 2001; Odom, Boyd, Hall, & Hume, 2010). These recommendations are motivated by the theory that interventions provided in early childhood are likely to yield the most optimal effects by capitalizing on the neuroplasticity of the developing brain (Dawson & Zanolli, 2003; Kolb & Gibb, 2011), and are rooted in early influential studies which suggested that intensive intervention yielded substantial cognitive gains, and that such gains varied according to age at the onset of intervention (e.g., Lovaas, 1987; McEachin, Smith & Lovaas, 1993). However, it is notable that some subsequent studies exploring putative predictors of treatment response have reported that age at intake was not significantly associated with intervention outcomes (e.g., Eikeseth, Smith, Jahr, & Eldevik, 2007; Eikeseth, Klintwall, Jahr, and Karlsson, 2012).

Types of intervention approaches. Several approaches to intervention aim to address the core and related challenges associated with ASD. These approaches vary in their underlying

theories on the nature of ASD and development, as well as in their procedures and instructional modalities.

Behavioral approaches. Behavioral interventions were among the first developed and clinically tested approaches for improving outcomes for children with autism (Ferster & DeMeyer, 1962). These approaches are derived from operant learning theory and are characterized by the discrete presentation of information (i.e., a stimulus), the prompted exhibition of target responses (i.e., desired academic, adaptive, and communicative behaviors), and the provision of extrinsic positive reinforcement (e.g., edible treats, toys, stickers, etc.) in the presence of those responses. Target skills are chosen based on functional areas of child need. Skills tend to be initially targeted in highly structured interactions within isolated clinical contexts (e.g., in the course of one-on-one interactions at a clinic with a therapist), but more natural settings and interaction partners (e.g., mainstream classrooms and other children) are gradually integrated as a child demonstrates progress. Initial studies suggested that Early Intensive Behavioral Intervention (EIBI) could yield marked improvements in cognitive and academic placement outcomes for children with ASD, especially when provided before school age and with sufficient intensity (Lovaas, 1987; McEachin, Smith, & Lovaas, 1993). In the wake of such research, a number of behavioral approaches were further developed and refined, and the Behavior Analyst Certification Board (BACB) was established to oversee the clinical certification associated with this approach. Other behavioral interventions include Discrete Trial Training (DTT), Picture Exchange Communication System (PECS), and Positive Behavioral Supports (PBS). Together, these interventions are sometimes loosely described as Applied Behavior Analysis (ABA) Therapy and now constitute the primary approach used in clinical

practice, according to parent and provider reports (Green et al., 2006; Stahmer, Collings, & Palinka, 2005).

Developmental approaches. At times viewed in contrast to the aforementioned traditional behavioral interventions are those derived from developmental theories of learning (e.g., Ospina et al., 2008; Prizant & Wetherby, 1998). Developmental interventions are rooted in constructivist theory, which posits that development is the result of children's active exploration of their physical and social surroundings. This exploration is far from being a solitary endeavor, as children are supported in social and language development by their interactions with more competent interaction partners such as caregivers (Bruner, 1982; Vygotsky, 1978). Foundational research on ASD within the developmental tradition has suggested that early deficits in social processes (joint attention being of particular importance) in children with ASD may in turn lead to difficulties in early caregiver-child social interactions. These early deficits are thus viewed as disrupting the primary context for subsequent language and social communication development. As such, developmental interventions focus on improving the synchrony, reciprocity, and duration of parent-child or child-child interactions as a pathway for ameliorating deficits in social communication and generating cascading improvements in developmentally related skills. These interventions are primarily delivered in the context of everyday routines such as play, and intervention goals are chosen based on the typical sequences of social communication and language development. Examples of classically developmental interventions include DIR/Floortime (Greenspan & Wieder, 2007) and Hanen models (Carter et al., 2011).

Naturalistic Developmental Behavioral Interventions (NDBIs). In 2015, several interventions were categorized as belonging to a third type of intervention approach which has theoretical underpinnings in both behavioral and developmental theories of learning and

development. Naturalistic Developmental Behavioral Interventions (NDBIs) involve the use of behavioral principles of learning to teach skills chosen from a developmental sequence in naturalistic environments and using natural rewards (Schreibman et al., 2015). Skills selected as relevant for intervention are those that allow the child to participate more fully within reciprocal interactions with the adult. These interventions are delivered primarily in the context of play, but control of interactions within this context is shared by both the child and the adult, through balanced turn-taking. Interventions categorized as NDBIs include the Early Start Denver Model (Rogers & Dawson, 2010); Enhanced Milieu Teaching (EMT; Kaiser, 1993); Pivotal Response Treatment (Koegel, Koegel, & Carter, 1999); and Joint Attention, Symbolic Play, Engagement, and Regulation (JASPER; Kasari, Freeman, & Paparella, 2006).

TEACCH. The TEACCH (Treatment and Education of Autistic and related Communication-handicapped Children) program was developed in 1972 by Eric Schopler and is based primarily in the state of North Carolina (Mesibov, Shea, & Schopler, 2004). We consider this specific intervention as distinct from other approaches because of the explicit focus on structured environmental design and self-monitoring, which is not the emphasis of any of the other interventions of interest to the present synthesis. The theoretical foundations of TEACCH are rooted neither in behavioral nor in developmental theories of learning. Rather, TEACCH procedures were designed according to Schopler's theorized profile of the learning strengths, preferences, and needs of individuals with ASD, which include relative visual strength and comfort with consistent routines. Thus, the TEACCH program is characterized by highly structured work routines and a heavy reliance on the visual presentation of information. TEACCH "work systems" organize individual student tasks to visually convey four pieces of information: (1) What activity the student will complete, (2) How many items need to be

completed, (3) How to identify when the work is finished, and (4) What will happen after task completion. TEACCH classrooms tend to feature carefully planned and structured environmental arrangements, work areas with minimal distractions, consistent routines, and the extensive use of visual schedules and supports.

Sensory-based interventions. Sensory-based interventions are motivated by the theory that sensory function is foundational in nature, and that sensory disruptions, particularly early in life, may produce cascading effects on development across a number of domains, ultimately yielding the constellation of core and related characteristics associated with ASD (e.g., Bahrick & Todd, 2012). Within this framework, it is hypothesized that targeted treatments may thus have the potential not only to ameliorate reported sensory differences, but also to translate to effects on higher-order social, communication, and cognitive skills in children with ASD (Cascio, Woynaroski, Baranek, & Wallace, 2016). The most well-known of these sensory-based approaches to treatment is Sensory Integration Therapy, in which children are presented with a series of individualized sensory-motor experiences intended to build foundational skills that will facilitate their engagement and participation in a range of activities of daily living (Ayres, 1979; Ayres, 2005). Other sensory-based interventions, as broadly conceptualized, may include activities such as brushing, swinging, the use of weighted vests and blankets to improve sensory processing, and music therapy and auditory integration training approaches that aim to scaffold motor, social, and emotional development (e.g., Baranek, 2002; Case-Smith and Arbesman, 2008). Sensory based approaches are most often provided by occupational therapists in clinical contexts but may also be delivered by caregivers, educators, and/or other service providers across a broader range of home and community settings.

Animal-assisted interventions. Animal-assisted interventions are those that rely on interactions with animals as the primary context for facilitating developmental change (e.g., O’Haire, 2013; 2017; Trzmiel, Purandare, Michalak, Zasadzka, & Pawlaczyk, 2018). In the ASD intervention literature, the intervention most prominently represented in this category is equine-assisted activities and therapy (EAAT; see Gabriels et al., 2012 for a review of related terminology). Proponents of EAAT contend that the activities of horse-riding and horse care provide a multisensory experience that allows children the opportunity to practice skills across multiple domains. More broadly, animal-assisted interventions are theoretically motivated by the possibility that human-animal interactions are highly motivating and provide calming contexts which may support improved psychological wellbeing and social function.

Technology-based interventions. Technology-based interventions employ one or more of a variety of technologies (e.g., computers, videos, video games, robots) as the primary medium for delivery of instruction. These interventions attempt to capitalize on the reported special interest that many autistic individuals have in computer technology (Grynszpan, Weiss, Perez-Diaz, & Gal, 2014) and predictable formats of information delivery (Baron-Cohen, Golan, & Ashwin, 2012), which allow users to control the pace of the interaction (Knight, McKissick, & Saunders, 2013). Examples of technology-based interventions include computer-assisted instruction and The Transporters™ DVD series (e.g., Young & Posselt, 2012).

Previous Syntheses of Intervention Literature

The National Professional Development Center (NPDC) on Autism Spectrum Disorders generated a list of 27 evidence-based practices for improving outcomes in individuals with ASD, based on prior reviews of single subject and group design research (Wong et al., 2015). Similarly, the National Standards Project (NSP, 2015) described 14 intervention practices as

established for children with ASD and an additional 18 as emerging, based on a review of single subject and group design literature. In 2011, Warren and colleagues systematically reviewed 34 group design studies examining interventions in children with ASD. Notably, only two of the studies included in the review by Warren et al. (2011) were randomized controlled trials (RCTs), and only one of those was rated as high quality. Very recent systematic reviews suggest the publication of RCTs has precipitously increased in ASD since the publication of the aforementioned synthesis by Warren and collaborators. For example, French and Kennedy (2017) systematically reviewed RCTs of interventions targeting any outcome in children with ASD below age 6, and found a total of 48 RCTs, 40 of which had been published since 2010.

Previous efforts to synthesize this literature have a number of shortcomings. First, NPDC and NSP review procedures attempted to synthesize evidence from RCTs, quasi-experimental studies, and single subject design studies (SSDs), when there is currently no agreed upon way of doing so. Though multiple methodologies can contribute to knowledge about effective practices, studies employing group designs, in particular high-quality RCTs, are the best equipped to control for alternative explanations and threats to internal validity. Syntheses that attempt to combine RCTs, quasi-experimental studies, and SSDs may overestimate the effectiveness of a given intervention approach. Inclusion of SSDs also limits the extent to which summary effects of intervention can be quantified with meta-analytic approaches. Though effect sizes that quantify change observed in SSDs have been proposed, many of these approaches fail to account for first order autocorrelation of data, ignore the logic of within study replication that is critical to interpretation of SSD data, and yield highly inflated and positively biased effect sizes which are not comparable to mean group differences that index treatment effects in group design (Wolery, Busick, Reichow, & Barton, 2010; Zimmerman et al., 2018).

Second, in previous reviews, limited consideration was given to the nature of outcomes measured. That is, prior syntheses of intervention literature have predominantly sought to ascertain whether various approaches to interventions are “evidence-based,” but they have largely failed to summarize the extent to which interventions effected meaningful change. Interventions that were shown to effect change that was overly specific to intervention targets were generally not distinguished from those that impacted scores on broader standardized assessments of developmentally advanced skills as administered by independent assessors. A synthesis is needed which asks not only “what works and for whom,” but also, “for what?”

Third, none of the prior reviews seeking to synthesize effects for the broad range of interventions geared towards young children with ASD attempted to identify the summary effects of varied interventions on any outcomes using meta-analytic tools. Although a narrative synthesis approach allows for tallying the number of studies that have shown an effect for a given outcome, they do not allow for deriving an estimate of the combined magnitude of the effect, or determining whether or not the combined effect is significantly different from zero. Additionally, narrative synthesis methods are unable to offer information about variables that may moderate effect sizes. Moderator effects offer vital information for understanding for whom interventions are effective, and for identifying study design features that result in potentially inflated effect sizes.

Crucial Quality Considerations

Although systematic reviews and meta-analyses are purported to provide the most reliable summary of evidence of intervention effects, their conclusions are limited by the quality of evidence which they summarize (Higgins et al., 2011; Murad, Asi, Alsawas, & Alahdab, 2016). Several aspects of study design pose risk of biasing outcomes. Thus, examination of any

set of intervention literature must include an assessment of several study-level quality indicators. We outline here those that are particularly important in studies of nonpharmacological interventions of children with ASD.

Random assignment. Though some have questioned the feasibility of conducting randomized controlled trials to test the effects of “real world” interventions with individuals with disabilities (Oliver et al., 2002), random assignment remains the most rigorous control for rival explanations of findings. Though random assignment does not ensure pretreatment statistical equivalence between groups on all variables, it is the best procedural guard against systematic differences between groups that would limit confidence in conclusions about causal associations between the intervention and dependent variables (Kasari, 2002). Historically, randomized tests of interventions have been exceptionally rare in ASD research (Warren et al., 2011). However, the recent proliferation of RCTs in this field suggests that random assignment is feasible and employed frequently enough to permit an evaluation of evidence from randomized trials versus quasi-experimental studies.

Independence of assessors. Detection bias refers to the risk of bias that arises when assessors are aware of the group assignment of individual participants. This type of bias manifests in different ways in studies of autism intervention, and the degree of risk may vary depending on the extent to which non-independent assessors are involved in outcome assessment. It is likely that detection bias poses the greatest threat when caregivers participate in outcome assessment, either as reporters or interaction partners, though the threat is still substantial in situations wherein outcomes are assessed or coded by professionals that are aware of group assignment.

Caregiver/teacher report. It is common for researchers to rely on parents or teachers to assess outcomes via standardized interview and/or report forms in pediatric psychology and adjacent fields. Because caregivers observe and engage with children for extended periods of time across a variety of contexts, they can draw on their cross-context knowledge of a child's abilities when reporting on an outcome, and may therefore produce scores that are more representative of a child's generalized abilities, compared to scores derived from brief assessments administered by unfamiliar examiners. However, parents and teachers are virtually always aware of the extent and nature of a child's participation in an intervention study. Moreover, they are likely to be personally invested in the outcome of intervention. This combination of awareness of group assignment and strong investment in positive outcomes can yield a "placebo by proxy" effect, which can positively bias results in favor of the treatment group (Grelotti & Kaptchuk, 2011). Prior placebo-controlled studies of pharmacological interventions such as secretin have demonstrated that these effects can be rather large (Williams, Wray, & Wheeler, 2012), and present even in simulated clinical trials where no intervention was provided (Jones, Carberry, Hamo, & Lord, 2017). Thus, outcomes from caregiver report are highly subject to systematic measurement error and may positively bias summary estimates of intervention effects.

Outcomes assessed in interactions with caregivers. Even in situations that do not involve standardized report, caregivers can exert undue influence on outcome measurement. This occurs when caregivers participate as interaction partners in observational measures of outcomes of interest. Autism researchers frequently use observational measurement to capture social communication and related skills in the natural contexts in which they arise. For example, scores of language and communication are often derived from free play sessions with parents, or from

interactions with teachers in the classroom. These scores are fundamentally dyadic; though they are often assumed to solely represent the skills or behavior of the child, they actually index the child's response to the interaction partner. When interaction partners are aware of the administration of a treatment, they may subconsciously or consciously shift their behavior to better elicit skill demonstration from the child. Though this threat arises often in studies of interventions on language and communication outcomes, it is not limited to measures of those domains. Therefore, outcomes measured in the context of natural interaction with caregivers are also subject to bias and may influence intervention effect sizes.

Outcomes assessed or coded by professionals aware of group assignment. Even unfamiliar professionals can influence outcomes when administering standardized assessments or coding observational measures of behaviors. A recent systematic review of medical literature that contained assessment of binary outcomes from both independent and non-independent assessors found that assessors that were aware of group assignment exaggerated odds ratios as much as 36% (Hróbjartsson et al., 2012).

Influential Outcome Characteristics

The Cochrane Collaboration has delineated a set of quality indicators that are applicable to intervention literature in most fields, but additional field-specific sources of bias also exist for autism early intervention literature. Further, when it comes to studies of intervention for children with ASD, we contend that various aspects of outcome measurement can also serve as sources of bias and should therefore be considered (Yoder, Bottema-Beutel, Woynaroski, Chandrasekhar, & Sandbank, 2013). We summarize two particularly important dimensions of outcome variables below (boundedness and proximity), and we review one additional source of bias related to study design that we hypothesize has the potential to influence effect sizes observed across studies of

treatment effects on outcomes of young children with ASD (correlated measurement error [CME] that arises when parents or teachers are trained in the intervention and then participate in the data collection).

Boundedness of outcomes to intervention context. Whether or not an intervention effects change that generalizes beyond the context of an intervention is a question of great importance. While the context of intervention is generally contrived and temporary, changes effected by intervention are often assumed to (or at least intended to) extend to natural environments and the routines of daily life. However, dependent variables vary in the extent to which they index generalized change. Those that are measured within the context of intervention, or in a context that is similar to intervention across several dimensions (i.e., materials, setting, interaction partners, interaction style), may reflect changes that are potentially bound to the intervention context. In contrast, dependent variables that are measured in a context that differs from the intervention on several dimensions should reflect highly generalized changes. For example, in the hypothetical study of an intervention that is administered during play with a therapist, outcomes measured in a play-based interaction with a familiar therapist and similar toys may index change that is bound to that context. The outcome measure does not afford any degree of confidence that the treatment has induced changes in child behavior that would generalize to other contexts. In contrast, outcomes measured using standardized assessment procedures (i.e., different interaction style and materials) administered by an unfamiliar examiner (i.e., different interaction partner) would likely reflect change that reaches across a wide range of contexts. Similarly, outcomes measured in the home environment in an interaction with a parent (i.e., different setting, interaction partner, and interaction style, assuming the parent has not been trained in the intervention), would serve as a naturalistic assessment of highly generalized

change in this hypothetical study. In theory, generalized change is more difficult to effect than context-bound change, so effect sizes for generalized outcomes are likely to be smaller relative to effect sizes of outcomes that are potentially context-bound.

Proximity of outcomes to intervention targets. Outcomes may also vary by their proximity to the targets or goals of the intervention. Ideally, interventions would be able to demonstrate change not only on outcomes that are directly taught or addressed by the intervention (i.e., *proximal* outcomes), but also on outcomes that are developmentally downstream from what is directly taught or addressed (i.e., *distal* outcomes). When interventions are able to demonstrate growth on distal outcomes, they are essentially providing evidence that the intervention is influencing children's development, which may mean that the intervention will continue to have effects long after the intervention has stopped. However, prior best evidence syntheses have shown that early interventions for children with ASD show much larger effects for proximal as compared to distal outcomes (Yoder et al., 2013).

Correlated measurement error in parent/teacher mediated interventions. In addition to other commonly cited sources of bias, studies of autism-specific interventions are frequently threatened by CME that occurs when parents or teachers are the interventionists and also participate in assessment procedures. Parents and teachers are primary figures in the lives of children with ASD, and this makes them ideal mediators of intervention. For this reason, researchers have developed a number of interventions that target parents and teachers as interventionists, and tested their effectiveness in parent- or teacher-training studies. Trouble arises when natural interaction partners are trained as interventionists over the course of a study while simultaneously participating as assessors, either by rating child outcomes via a standardized report, or by serving as the interaction partner in an observational assessment

context. The risk of bias posed by this specific study design flaw extends beyond that posed by detection bias related to the non-independence of assessors. This is because, in addition to being aware of group assignment, the assessors and assessment context has also changed from pre to post intervention in a manner that favors the intervention group. For example, a study might test the effects of parent-training for improving communication in children with ASD by examining the frequency of child communication during free play with parents. Prior to intervention, the assessment context in both groups would feature a parent naive to strategies for eliciting communication. However, after intervention, the assessment context in the treatment group would feature a parent who is more adept at eliciting communication while the assessment context in the control group remained the same. Though these two assessment contexts seem identical, they are fundamentally different. Though studies of parent- and teacher-led interventions are not unique to this population, they are well-represented in autism intervention literature. As such, any assessment of study quality should include an evaluation of the potential influence of this field-specific source of bias.

Study Purpose and Research Questions

The purpose of this study is to gather and synthesize all available studies of nonpharmacological interventions targeting any outcome in children with ASD below the age of 8 years. Our specific research questions were:

1. Across all eligible quasi-experimental and experimental studies, are summary effects positive and significant for targeted outcomes for each of seven intervention types (behavioral, developmental, NDBI, TEACCH, sensory-based, animal-assisted, and technology-based)?

2. Are summary effects positive and significant for targeted outcomes for each of the aforementioned seven intervention types when only outcomes from studies with basic quality controls (i.e., random assignment, independent assessors) are included?
3. Across intervention and outcome types, are summary effects for proximal outcomes larger than summary effects for distal outcomes?
4. Across intervention and outcome types, are summary effects for outcomes that measure context-bound behaviors larger than summary effects for outcomes that measure more highly generalized characteristics?

Method

Search

Search Terms and Databases. To gather the peer-reviewed literature included in the current meta-analysis, the following nine online databases were searched: Academic Search Complete, CINAHL Plus with Full Text, Education Source, Educational Administration Abstracts, ERIC, MEDLINE, PsycINFO, Psychology and Behavioral Sciences Collection, and SocINDEX with Full Text. Search terms were used in various combinations to capture the diagnostic criteria and intervention designs included within the search. The individual databases were searched using the following terms: *autis**, *ASD*, *PDD*, *Aspergers*, *intervention*, *therapy*, *teach**, *treat**, *program*, *package*, *assign**, *control group*, *BAU*, “wait list”, *RCT*, *random**, *quasi*, “treatment group”, “intervention group”, “group design”, and *trial*. This initial search yielded 12,933 results from academic journals, dissertations, books, reports, conference materials, and reviews.

To gather grey literature, or studies not published in peer-reviewed journals, investigators who received federal grants to study autism were identified through a search of the National

Database for Autism Research (NDAR), the National Institutes of Health (NIH) Matchmaker, and Institute of Education Sciences (IES) websites. A list of researchers ($n = 106$) was generated, and 90 of these investigators were emailed with a request for eligible data. The contact information for the remaining investigators could not be found.

Screening process. A preliminary screen of abstracts was first completed using abstractkr (Wallace, Small, Brodley, Lau, & Thomas, 2012). Studies were screened at the full-text level if they met the following inclusion criteria: (a) published in English, (b) published from 1970 - present, (c) group design that included both an intervention and control group, (d) a simple majority of participants were reported to have a diagnosis of ASD, and (e) the average age of included participants was between 0 and 8 years. In many instances, though studies met inclusion criteria, insufficient information was provided to enable the extraction of unadjusted effect sizes. In these cases, authors were identified and emailed with a request to provide unadjusted post-intervention means and standard deviations. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram in Figure 1 summarizes the search process and provides justifications for exclusion of articles.

Coding Procedures

Included studies were coded for participant characteristics, intervention characteristics, study characteristics (including quality indicators), outcome characteristics, and effect size information. The coding manual is available upon request from the first author.

Participant characteristics. Participant characteristics coded from studies included average age of participant samples in months, percentage of sample that was male, and average language age in months (either receptive, expressive, or total) whenever it was reported.

Intervention characteristics. Intervention approaches were categorized based on the specific techniques used and the underlying philosophies that motivated the approach. A set of candidate categories (behavioral, developmental, NDBI, sensory-based, technology-based, cognitive behavior therapy, other) were drafted in the first instantiation of the coding manual for this synthesis based on authors' knowledge of intervention literature. Based on the results of our literature search and screening process, as well as the range of intervention approaches encountered across our team's initial training on coding precision and reliability, intervention categories were further refined to include 'animal-assisted therapy'. This intervention approach was found to be motivated by a distinct theoretical framework and to have amassed a sufficient number of group design studies to permit prior systematic review and meta-analysis (O'Haire, 2013; 2017; Trzmiel, Purandare, Michalak, Zasadzka, & Pawlaczyk, 2018). Thus, interventions were initially coded as animal-assisted therapy, behavioral, developmental, NDBI, cognitive behavior therapy, sensory-based, technology-based, or other. After completion of coding, the set of interventions coded as 'other' were re-examined to determine whether there existed a sufficient set of similar studies (e.g., 5 or more) that could be meaningfully combined to comprise an additional category. This was the case for studies of the TEACCH intervention. Studies of TEACCH that were initially coded as 'other' were, therefore, re-coded as 'TEACCH'.

Animal-assisted therapy. Interventions coded as animal-assisted therapy were those mediated through the presence of an animal. Equine Assisted Therapy was an example listed in the coding manual.

Behavioral. Interventions were coded as behavioral if authors described the intervention as being heavily situated in operant theories of learning, or if they relied heavily on behavior analytic techniques, such as didactic instruction, prompting, shaping, and extrinsic

reinforcement. Examples of behavioral interventions listed in the coding manual included EIBI, PECS, Discrete Trial Training, Verbal Behavior, Autism Partnership, and the Lovaas Model.

Developmental. Interventions were coded as developmental if they were described as being motivated by constructivist theories of learning, or if they were heavily child-led and implemented according to a typical developmental sequence, with the goal of facilitating the development of foundational skills that would translate to gains in developmentally downstream domains. Examples of developmental interventions featured in the coding manual included DIR/Floortime, Hanen models, and Responsive Teaching.

NDBIs. Interventions were coded as NDBIs if they were one of any of the named interventions in the consensus paper on this intervention approach (Schreibman et al., 2015), or if they combined adult-led, behavioral teaching methods with child-led routines and taught to a natural developmental progression within naturalistic settings. These included Incidental Teaching, Pivotal Response Treatment, Early Start Denver Model, Enhanced Milieu Teaching, Reciprocal Imitation Training, Project ImPACT, JASPER, SCERTS, Early Achievements, and Prelinguistic Milieu Teaching. Although Prelinguistic Milieu Teaching is not explicitly listed as an example of an NDBI in the consensus paper by Schreibman and colleagues, we contacted Paul Yoder, a leading researcher of this intervention, while drafting the coding manual for this meta-analysis to verify that this would be the appropriate category for this intervention approach (Personal communication, March 29, 2018).

Cognitive behavior therapy. Interventions were coded as cognitive behavior therapy if a study explicitly named the intervention as such.

Sensory-based interventions. Interventions were coded as sensory-based if they incorporated targeted exposure to sensory or multisensory (e.g., auditory, visual, tactile,

olfactory) stimuli. Examples listed in the coding manual included sensory integration, music therapy, massage, acupuncture, auditory integration, and weighted blankets. This category was drafted based on precedent across prior reviews of sensory-based interventions (Baranek, 2002; Case-Smith and Arbesman, 2008; Weitlauf et al., 2017).

Technology-based interventions. Interventions were coded as technology-based if the intervention was primarily delivered on a computer or electronic device (i.e., iPad, DVD).

TEACCH. Interventions were re-coded as TEACCH if a study explicitly identified using this method.

Other. Interventions that did not fit into the previously defined categories were coded as other.

Study characteristics. Study-level characteristics that were coded include design type (i.e., randomized control trial [RCT] or quasi-experimental), publication status (i.e., indexed, non-indexed, unpublished), and several features of study quality. Studies were coded as a randomized controlled trial if the text indicated that participants were randomly assigned to an intervention group and a control group or contrasting treatment, or if the authors referred to the study as “randomized.” Studies were coded as quasi-experimental when authors made no indication that the process of group allocation was random. If a contrasting treatment model was used, the group receiving the treatment that was hypothesized by the authors to effect greater change was considered the treatment group. In the case of studies testing multiple active treatment groups compared to a passive control, treatment characteristics and effects were coded separately in contrast to control.

Publications were coded for whether they were published or unpublished. Published studies included indexed and non-indexed journals, and unpublished studies included

dissertations and theses. Despite our extensive attempts to locate, obtain, and include unpublished data apart from dissertations and theses, no researchers provided us with data or effect sizes from unpublished reports.

Studies were coded for several indicators of study quality. These indicators included those specified by the Cochrane Collaboration's tool for assessing risk of bias (e.g., selection bias, performance bias, detection bias, reporting bias; Higgins et al., 2011), as well as additional indicators which we proposed in prior work (e.g., potential presence of CME related to parent/teacher training, sufficient number of participants to justify statistical analysis, reliance on parent or teacher report; Yoder et al., 2013). Selection bias related to insufficient randomization procedures and allocation concealment was coded as "high", "low", or "unclear" for studies coded as randomized controlled trials, and as "not applicable" for quasi-experimental studies. For subsequent Cochrane quality indicators, risk of bias was coded as "high" or "low" if studies explicitly indicated or provided sufficient information to ascertain the presence or absence of such risk, and as "unclear" if information related to risk potential was not detailed. Risk of selective reporting bias was coded as high if outcomes were reported to have been collected at post but were not reported in results, or if an entire assessment was administered but only selective subscores were reported without sufficient justification. Performance bias risk was assessed in consideration of the participants' and families' awareness of their group assignment. Detection bias accounted for the independence of assessors and coders. We elected to include interaction partners in naturalistic observational measures as "assessors," given that they may transiently influence child behavior during interactions. Attrition bias was coded with respect to the number of participants recruited and the number of participants included in analysis.

Specifically, attrition bias was considered low if attrition was lower than 20% or if intent-to-treat analysis was utilized.

Outcome characteristics. In addition to the above quality indicators, we coded for outcome-level quality indicators that are especially important for research on intervention in young children with ASD. These quality indicators captured the boundedness and distality of outcomes, as well as the potential presence of CME related to parent or teacher training. Outcomes were coded as context-bound if they were measured in or very near the context of the intervention, and as generalized if they were measured in a context that differed from the context of intervention on multiple dimensions (e.g., interaction partners, materials, setting, interaction style). Outcomes taken from standardized parent/teacher reports were coded as potentially context-bound if reporters were also the primary mediators of intervention, based on the rationale that their report could be based on their observance of the outcome as it occurred within the context of the intervention they provided. Outcomes were coded as proximal if they indexed skills that were directly taught, modeled, or prompted during the intervention, and otherwise as distal. Outcomes indexed by developmentally scaled assessments were automatically coded as distal, based on the reasoning that these assessments are meant to tap generalized development rather than specific skills. We recognize that it is possible for an intervention to directly target specific items of a developmentally scaled assessment, but reasoned that in the absence of an extremely detailed description of intervention procedures, we should assume these assessments captured constructs beyond what was directly taught in intervention. Decision trees used to judge distality and boundedness are presented in Figures 2 and 3, respectively. Correlated measurement error related to parent/teacher training was coded as potentially present when parents or teachers operated as both the mediators of intervention as well as the outcome assessors.

Outcome categorization. Each dependent variable was categorized as either a core feature of ASD (i.e., social communication; restricted/repetitive patterns of behaviors, interests, or activities; sensory) or a related outcome (i.e., language, motor, adaptive, cognitive, academic, play, sleep, brain imaging, social emotional/challenging behavior). If outcomes were reported at multiple time points, immediate and follow-up outcomes were coded separately.

Effect size information. Unadjusted means, SDs, and *ns* were extracted from all eligible studies that reported a group difference between participants receiving the specified intervention and those not receiving the specified intervention. Group difference effect sizes were calculated for each outcome using the standardized mean difference (*d*), as derived via the Campbell Collaboration Practical Meta-Analysis Effect Size Calculator (Lipsey & Wilson, 2001) and then converted to the effect size metric used for analyses, Hedge's *g* (*g*). Effect sizes were reported in such a way that higher *g* scores indicated superior performance in the treatment group.

We were unable to extract effect sizes from some eligible studies due to insufficient information (e.g., authors did not report means and SDs, reported only mean change scores, or reported means and SDs that were adjusted for baseline covariates and therefore could not be meta-analyzed with unadjusted means and SDs). When this occurred for articles published within the last ten years, we contacted the corresponding author(s) in an attempt to obtain either the unadjusted post means and SDs, or any other statistical information that would allow us to calculate the standardized mean difference between treatment and control/contrast groups after intervention. Fifty-five studies did not have sufficient information to allow effect size extraction for all outcomes. In the case of nine of these studies, effect size extraction was possible for some but not all outcomes, so eligible outcomes were included from those studies. Authors responded and supplied effect size information for 14 additional studies.

Reliability

A primary coder (the first author) read and coded all studies. All studies were also independently coded for reliability by one coder from a team of nine. Both coding sheets were then sent to a separate coding auditor who examined codesheets for discrepancies and reported any disagreements between coders. Original primary and reliability codes were then saved for reliability analyses in a separate folder, and all disagreements were addressed in discrepancy discussions between the primary and reliability coders. Discrepancies were considered resolved once both coders agreed to a final consensus code, which was then added to the dataset used for the final analyses. Therefore, we are able to report reliability data from the original coding and also confirm that *all* disagreements were resolved prior to statistical analysis.

All reliability calculations were completed in R studio (R Core Team, 2017) using the *irr* package (Gamer, Lemon, Fellows, & Singh, 2012). Reliability was indexed using unweighted kappa for all categorical variables (Cohen, 1960) and one-way random intraclass correlation coefficients for all continuous variables (ICC; Shrout & Fleiss, 1979). Kappas ranged from 0.602-0.923, and average kappa across all categorical variables was 0.751. ICCs ranged from 0.676-0.999, and average ICC across all continuous variables was 0.916.

Analysis

All analyses were conducted in R (R Core Team, 2017). To account for the nesting of multiple effect sizes within overlapping participant samples, we used robust variance estimation (RVE) with small sample adjustments when synthesizing effect sizes and conducting meta-regressions (Hedges, Tipton, & Johnson, 2010; Tanner-Smith, Tipton, & Polanin, 2016). These procedures account for the non-independence of effect size statistics drawn from overlapping

samples, and provide test statistics and confidence intervals that are adjusted based on how the effect sizes are clustered.

Effect sizes were aggregated based on type of outcome (see Outcome characteristics) within each type of intervention (see Intervention characteristics). Aggregating the results in this manner provided a summary statistic for the effect of each intervention type on each outcome type. Meta-regression analyses were conducted on the coded variables of distality and boundedness (see Outcome characteristics) to determine whether the magnitude of the effects across intervention and outcome types were moderated by these categorical characteristics related to measurement. The threshold level of significance for these tests was set at $p < .10$, given that we had clear directional hypotheses for each potential moderator, meriting one-tailed tests of significance. To examine the potential presence of publication bias, we examined funnel plots of effect size estimates against their standard errors, and corresponding Egger's tests of funnel plot asymmetry, for each summary effect estimate. Due to the large number of significance tests this demanded, we applied the Benjamini-Yekutieli false discovery rate correction to the significance values from the Egger's tests to correct for spurious findings using the Hmisc package in R (Harrell, 2018). The Robumeta package in R (Fisher, Tipton, & Zhipeng, 2017) was used to conduct these analyses while the Metafor package (Viechtbauer, 2010) was used to graph the forest plots and funnel plots.

Results

Descriptives of Included Study Samples and Outcomes

The search and screening process yielded 1,615 effect sizes gathered from 130 independent study samples (from a total of 149 reports) representing 6,240 participants. Across all studies, the average age of participants was 54.21 months ($SD = 18.98$), the average

proportion of male participants per sample was 0.84 ($SD = 0.07$), and the average language age of participants in studies for which it was reported was 22.68 months ($SD = 11.91$). An average of 12.4 outcomes were reported for a single study sample ($MIN = 1$, $MAX = 100$, $MDN = 8$). Participant characteristics according to intervention type are reported in Table 1. There were 27 studies of behavioral interventions, 14 studies of developmental interventions, 26 studies of NDBIs, seven studies of sensory-based interventions, ten studies of technology-based interventions, and six studies of TEACCH included in the synthesis. The RVE approach requires that at least five studies contribute to the generation of effect sizes, so the studies representing animal-assisted intervention ($n = 4$), cognitive behavioral therapy ($n = 2$), and other varied approaches that could not be meaningfully combined into intervention types ($n = 29$) were excluded from summary effect estimation.

Study Quality

Figures 4 and 5 illustrate the proportion of studies or outcomes that received each quality rating (i.e., low risk of bias, high risk of bias, unable to determine) for seven key quality indicators, according to intervention type. These figures include only studies that contributed to summary effect estimation. Because it is almost always impossible for participants to be naive to intervention delivery in studies of nonpharmacological interventions for ASD, performance bias was rated as high for all but one study included in summary effect estimation and, thus, is not reported separately for each intervention type (see Corbett, Schickman, & Ferrer, 2008 for the lone exception).

Behavioral intervention studies. Figure 4 reflects information regarding quality indicator ratings for studies of behavioral interventions. Notably, only 29.63% of studies of behavioral interventions were RCTs. Detection bias was rated as high for 77.05% of outcomes in

behavioral studies. High detection bias in this set of studies was largely driven by an overreliance on reports completed by individuals who were aware of intervention assignment – 60.33% of outcomes were based on parent or teacher report. Correlated measurement error related to parent/teacher training threatened 53.77% of outcomes reported in behavioral studies. Since many of the studies relied on standardized report forms, and because most of these studies only loosely described intervention targets, 86.23% of outcomes tracked in behavioral intervention studies were categorized as distal to the intervention targets. Half (50.49%) of outcomes were categorized as generalized, and 10.49% were classified as context bound. The remaining 39.02% of outcomes were categorized as potentially context bound, because they were derived from caregiver reports in studies where caregivers participated as interventionists (meaning that it is unclear if the outcome could be demonstrated in interactions with individuals who were not trained as interventionists). Bias related to substantial attrition (i.e., > 20% of the study sample) was rated as high for 15.41% of all outcomes.

Developmental intervention studies. Figure 4 reflects quality indicator ratings for studies of developmental interventions. A large majority (78.57%) of included developmental studies were RCTs. Detection bias was rated as high for 53.97% of outcomes, but this was not due entirely to over-reliance on caregiver report. Nearly a third (29%) of outcomes were taken from parent/teacher report. The remainder of outcomes flagged for high detection bias (approximately half of the outcomes tracked in these studies) reflects the common practice of measuring language and communication outcomes in the context of interactions with natural communication partners (primarily parents, who were aware of group assignment). CME related to parent/teacher training threatened three quarters (75%) of all outcomes in developmental studies. Since many of the developmental interventions were explicitly described as targeting

language and social communication, and many of the outcomes were observational measures of language and social communicative behaviors, approximately half (53.57%) of outcomes were categorized as proximal to intervention targets. Approximately a quarter (27.84%) of outcomes were categorized as generalized, a quarter (25%) were categorized as potentially context-bound, and approximately one half (47.16%) were categorized as context-bound. Over a third (34.66%) of all outcomes were subject to high bias from substantial attrition.

Naturalistic developmental behavioral intervention studies. Figure 4 illustrates quality indicator ratings for included studies of NDBIs. A large majority (76.92%) of included studies of NDBIs were RCTs. Detection bias was rated as high for 59.42% of outcomes. This was due, in part, to the common use of observational measures of skills coded from natural interactions with interaction partners who were aware of group assignment. Only 17% (the lowest of any intervention type) of outcomes were collected from parent/teacher report. However, CME related to parent/teacher training threatened 47.09% of outcomes, due to a prevalence of parent-training studies which included outcomes derived from parent-child interactions. Because many NDBIs were described as specifically targeting symbolic play, early social communication, and language, researcher-created measures of these skills were coded as proximal to intervention targets. Thus, nearly half (47.59%) of outcomes in NDBI studies were categorized as proximal. Nearly a quarter (22.22%) were categorized as generalized, 52.41% were categorized as potentially context-bound, and another quarter (26.36%) were categorized as context-bound. Only 7.25% of outcomes were subject to bias from high attrition.

Sensory-based intervention studies. Figure 5 reflects quality indicator ratings for sensory-based intervention studies that were included in summary effect size estimation. All of the seven studies included in effect size estimation were RCTs. Since language was the only

outcome category for which there were a sufficient number of sensory-based intervention studies to permit summary effect size estimation, the following outcome-level quality indicator ratings apply only to the language outcomes ($n = 13$) tracked in these studies. Detection bias was rated as high for nearly half (46.15%) of all language outcomes. Nearly a third (30.77%) of all outcomes were based on parent/teacher report, and these same outcomes were also subject to CME related to parent training. The overwhelming majority (92%) of outcomes were categorized as distal, because few sensory-based interventions were described as directly targeting language. Nearly a third (30.77%) were categorized as generalized, 53.86% were categorized as potentially context-bound, and 15.38% of outcomes were categorized as context-bound. Attrition bias was rated as high for 15.38% of outcomes.

TEACCH studies. Figure 5 illustrates quality indicator ratings for studies of TEACCH that were included in summary effect size estimation ($n = 6$). Only two (33%) of these studies were RCTs. Detection bias was rated as high for the majority (81.81%) of outcomes, and this was largely driven by an over-reliance on parent/teacher report, from which 77.27% of outcomes were derived. CME related to parent/teacher training threatened half (50%) of all outcomes. Given that the explicit individual intervention targets of TEACCH were not thoroughly described, and that the majority of outcomes were taken from standardized parent/teacher reports, almost all (95.45%) outcomes were assumed to be distal. Nearly half (45.45%) of outcomes were categorized as generalized, half (50%) were categorized as potentially context-bound, and the remaining 4.54% were categorized as context-bound. None (0%) of the studies reported substantial attrition.

Technology-based intervention studies. Figure 5 illustrates quality indicator ratings for studies of technology-based interventions. Of ten technology-based intervention studies included

in summary effect estimation, eight (80%) were RCTs. Detection bias was rated as high for 64.28% of all outcomes. Over a third (38.1%) of outcomes were taken from parent/teacher report. CME related to parent/teacher training threatened 30.95% of outcomes. Over half (53.57%) of outcomes were categorized as distal. Nearly a third (30.95%) of outcomes were categorized as generalized, nearly half (47.62%) were categorized as potentially context-bound, and 21.43% were categorized as context-bound. Bias related to substantial attrition was rated as high for 15.38% of outcomes.

Summary Effects by Intervention and Outcome Type

Summary effects across all studies without consideration of quality indicators.

Figure 6 reflects summary effect size estimates within interventions and outcome types. These estimates were derived using all available effect sizes, both from quasi-experimental studies and RCTs. Summary effects were computed when effect sizes associated with a given outcome and intervention type were available from at least five independent participant samples. Thus, we were able to estimate the summary effects of behavioral interventions on adaptive outcomes, cognitive outcomes, language outcomes, motor outcomes, social communication outcomes, social emotional/challenging behavior outcomes, and outcomes quantifying broader autism symptomatology. Summary effects for behavioral interventions across outcome types ranged from 0.24 to 0.46 and were all statistically significant. For developmental interventions, only language and social communication outcomes were measured in a sufficient number of studies to permit the estimation of summary effects. The summary effects of developmental interventions on these outcomes were 0.06 and 0.30, respectively, and only the estimate for social communication was statistically significant. The summary effects of NDBIs were separately estimated for adaptive outcomes, cognitive outcomes, language outcomes, play outcomes,

restrictive and repetitive behaviors, social communication outcomes, social emotional/challenging behavior outcomes, and outcomes that quantified broader autism symptomatology. These summary effects ranged from -0.01 to 0.35. The summary effect estimates of NDBIs on cognition, language, play, and social communication outcomes were statistically significant. For sensory-based interventions, only language outcomes were measured in a sufficient number of studies to permit the estimation of summary effects. This summary effect estimate was 0.28, and was not significant. For TEACCH, summary effects could be generated only for social communication outcomes. This summary effect estimate was -0.11 and was not significant. For technology-based interventions, the most frequently tracked outcomes were social communication and social emotional/challenging behavior. Summary effect estimates for these outcomes were 0.05 and 0.42, respectively, and neither were significant.

Summary effects from RCTs. Figure 7 reflects summary effect size estimates derived exclusively from outcomes extracted from RCTs, according to intervention and outcome type. There were not enough RCTs of behavioral interventions to permit summary effect estimation for any outcome type. For developmental interventions, the summary effect across social communication outcomes from RCTs was 0.27 and significant. For NDBIs, a sufficient number of RCTs permitted the estimation of summary effects on cognition, language, play, and social communication. These estimates ranged from 0.18 to 0.42, and were significant for language, play, and social communication. All of the studies tracking the effect of sensory-based interventions on language outcomes were RCTs. Therefore, this summary effect estimate remains identical to that of the initial model. For technology-based interventions, there were only enough RCTs to permit estimation of a summary effect for social communication. This was 0.06

and was not significant. There were no RCTs examining the effects of the TEACCH intervention on any outcome.

Summary effects from RCTs excluding outcomes from caregiver reports. Figure 8 reflects summary effects estimated exclusively from outcomes that were extracted from RCTs and that were not based on caregiver report. For developmental interventions, a sufficient number of studies and outcomes permitted the estimation of a summary effect for social communication, which was 0.31 and statistically significant. For NDBIs, summary effect estimation was possible for cognition, language, play, and social communication. These effects ranged from 0.18 to 0.47, and were significant in the cases of play and social communication outcomes. For sensory-based interventions, summary effect estimation was possible for language only. This estimate was 0.28 and was not significant.

Summary effects from RCTs excluding all outcomes subject to a high threat of detection bias. Figure 9 reflects summary effects estimated exclusively from outcomes that were extracted from RCTs where assessors were unaware of group assignment. There were enough studies/effect sizes of this nature to permit estimation of the summary effects of NDBIs on language and social communication only. These estimates were 0.17 and 0.17, respectively, and were not significant.

Publication Bias Analyses

Funnel plots and Egger's test results are included in the supplementary materials accompanying this report. Corrected *p*-values for Egger's tests for funnel plot asymmetry were significant for adaptive and social communication outcomes from studies of NDBIs, suggesting that publication bias may have threatened these summary estimates.

Moderator Analyses

Meta-regression analyses across the entire dataset suggested that summary effects were significantly larger for outcomes that were proximal compared to those that were distal ($\beta = 0.171, p = 0.024$). Boundedness was also a significant source of effect size variance; effect sizes coded as generalized ($\beta = -0.170, p = .076$) were smaller than those coded as potentially context-bound or context bound ($\beta = -0.115, p = 0.22$).

Discussion

The purpose of this study was to locate, evaluate, and synthesize all available quasi-experimental and RCT investigations of nonpharmacological interventions for children with ASD in terms of methodological quality and summary effect. Results suggest that some intervention approaches show promise for improving a range of outcomes, while others have amassed relatively limited evidence of effectiveness to date. The number of RCT investigations in this area have increased precipitously, but low methodological rigor remains a concern.

Promising Intervention Types

We consider intervention types for which significant summary effects were shown for at least one outcome, when two important quality indicators were taken into account (randomization and abstention from using caregiver reports) to be ‘promising’. NDBIs and developmental interventions meet these criteria.

NDBI. This is the first paper to report summary effects of NDBIs since the 2015 consensus paper that established this new category of intervention as a blend of traditional behavioral and developmental approaches. By far, NDBIs have emerged as the intervention type most supported by evidence from RCTs. These studies suggest NDBIs may be particularly useful for supporting development of social communication, language, and play skills. Studies of NDBIs were also the least likely to rely on caregiver report as a primary index of intervention

effectiveness. However, we note that when outcomes subject to all forms of detection bias were excluded from summary effect estimation, there was no category of outcomes for this intervention type that reached significance. In addition, our results suggest that publication bias may have threatened overall summary estimates for adaptive and social communication outcome types. However, asymmetry in these funnel plots may also be due to other methodological design flaws, such as the presence of detection bias.

Developmental. Evidence suggests that developmental interventions may be particularly effective for supporting the acquisition of social communication skills, which represents a core challenge for young children with ASD. This conclusion is supported even when outcomes from quasi-experimental studies and caregiver report are excluded. However, a substantial portion of outcomes were subject to high detection bias due to interaction partners or assessors that were aware of group assignment. When these outcomes are excluded, the remaining studies are too few in number to permit summary effect estimation for any outcome type. A key assumption of developmental interventions is that targeted gains in social communication will facilitate cascading developments in the domain of language. This assumption was not supported by our meta-analysis, as the summary effect of developmental interventions on language outcomes was not significant. However, we did locate compelling evidence suggesting that early targeted improvements in the synchrony of parent-child interactions can yield longitudinal improvements in the core challenges associated with ASD, which are detectable with standardized, independently administered assessments (Green et al., 2010; Pickles et al., 2016). Green and colleagues (2010) study of the Preschool Autism Communication Trial (PACT) supports the notion that proximal changes effected by intervention can facilitate long-term change in developmentally distal outcomes, even in the absence of continued intervention. It also provides

an example of methodological rigor to which the field should aspire, as it employed random assignment, pre-registered analyses, independent evaluators, and clearly defined proximal and distal outcomes.

Intervention Types with Some Evidence of Effectiveness

Behavioral. Behavioral intervention, specifically EIBI and related variants, is the most commonly recommended intervention approach for children with ASD, with many states specifying behavioral interventions explicitly in insurance coverage mandates (“Autism and insurance coverage,” 2018). Indeed, the large number of behavioral intervention studies ($n = 27$) that met our search criteria also suggests this is the most studied intervention approach for this population. Considered as a whole, without regard to quality of evidence, these studies support the effectiveness of behavioral interventions for improving a wide range of outcomes for children with ASD. However, only a fraction of past studies exploring the effects of traditional behavioral interventions were RCTs, and the majority of outcomes contributing to summary effect sizes were taken from caregiver report. Thus, the relatively low quality of this set of intervention literature limits our confidence in the accuracy of the summary effect sizes estimated in the initial model. A notable exception is the sole RCT which examined the effects of EIBI on standardized measures of cognition and language administered by independent evaluators (Smith, Groen, & Wynn, 2000). Though the positive results of this study are encouraging, they have persisted without replication for nearly 20 years. The dramatic increase in published RCTs in the intervening years since this study’s publication stand as proof that high quality group experimental investigations of autism-specific interventions are both possible and necessary in order to to unquestionably establish the effectiveness of interventions that are so routinely recommended. In the meantime, clinicians are encouraged to expand their knowledge

and skills to include naturalistic approaches that center the principles of early childhood development. States with insurance mandates that explicitly cover traditional behavioral interventions should furthermore revise their policies to also include NDBI and developmental approaches, given that these approaches have now accrued substantial evidence for effects in young children on the autism spectrum from recently-published RCTs.

Intervention Types with Little Evidence of Effectiveness

Sensory. Several previous systematic reviews have concluded that sensory-based interventions have amassed little evidence supporting their effectiveness to date (e.g., Barton, Reichow, Schnitz, Smith, & Sherlock, 2015; Case-Smith, Weaver, & Fristad, 2015). Our results are consistent with these conclusions. Relatively few group design studies of sensory-based interventions specifically focused on young children with ASD (i.e., with a mean age < 8 years) were located. Furthermore, there were not a sufficient number of studies measuring and reporting sensory outcomes in a manner that permitted extraction of effect size information and estimation of the summary effect of this intervention approach on what would presumably be the most proximal outcome (i.e., improvements in sensory function). This is particularly concerning in light of the fact that sensory differences are highly prevalent in this population (e.g., Ausderau, Sideris, Furlong, Little, Bulluck, & Baranek, 2014; Ben-Sasson, Hen, Fluss, Cermak, Engel-Yeger, & Gal, 2009; Leekam, Nieto, Libby, Wing, & Gould, 2007) and have been found to be associated with some aspects of child stress (Corbett, Schupp, Levine, & Mendoza, 2009). Unfortunately, across all included studies, we found no evidence that *any* intervention type had the potential to influence sensory outcomes in children with ASD. When we were able to estimate summary effects of sensory-based interventions, as was the case for language outcomes, the relative paucity of studies limited the precision of our estimates. Though the summary effect

estimate for sensory-based interventions on language outcomes is similar in magnitude to those of behavioral and NDBI approaches, this estimate is surrounded by a much wider confidence band, which overlaps with zero (i.e., the effect is not significant).

It should be noted that our category of sensory-based interventions was broad and included intervention approaches as distinct as Sensory Integration Therapy, Tomatis Sound Therapy™, and music therapy. The heterogeneity of these intervention approaches may limit the conclusions that can be drawn from this summary effect size estimate, as the theoretical underpinnings and clinical procedures do vary across approaches. It may be useful to consider the evidence for each of these intervention approaches separately, though the limited number of studies for each prevented us from computing subgroup effect sizes here. However, we did not come across any noteworthy high quality studies that suggested that any of the aforementioned intervention approaches had markedly positive effects on outcomes (though see Schaaf et al., 2014 which unfortunately did not report outcome data in a manner that would permit derivation of effect size information for synthesis). We did locate two exceptionally high quality studies demonstrating null effects of two sensory-based interventions, music therapy (Bieleninik et al., 2017) and auditory stimulation (Corbett et al., 2008). Therefore, our conclusion that there is limited high quality evidence to date to support sensory-based interventions for young children with ASD is based on our quantitative findings as well as our more fine-grained qualitative observations about this set of literature. Given that sensory features are now a core diagnostic criteria of ASD (APA, 2013), and given the already widespread implementation of sensory-based interventions for this population (e.g., Goin-Kochel, Mackintosh, & Myers, 2009; Schaaf & Case-Smith, 2014), we suggest that more rigorous research of these interventions be conducted to precisely determine their effects for children with ASD.

TEACCH. Though TEACCH was among the first interventions designed specifically for individuals with ASD, it also remains relatively under-studied compared to several other intervention approaches geared towards this population. Few eligible studies of TEACCH were located, and most were quasi-experimental. This may be because TEACCH is often conceptualized as a classroom wide intervention, necessitating large, cluster-randomized trials that are substantially more expensive to implement than clinically-based RCTs. The negative summary effect estimated across these studies suggests that there is limited evidence to support the effectiveness of TEACCH for improvement of social communication skills, and almost no evidence to support the effectiveness of TEACCH for the improvement of other core and related symptoms of ASD.

Technology-based interventions. Although assistive technology is an important support that must be accessible to autistic individuals, early interventions mediated *entirely* through technology have little evidence to support their effectiveness for improving social communication or social emotional outcomes in children with ASD. Both of the summary effect sizes for these outcome types had confidence intervals which included zero. The majority of technology-based interventions represented in this meta-analysis were DVDs or video games that targeted social emotional learning and social communication skills. The limited effectiveness of these interventions may be attributable to the near or total absence of a human interaction partner in these intervention contexts. Though technological supports have characteristics that might make them particularly useful to autistic people (e.g., predictable formats of information delivery, self-paced usage, highly motivating), these supports likely need to be integrated into interpersonal interactions, which could include computer-mediated interpersonal interactions, rather than replacing interaction partners entirely in learning situations. This may be particularly

true when the targeted developmental achievements are social in nature. In fact, the integration of technological supports into other interaction-based interventions is an approach that is supported by high-quality studies. For example, Kasari and colleagues (2014) integrated speech generating devices (SGD) into their JASP-EMT early intervention approach, and found gains on a variety of communication outcomes for preschoolers who were initially minimally verbal, compared to those that received the same intervention without use of the SGD. In this study, technology was integrated into an already well-developed intervention, that had amassed some degree of empirical support.

This may be a sensible path forward for conceptualizing the utility of new technologies for early intervention. That is, technology may be most useful when it is integrated into previously developed and validated approaches as a means to expand the populations of children with ASD for whom the intervention is accessible, rather than as an intervention in its own right. In this regard, it is important to consider that the ultimate use of technology is usually separable from the means by which children are taught to use it, so even the most intuitively designed technologies will still need to be paired with a validated teaching approach to ensure that children are able to learn to use the technology in a meaningful way.

Animal-assisted interventions. Although we did locate studies of animal-assisted interventions, there were too few to permit estimation of summary effect sizes for any outcomes. The two interventions represented in these studies were EAAT and canine assistance. Several of these studies relied on caregiver report to index change, and two were flagged for possible unreported conflicts of interest, as the authors currently provide the interventions in question for profit (Bass, Duchowny, & Llabre, 2009; Page, 2012). Therefore, there is little quality evidence

to support the effectiveness of animal-assisted interventions for any outcomes for children with ASD at this time.

Issues Related to Quality Indicators

The results of this study indicate that study quality remains an issue plaguing intervention research in young children with ASD. Three issues appear especially important to point out, including the preponderance of quasi-experimental group designs, reliance on caregiver/teacher report, and correlated measurement error due to interaction partners or assessors who participated in the intervention.

Although it is well established that randomized controlled trials offer the best protection against alternative explanations for intervention effects, quasi-experimental studies continue to be relied upon in autism intervention research. There are some circumstances wherein quasi-experimental methods may be appropriate, such as studies aiming to move established interventions into community settings where groups are already intact and randomizing participants would be prohibitively costly (e.g., Vivanti et al., 2014). However, our results suggest that we do not yet have intervention types that can be considered ‘established’ to an extent that would warrant this strategy. Since there were too few studies to permit the estimation of summary effects once study design and performance bias were taken into account, we suggest that researcher and funding resources should continue to focus on establishing study efficacy using the highest quality designs.

Another area of particular concern is continued reliance on parent/teacher report. These measures are nearly impossible to administer in such a way that the respondent is unaware of the child’s participation in an intervention. Indeed, research has shown that when caregivers complete such measures, an intervention ‘effect’ will be demonstrated if they believe their child

is receiving an intervention even when no intervention has actually occurred (Jones, Carberry, Hamo, & Lord, 2017). We therefore suggest that early intervention researchers should not rely on such measures, and instead seek alternative measurement systems that can be administered and scored by assessors who are unaware of group assignment.

Finally, correlated measurement error that occurs when parents or teachers are trained in an intervention and also participate as assessors is a common threat to validity that has received little attention from the field. Continued use of observational measures taken from interactions with trained caregivers may be fruitful for mediation analyses, in order to verify that post-treatment group differences in developmentally distal and generalized outcomes are explained at least in part by changes in reciprocal interactions with caregivers within the context of intervention. However, researchers should recognize that these measures are biased in favor of the intervention group, and should therefore not rely on them as a primary index of intervention effects. Researchers should also employ valid, standardized, independently-administered assessments as primary outcomes whenever possible. While changes in interactions between a trained caregiver and child may be important to measure if those interactions are expected to be the ‘mechanism’ through which the child achieves later developmental milestones, these interactions may not themselves index improvements in the child’s interactional repertoire. If researchers consider interactions with a familiar person as the most valid context for outcome assessment, they can avoid this threat to validity by relying on observational measures taken from interactions with familiar but untrained interaction partners (e.g., untrained teachers, untrained parents, untrained siblings, or untrained peers). Use of untrained interaction partners that are also naive to group assignment will further help researchers address the added threat of detection bias.

Understanding Intervention Outcomes - Boundedness and Proximity

Replicating previous research syntheses (Yoder et al., 2013; Fuller & Kaiser, 2019) and confirming our hypotheses, effect sizes were larger for indices of context-bound behaviors as compared to generalized child characteristics. This finding confirms that interventions (broadly considered) produce larger effects on behaviors that are potentially bound to the treatment context, which are likely easier to change, than on more highly generalized characteristics of young children with ASD. In certain circumstances, context-bound behavior change may be considered important. For example, if a study aims to improve children's classroom engagement, many would consider it acceptable if these effects did not generalize beyond the classroom, as the effects are likely only relevant in classroom contexts.

However, many stakeholders may expect interventions aiming to improve child characteristics associated with longer-term development (e.g., social communication) to produce gains that generalize to contexts beyond intervention settings. If developmentally important effects cannot be demonstrated outside intervention settings, it is unlikely that they will continue to be a part of the child's behavioral repertoire, in any context, once the intervention has stopped. Unfortunately, researchers do not always indicate whether their measurement system was restricted to detecting context-bound behaviors, or if it was able to detect gains in generalized child characteristics. We encourage researchers to make this distinction clear when presenting their study design, and when describing potential limitations in the case of studies that exclusively examine context-bound behavior change.

Our hypothesis was also confirmed in regards to proximity; effect sizes for proximal outcomes were larger than effect sizes for distal outcomes. Parallel to our findings on boundedness, this indicates that interventions are more effective at achieving gains on outcomes

that reflect what was directly addressed in the intervention than gains on outcomes that are broader or beyond what was directly taught. Evidence of distal effects provide some evidence that the intervention is tapping into a developmental pathway, which can give researchers confidence that the intervention will continue to influence children's development after the intervention period is over.

There are some caveats to our approach in categorizing outcome proximity. One is that this concept is likely more accurately described as continual rather than binary. There are degrees of proximity and distality that we were not able to capture by restricting our coding to only two categories. A second caveat is that we were limited to the information about the intervention provided by study authors, which was often quite sparse. When delineating the focus of the intervention, authors did not always clarify if they were describing the immediate targets of the intervention, or a developmentally downstream target. Similarly, many studies did not offer a detailed description of the intervention, which hampered our ability to determine which outcomes were directly addressed by intervention procedures. Finally, proximity and distality are conflated with type of measurement system. Norm-referenced, standardized measures generally assess broad contexts which by definition cannot be directly targeted by intervention procedures and are therefore categorized as distal. On the other hand, observational measures of particular behaviors are often designed by researchers specifically to detect the most immediate effects of intervention (e.g., observational measures of joint engagement for interventions that seek to increase the amount of time children spend jointly engaged), which would be categorized as proximal. Thus, proximal measures may be more sensitive to change than distal measures, while distal measures are likely more construct valid than researcher-created proximal measures.

Interpreting Findings in Light of the Exclusion of Evidence from SSDs

It should be reiterated that we exclusively synthesized findings from randomized and nonrandomized group design studies of interventions for children with ASD. By excluding studies with single group pretest-posttest designs and SSDs, we have omitted a substantial body of research that has been used to draw conclusions about evidence-based practice, particularly in regards to the effectiveness of behavior analytic approaches. In fact, as of 2015, the majority of the available studies of intervention techniques for children with autism employed SSD (Wong et al., 2015), though our review and other reviews published since attest to the recent precipitous increase in group design literature published in this field (French & Kennedy, 2017).

Our decision to exclude SSDs from this meta-analysis was rooted primarily in the lack of adequate and agreed upon effect size metrics for synthesizing effects (Kratochwill et al., 2013). However, we believe there are additional insights to be gained from limiting our conclusions specifically to evidence offered by group design studies. Though SSDs are well-equipped to identify effective techniques for teaching specific targeted skills, group design studies are particularly useful for determining whether interventions can facilitate gains in generalized development. The repeated measurement that is a hallmark of SSDs may allow investigators to understand variability in specific behaviors associated with careful and controlled changes in the independent variable, but it limits reliance on validated standardized assessments as outcome measures. Such assessments, though often time consuming to administer, are likely better equipped to tap improvements in generalized development than researcher-created operationalizations of specific behaviors. Thus, if we wish to evaluate whether intervention facilitates developmental progress in young children with autism on average, an evaluation of group design studies may, arguably, be more methodologically suited for this purpose. However,

even though group design studies may be preferable in this regard, ours and other recent work has shown that a substantial portion of the outcome measures used in clinical trials were overly specific to the intervention context and targets (Provenzani et al., 2019). Thus, fragmented measurement approaches continue to limit the conclusions that can be drawn regarding the effectiveness of autism interventions, both in SSDs and group design studies. This remains a limitation, both for the body of evidence as a whole, and our conclusions here.

Recommendations for Primary Intervention Research

Given the results of this series of meta-analyses, we propose several recommendations. While our confidence in summary effect estimates for *any* intervention type is hindered by a lack of high quality studies, we do have single examples of studies that meet the majority of quality indicators (e.g., Green et al., 2010). This suggests that designing a high-quality study is not an unreachable challenge for early intervention researchers. It would perhaps incentivize future high-quality research if funding agencies held investigators to a higher standard and required basic quality features such as randomized trials and measurement systems that can be administered in such a way that assessors remain naive to treatment status. At the very least, caregiver and teacher reports should likely be discarded altogether, as it is already clear that they introduce bias and render findings largely uninterpretable (Jones et al., 2017). For some domains, this may mean that new measures will need to be developed and validated that are low-cost to administer and adequately sensitive to change.

A second recommendation, also related to measurement systems, is that researchers should provide detailed descriptions of each measure (especially if they are researcher-created), and the assessment process in which each measure is used. This will allow for an adequate assessment of the kinds of bias introduced or avoided by particular approaches to measurement,

and will allow for a determination of whether measures are capturing context-bound behavior change or generalized characteristics. To make this latter determination, aspects such as the measurement context, who administered the measure, and the materials and activities used during measurement should be made clear.

Third, we were quite struck by how little information many studies contained in regards to the intervention that was tested. Though it is not necessary that every study on a given intervention provide minute detail of the procedures, it would be helpful if there were at least one manualized protocol available for each intervention that describes the full set of strategies and activities involved in implementing the intervention. This would encourage independent replication of intervention studies, and would allow for a determination of whether the outcomes measured were proximal or distal to the intervention procedures. To make this distinction, researchers need to go beyond describing the aims of the intervention- they need to specifically describe the protocol in such a way that the immediate outcomes of implementing the intervention are readily discernible.

Fourth, fifty studies were excluded because relevant effect size information was not published or extractable. In many cases, this was due to exclusive reporting of change scores or post-intervention means adjusted for various baseline covariates, which should not be meta-analyzed alongside standardized mean differences extracted from unadjusted means (Deeks, Higgins, & Altman, 2008). Though we contacted authors in every case wherein studies were less than 10 years old, many failed to respond. Therefore, we recommend that authors reporting results that control for covariates include unadjusted means and SDs in supplementary materials, to facilitate future attempts at meta-analysis.

Finally, we suggest that an “optimal” intervention design would include paired proximal and distal measures (or perhaps even include a third, far distal measure) that are expected to be developmentally connected and malleable to change. The proximal measure should be selected to capture the immediate effects of the intervention, while the distal measure should be selected to measure effects hypothesized to be developmentally downstream from proximal effects. Mediation analyses, in which the proximal measure is the mediating variable and the distal measure is the outcome variable, could then confirm whether the proposed developmental pathway between proximal and distal effects was activated by participation in the intervention. This would allow for a better understanding of the mechanisms or ‘active ingredients’ through which interventions achieve cascading developmental gains.

Limitations and Future Meta-analytic Research

There are at least three limitations to consider when interpreting the results of this study. First, despite our best efforts, we were unable to collect any unpublished effect sizes or datasets apart from dissertations and theses. This could mean that the effect size estimates presented here are larger than the ‘true’ effects (an interpretation supported by inspection of funnel plots). Our attempts to gather unpublished effect sizes included searching NIH, NDAR, and IES databases, and requesting data directly from investigators who were reported to have received funding for group design intervention research in children with ASD. However, we did not receive any unpublished data from any researchers, suggesting there may be reticence among researchers to share their unpublished data. This is unfortunate, as access to unpublished data is critical for accurately estimating effect sizes, and accurately assessing the ‘state of the science’. Further, data sharing practices are critical to ensure replicability of findings (Nuijten, 2018).

A second limitation to consider is that there were too few studies to adequately synthesize effect sizes for all outcomes and intervention types. This was especially true when quality indicators were taken into consideration. Researchers will need to commit to conducting high-quality intervention research in order for future syntheses to accurately draw conclusions about intervention effectiveness for outcomes of interest in children with autism.

Finally, the heterogeneity of variables within each “outcome type” and treatments represented within each “intervention type” may limit the interpretability of our summary effect estimates. While we note that variables and intervention approaches were similar enough to be categorized with high reliability – kappa coefficients for outcome type and intervention type coding were 0.862 and 0.907, respectively – categorization of items that differ on a continuum will always result in the loss of information, and this information may be important for understanding key components that drive intervention effects. For example, the same intervention provided with different intensities (i.e., number of hours per week) may yield different effects. Similarly, intervention effects may differ for variables that share a domain but are distinct (e.g., social communication variables such as responding to joint attention and initiating joint attention). More fine-grained analyses within each outcome type could allow us to answer questions about putative moderators, as well as to calculate subgroup effect sizes for identical outcome types across studies (e.g., Vineland scores), or identical interventions (e.g., PECS) as the literature base on treatment effects in children with ASD continues to grow.

Conclusions

The current study differs from existing reviews on intervention in children with ASD in two important ways. First, this study is one of few attempts to consider all intervention types and intervention outcomes as broadly as possible. This allows us to report the state of the science in

regards to which interventions have accrued the most convincing evidence of effectiveness for young children with ASD, and to report on the full range of outcomes that these interventions are able to influence. Second, this study accounts for rigorous quality criteria that are common considerations in other areas of psychology, but that are applied less often to evaluations of autism research (e.g., Reichow, Volkmar, & Cicchetti, 2008). Finally, several syntheses that are similar to ours in scope consider some of the design factors of included studies in order to classify intervention types according to levels of evidence (e.g., Wong et al., 2015). However, these syntheses have not provided an examination of intervention effects according to characteristics of the outcome variable, which prevents researchers from drawing conclusions in regards to whether interventions are able to influence generalized characteristics that extend beyond the skills directly targeted by the interventions. Our findings echo recent sentiments from intervention researchers who are heartened by the relative increase in RCTs over the past 15 years, but also raise concerns in regards to the availability of high quality study designs that reliably and consistently link established interventions with meaningful child outcomes (Charman, 2019).

Even given these concerns, the evidence base regarding intervention for children with ASD has been rapidly transforming. The last decade has seen the publication of over 100 group design studies of intervention, including at least 50 RCTs. These studies attest to the fact that access to intervention in early childhood can yield a range of positive outcomes for the children receiving it. NDBIs have emerged as a new intervention category with significant summary effects even when several quality indicators are taken into account. High quality studies also suggest that developmental intervention can improve some core challenges associated with ASD, particularly difficulties in social communication. Traditional behavioral intervention approaches

show some evidence of effectiveness, but methodological rigor remains a pressing concern for this body of research. There is little evidence to date, however, to support the effectiveness of several other interventions that are geared towards young children with autism, including TEACCH, sensory-based interventions, animal-assisted interventions, and interventions mediated solely through technology (though approaches that integrate technology, such as high-tech augmentative and alternative communication devices, into more established interventions appear promising). More high quality randomized-controlled trials that feature independently-administered assessments are needed to unquestionably establish the efficacy of any intervention type. Finally, researchers should consider the characteristics (i.e., distality and boundedness) of outcomes being tracked in intervention studies and interpret findings accordingly to permit a more ready assessment of the extent to which any particular treatment approach is likely to yield desired effects on developmental trajectories of young children affected by autism.

References

- *Adkins, K. W., Molloy, C., Weiss, S. K., Reynolds, A., Goldman, S. E., Burnette, C., ... & Malow, B. A. (2012). Effects of a standardized pamphlet on insomnia in children with autism spectrum disorders. *Pediatrics*, 130, S139-S144.
- *Aldred, C., Green, J., & Adams, C. (2004). A new social communication intervention for children with autism: Pilot randomised controlled treatment study suggesting effectiveness. *Journal of Child Psychology and Psychiatry*, 45, 1420-1430.
- American Psychological Association (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Autism and insurance coverage; state laws. (2018). Retrieved from <http://www.ncsl.org/research/health/autism-and-insurance-coverage-state-laws.aspx>
- Ayres, A. J. (1979). *Sensory integration and the child*. Los Angeles, CA: Western Psychological Services.
- Ayres, A. J. (2005). *Sensory integration and the child, 25th anniversary edition*. Los Angeles: Western Psychological Services.
- *Azarbehi, A. C. (2009). *The effectiveness of early intervention programs for children with autism: A one-year follow-up study of intense behavioural intervention versus preschool integration* (Unpublished doctoral dissertation). Academic Unit of Psychology of The University of New Brunswick, Fredericton, New Brunswick.
- Bahrack, L. E., & Todd, J. T. (2012). Multisensory processing in autism spectrum disorders: Intersensory processing disturbance as a basis for atypical development. In B. E. Stein and M. T. Wallace (Eds), *The new handbook of multisensory processes* (pp. 1453–1508). Cambridge, MA: MIT Press.

- *Bailey, B., Arciuli, J., & Stancliffe, R. J. (2017). Effects of ABRACADABRA literacy instruction on children with autism spectrum disorder. *Journal of Educational Psychology, 109*, 257-268.
- Baio J, Wiggins L, Christensen DL, et al. (2018). Prevalence of autism spectrum disorder among children aged 8 years — autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *Morbidity and Mortality Weekly Report Surveillance Summaries, 67*(No. SS-6):1–23.
- Barton, E. E., Reichow, B., Schnitz, A., Smith, I. C., & Sherlock, D. (2015). A systematic review of sensory-based treatments for children with disabilities. *Research in Developmental Disabilities, 37*, 64-80.
- Baron-Cohen, S., Golan, O., & Ashwin, E. (2012). Educational cognitive neuroscience: Designing autism-friendly methods to teach emotion recognition. *Neuroscience in Education: The Good, the Bad and the Ugly, Oxford University Press, Oxford*, 299-311.
- *Bass, M. M., Duchowny, C. A., & Llabre, M. M. (2009). The effect of therapeutic horseback riding on social functioning in children with autism. *Journal of Autism and Developmental Disorders, 39*, 1261-1267.
- Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The Annals of Statistics, 29*(4), 1165-1188.
- *Bieleninik, Ł., Geretsegger, M., Mössler, K., Assmus, J., Thompson, G., Gattino, G., ... & Suvini, F. (2017). Effects of improvisational music therapy vs enhanced standard care on symptom severity among children with autism spectrum disorder: The TIME-A randomized clinical trial. *Jama, 318*, 525-535.

- *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, 366-380.
- Boyd, B. A., Odom, S. L., Humphreys, B. P., & Sam, A. M. (2010). Infants and toddlers with autism spectrum disorder: Early identification and early intervention. *Journal of Early Intervention*, 32, 75-98.
- *Boyd, B. A., Watson, L. R., Reszka, S. S., Sideris, J., Alessandri, M., Baranek, G. T., ... & Belardi, K. (2018). Efficacy of the ASAP intervention for preschoolers with ASD: A cluster randomized controlled trial. *Journal of Autism and Developmental Disorders*, 48, 3144-3162.
- *Bremer, E., Balogh, R., & Lloyd, M. (2015). Effectiveness of a fundamental motor skill intervention for 4-year-old children with autism spectrum disorder: A pilot study. *Autism*, 19, 980-991.
- *Brian, J. A., Smith, I. M., Zwaigenbaum, L., & Bryson, S. E. (2017). Cross-site randomized control trial of the Social ABCs caregiver-mediated intervention for toddlers with autism spectrum disorder. *Autism Research*, 10, 1700-1711.
- Bruner, J. (1982). The organization of action and the nature of the adult-infant transaction. In E. Tronick (Ed.), *Social interchange in infancy: Affect, cognition, and communication*. Baltimore, MD: University Park Press.
- * 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, 741-752.

Cascio, C. J., Woynaroski, T., Baranek, G. T., & Wallace, M. T. (2016). Toward an interdisciplinary approach to understanding sensory function in autism spectrum disorder. *Autism Research*, 9, 920-925.

Case-Smith, J., Weaver, L. L., & Fristad, M. A. (2015). A systematic review of sensory processing interventions for children with autism spectrum disorders. *Autism*, 19, 133-148.

* Casenhiser, D. M., Binns, A., McGill, F., Morderer, O., & Shanker, S. G. (2015). Measuring and supporting language function for children with autism: Evidence from a randomized control trial of a social-interaction-based therapy. *Journal of Autism and Developmental Disorders*, 45, 846-857.

* Casenhiser, D. M., Shanker, S. G., & Stieben, J. (2013). Learning through interaction in children with autism: Preliminary data from a social-communication-based intervention. *Autism*, 17, 220-241.

* Chang, Y. C., Shire, S. Y., Shih, W., Gelfand, C., & Kasari, C. (2016). Preschool deployment of evidence-based social communication intervention: JASPER in the classroom. *Journal of Autism and Developmental Disorders*, 46, 2211-2223.

Charman, T. (2019). Editorial: Trials and tribulations in early autism intervention research. *Journal of the American Academy of Child and Adolescent Psychiatry*. Advance online publication.

* Cheldavi, H., Shakerian, S., Boshehri, S. N. S., & Zarghami, M. (2014). The effects of balance training intervention on postural control of children with autism spectrum disorder: Role of sensory information. *Research in Autism Spectrum Disorders*, 8, 8-14.

- *Chiang, C. H., Chu, C. L., & Lee, T. C. (2016). Efficacy of caregiver-mediated joint engagement intervention for young children with autism spectrum disorders. *Autism*, 20, 172-182.
- *Clionsky, L. N. (2012). *Child directed interaction training for young children with autism spectrum disorders: The impact on child language, social skills, adaptive skills, and problematic behaviors* (Unpublished doctoral dissertation). University of Florida, Gainesville, FL.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20, 37-46.
- *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, S145-S155.
- Corbett, B. A., Schupp, C. W., Levine, S., & Mendoza, S. (2009). Comparing cortisol, stress, and sensory sensitivity in children with autism. *Autism Research*, 2(1), 39-49.
- *Corbett, B. A., Shickman, K., & Ferrer, E. (2008). Brief report: The effects of Tomatis sound therapy on language in children with autism. *Journal of Autism and Developmental Disorders*, 38, 562-566.
- *D'Elia, L., Valeri, G., Sonnino, F., Fontana, I., Mammone, A., & Vicari, S. (2014). A longitudinal study of the TEACCH program in different settings: The potential benefits of low intensity intervention in preschool children with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 44, 615-626.
- *Dawson, G., Jones, E. J., Merkle, K., Venema, K., Lowy, R., Faja, S., ... & Webb, S. J. (2012). Early behavioral intervention is associated with normalized brain activity in young

- children with autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51, 1150-1159.
- *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, e17-e23.
- Dawson, G., & Zanolli, K. (2003). Early intervention and brain plasticity in autism. In M. Rutter (Ed.), *Autism: Neural bases and treatment possibilities* (pp. 266-280). London: Novartis.
- Deeks, J. J., Higgins, J. P., & Altman, D. G. (2008). Analysing data and undertaking meta-analyses. *Cochrane handbook for systematic reviews of interventions: Cochrane book series*, 243-296.
- *Drew, A., Baird, G., Baron-Cohen, S., Cox, A., Slonims, V., Wheelwright, S., ... & Charman, T. (2002). A pilot randomised control trial of a parent training intervention for pre-school children with autism. *European Child & Adolescent Psychiatry*, 11, 266-272.
- *Duifhuis, E. A., Den Boer, J. C., Doornbos, A., Buitelaar, J. K., Oosterling, I. J., & Klip, H. (2017). The effect of pivotal response treatment in children with autism spectrum disorders: A non-randomized study with a blinded outcome measure. *Journal of Autism and Developmental Disorders*, 47, 231-242.
- *Durand, V. M., Hieneman, M., Clarke, S., Wang, M., & Rinaldi, M. L. (2013). Positive family intervention for severe challenging behavior I: A multisite randomized clinical trial. *Journal of Positive Behavior Interventions*, 15, 133-143.
- *Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2007). Outcome for children with autism who began intensive behavioral treatment between ages 4 and 7: A comparison controlled study. *Behavior Modification*, 31, 264-278.

- *Eikeseth, S., Klintwall, L., Jahr, E., & Karlsson, P. (2012). Outcome for children with autism receiving early and intensive behavioral intervention in mainstream preschool and kindergarten settings. *Research in Autism Spectrum Disorders*, 6, 829-835.
- * Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2002). Intensive behavioral treatment at school for 4-to 7-year-old children with autism: A 1-year comparison controlled study. *Behavior Modification*, 26, 49-68.
- *Elder, L. M. (2011). *The relationship between parent training and parent-child interaction in autism* (Unpublished doctoral dissertation). University of Washington, Seattle, WA.
- * 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, 210-220.
- *ElGarhy, S., & Liu, T. (2016). Effects of psychomotor intervention program on students with autism spectrum disorder. *School Psychology Quarterly*, 31, 491-506.
- *Espanola, E. (2015). *A randomized comparison of two instructional sequences for imitation intervention for children with autism spectrum disorders* (Unpublished doctoral dissertation). Florida International University, Miami, FL.
- *Estes, A., Munson, J., Rogers, S. J., Greenson, J., Winter, J., & Dawson, G. (2015). Long-term outcomes of early intervention in 6-year-old children with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 54, 580-587.
- *Fava, L., Strauss, K., Valeri, G., D'Elia, L., Arima, S., & Vicari, S. (2011). The effectiveness of a cross-setting complementary staff-and parent-mediated early intensive behavioral intervention for young children with ASD. *Research in Autism Spectrum Disorders*, 5, 1479-1492.

- *Fernell, E., Hedvall, Å., Westerlund, J., Carlsson, L. H., Eriksson, M., Olsson, M. B., ... & Gillberg, C. (2011). Early intervention in 208 Swedish preschoolers with autism spectrum disorder. A prospective naturalistic study. *Research in Developmental Disabilities, 32*, 2092-2101.
- Ferster, C. B., & DeMeyer, M. K. (1962). A method for the experimental analysis of the behavior of autistic children. *American Journal of Orthopsychiatry, 32*, 89–98.
- Fisher, Z., Tipton, E., & Zhipeng, H. (2017). robumeta: Robust variance meta-regression (Version 2.0). Retrieved from <https://CRAN.R-project.org/package=robumeta>
- *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*, 673-682.
- Fletcher-Watson, S. (2018, June). Is early intervention compatible with neurodiversity? [Blog post]. Retrieved from <https://dart.ed.ac.uk/intervention-neurodiversity/>
- *Fletcher-Watson, S., Petrou, A., Scott-Barrett, J., Dicks, P., Graham, C., O'Hare, A., ... & McConachie, H. (2016). A trial of an iPad™ intervention targeting social communication skills in children with autism. *Autism, 20*, 771-782.
- *Fridenson-Hayo, S., Berggren, S., Lassalle, A., Tal, S., Pigat, D., Meir-Goren, N., ... & Golan, O. (2017). 'Emotiplay': A serious game for learning about emotions in children with autism: Results of a cross-cultural evaluation. *European Child & Adolescent Psychiatry, 26*, 979-992.
- French, L., & Kennedy, E. M. M. (2017). Annual research review: Early intervention for infants and young children with, or at-risk of, autism spectrum disorder: A systematic review. *Journal of child psychology and psychiatry, and allied disciplines, 59*, 444-456.

- Fuller, E. A., & Kaiser, A. P. (2019). The effects of early intervention on social communication outcomes for children with autism spectrum disorder: A meta-analysis. *Journal of Autism and Developmental Disorders*. Advance Online Publication.
- Gabriels, R. L., Agnew, J. A., Holt, K. D., Shoffner, A., Zhaoxing, P., Ruzzano, S., ... & Mesibov, G. (2012). Pilot study measuring the effects of therapeutic horseback riding on school-age children and adolescents with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 6, 578-588.
- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2012). irr: Various coefficients of interrater reliability and agreement. R package v. 0.84.
- Gernsbacher, M. A. (2017). Editorial perspective: The use of person-first language in scholarly writing may accentuate stigma. *Journal of Child Psychology and Psychiatry*, 58, 859-861.
- *Golan, O., Ashwin, E., Granader, Y., McClintock, S., Day, K., Leggett, V., & Baron-Cohen, S. (2010). Enhancing emotion recognition in children with autism spectrum conditions: An intervention using animated vehicles with real emotional faces. *Journal of Autism and Developmental Disorders*, 40, 269-279.
- *Grahame, V., Brett, D., Dixon, L., McConachie, H., Lowry, J., Rodgers, J., ... & Le Couteur, A. (2015). Managing repetitive behaviours in young children with Autism Spectrum Disorder (ASD): Pilot randomised controlled trial of a new parent group intervention. *Journal of Autism and Developmental Disorders*, 45, 3168-3182.
- *Green, J., Charman, T., McConachie, H., Aldred, C., Slonims, V., Howlin, P., ... & Barrett, B. (2010). Parent-mediated communication-focused treatment in children with autism (PACT): A randomised controlled trial. *The Lancet*, 375, 2152-2160.

- Green, V. A., Pituch, K. A., Itchon, J., Choi, A., O'Reilly, M., & Sigafoos, J. (2006). Internet survey of treatments used by parents of children with autism. *Research in Developmental Disabilities, 27*, 70–84.
- Greenspan, S. I. & Wieder, S. (2007). The developmental individual-difference, relationship-based (DIR/Floortime) model approach to autism spectrum disorders. In E. Hollander & E. Anagnostou (Eds.) *Clinical manual for the treatment of autism* (pp. 179-209). Arlington, VA, US: American Psychiatric Publishing, Inc.
- Grelotti, D. J., & Kaptchuk, T. J. (2011). Placebo by proxy. *BMJ (Clinical research ed.)*, 343, d4345.
- Grynszpan, O., Weiss, P. L., Perez-Diaz, F., & Gal, E. (2014). Innovative technology-based interventions for autism spectrum disorders: A meta-analysis. *Autism, 18*, 346-361.
- *Gulsrud, A. C., Kasari, C., Freeman, S., & Paparella, T. (2007). Children with autism's response to novel stimuli while participating in interventions targeting joint attention or symbolic play skills. *Autism, 11*, 535-546.
- *Gulsrud, A. C., Hellemann, G. S., Freeman, S. F., & Kasari, C. (2014). Two to ten years: Developmental trajectories of joint attention in children with ASD who received targeted social communication interventions. *Autism Research, 7*, 207-215.
- *Hardan, A. Y., Gengoux, G. W., Berquist, K. L., Libove, R. A., Ardel, C. M., Phillips, J., ... & Minjarez, M. B. (2015). A randomized controlled trial of Pivotal Response Treatment Group for parents of children with autism. *Journal of Child Psychology and Psychiatry, 56*(8), 884-892.
- Harrell, F. E. Hmisc: Harrell Miscellaneous (Version 4.1-1). Retrieved from <https://CRAN.R-project.org/package=Hmisc>

- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*, 39-65.
- *Hemdi, A., & Daley, D. (2017). The Effectiveness of a Psychoeducation Intervention delivered via WhatsApp for mothers of children with Autism Spectrum Disorder (ASD) in the Kingdom of Saudi Arabia: A randomized controlled trial. *Child: Care, Health and Development, 43*, 933-941.
- Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., ... & Sterne, J. A. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal, 343*, d5928.
- Houting, J. den. (2019). Neurodiversity: An insider's perspective. *Autism, 23*(2), 271–273.
- *Howlin, P., & Rutter, M. (1989). Mothers' speech to autistic children: A preliminary causal analysis. *Journal of Child Psychology and Psychiatry, 30*, 819-843.
- Howling, P., & Magiati, I. (2017). Autism spectrum disorder: Outcomes in adulthood. *Current Opinion in Psychiatry, 30*(2), 69-76.
- Hróbjartsson, A., Thomsen, A. S. S., Emanuelsson, F., Tendal, B., Hilden, J., Boutron, I., ... & Brorson, S. (2012). Observer bias in randomised clinical trials with binary outcomes: Systematic review of trials with both blinded and non-blinded outcome assessors. *BMJ, 344*, e1119.
- *Hudson, R. F., Sanders, E. A., Greenway, R., Xie, S., Smith, M., Gasamis, C., ... & Hackett, J. (2017). Effects of emergent literacy interventions for preschoolers with autism spectrum disorder. *Exceptional Children, 84*, 55-75.

- *Iadarola, S., Shih, W., Dean, M., Blanch, E., Harwood, R., Hetherington, S., ... & Smith, T. (2018). Implementing a manualized, classroom transition intervention for students with ASD in underresourced schools. *Behavior Modification*, 42, 126-147.
- *Ichikawa, K., Takahashi, Y., Ando, M., Anme, T., Ishizaki, T., Yamaguchi, H., & Nakayama, T. (2013). TEACCH-based group social skills training for children with high-functioning autism: A pilot randomized controlled trial. *BioPsychoSocial Medicine*, 7, 14.
- *Ingersoll, B. (2010). Brief report: Pilot randomized controlled trial of reciprocal imitation training for teaching elicited and spontaneous imitation to children with autism. *Journal of Autism and Developmental Disorders*, 40, 1154-1160.
- *Ingersoll, B. (2012). Brief report: Effect of a focused imitation intervention on social functioning in children with autism. *Journal of Autism and Developmental Disorders*, 42, 1768-1773.
- *Ingersoll, B., Wainer, A. L., Berger, N. I., Pickard, K. E., & Bonter, N. (2016). Comparison of a self-directed and therapist-assisted telehealth parent-mediated intervention for children with ASD: A pilot RCT. *Journal of Autism and Developmental Disorders*, 46, 2275-2284.
- *Jalili, M., Jahangiri, N., Yazdi, A. A., & Ashrafzadeh, F. (2014). The effects of imitative vs. cognitive methods on the speech development of children with autism. *Iranian Journal of Child Neurology*, 8, 37-46.
- *Jocelyn, L. J., Casiro, O. G., Beattie, D., Bow, J., & Kneisz, J. (1998). Treatment of children with autism: A randomized controlled trial to evaluate a caregiver-based intervention program in community day-care centers. *Journal of Developmental and Behavioral Pediatrics*, 19, 326-334.

- *Johnson, C. R., Turner, K. S., Foldes, E., Brooks, M. M., Kronk, R., & Wiggs, L. (2013). Behavioral parent training to address sleep disturbances in young children with autism spectrum disorder: A pilot trial. *Sleep Medicine, 14*, 995-1004.
- Jones, R. M., Carberry, C., Hamo, A., & Lord, C. (2017). Placebo-like response in absence of treatment in children with Autism. *Autism Research, 10*, 1567-1572.
- *Jouen, A. L., Narzisi, A., Xavier, J., Tilmont, E., Bodeau, N., Bono, V., ... & Muratori, F. (2017). GOLIAH (Gaming Open Library for Intervention in Autism at Home): A 6-month single blind matched controlled exploratory study. *Child and Adolescent Psychiatry and Mental Health, 11*, 17.
- *Kaale, A., Smith, L., & Sponheim, E. (2012). A randomized controlled trial of preschool-based joint attention intervention for children with autism. *Journal of Child Psychology and Psychiatry, 53*, 97-105.
- *Kaale, A., Fagerland, M. W., Martinsen, E. W., & Smith, L. (2014). Preschool-based social communication treatment for children with autism: 12-month follow-up of a randomized trial. *Journal of the American Academy of Child & Adolescent Psychiatry, 53*, 188-198.
- Kaiser, A. P. (1993). Parent-implemented language intervention: An environmental system perspective. In A. P. Kaiser & D. B. Gray (Eds.), *Enhancing children's communication: Research foundations for intervention* (Vol. 2, pp. 63-84). Baltimore: Brookes.
- Kaiser, A. P., & Roberts, M. Y. (2012). Parent-implemented enhanced milieu teaching with preschool children who have intellectual disabilities. *Journal of speech, language, and hearing research : JSLHR, 56*, 295-309.

- *Kalyva, E., & Avramidis, E. (2005). Improving communication between children with autism and their peers through the 'Circle of Friends': A small-scale intervention study. *Journal of Applied Research in Intellectual Disabilities*, 18, 253-261.
- *Kamps, D., Heitzman-Powell, L., Rosenberg, N., Mason, R., Schwartz, I., & Romine, R. S. (2016). Effects of Reading Mastery as a small group intervention for young children with ASD. *Journal of Developmental and Physical Disabilities*, 28, 703-722.
- Kasari, C. (2002). Assessing change in early intervention programs for children with autism. *Journal of Autism and Developmental Disorders*, 32, 447-461.
- *Kasari, C., Freeman, S., & Paparella, T. (2006). Joint attention and symbolic play in young children with autism: A randomized controlled intervention study. *Journal of Child Psychology and Psychiatry*, 47, 611-620.
- *Kasari, C., Gulsrud, A. C., Wong, C., Kwon, S., & Locke, J. (2010). Randomized controlled caregiver mediated joint engagement intervention for toddlers with autism. *Journal of Autism and Developmental Disorders*, 40, 1045-1056.
- *Kasari, C., Gulsrud, A., Paparella, T., Hellemann, G., & Berry, K. (2015). Randomized comparative efficacy study of parent-mediated interventions for toddlers with autism. *Journal of Consulting and Clinical Psychology*, 83, 554-563.
- Kasari, C., Kaiser, A., Goods, K., Nietfeld, J., Mathy, P., Landa, R., ... & Almirall, D. (2014). Communication interventions for minimally verbal children with autism: A sequential multiple assignment randomized trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 53, 635-646.

- *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*, 125-137.
- Kasari, C., & Smith, T. (2013). Interventions in schools for children with autism spectrum disorder: Methods and recommendations. *Autism, 17*, 254-267.
- *Keen, D., Couzens, D., Muspratt, S., & Rodger, S. (2010). The effects of a parent-focused intervention for children with a recent diagnosis of autism spectrum disorder on parenting stress and competence. *Research in Autism Spectrum Disorders, 4*, 229-241.
- Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism, 20*, 442-462.
- Knight, V., McKissick, B. R., & Saunders, A. (2013). A review of technology-based interventions to teach academic skills to students with autism spectrum disorder. *Journal of autism and developmental disorders, 43*, 2628-2648.
- Koegel, R. L., Koegel, L. K., & Carter, C. M. (1999). Pivotal teaching interactions for children with autism. *School Psychology Review, 28*, 576-594.
- Kolb B, & Gibb R. (2011). Brain plasticity and behaviour in the developing brain. *Journal of the Canadian Academy of Child & Adolescent, 20*, 265-276.
- *Kovshoff, H., Hastings, R. P., & Remington, B. (2011). Two-year outcomes for children with autism after the cessation of early intensive behavioral intervention. *Behavior Modification, 35*, 427-450.

- Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-case intervention research design standards. *Remedial and Special Education, 34*(1), 26-38.
- *LaGasse, A. B. (2014). Effects of a music therapy group intervention on enhancing social skills in children with autism. *Journal of Music Therapy, 51*, 250-275.
- *Landa, R. J., Holman, K. C., O'Neill, A. H., & Stuart, E. A. (2011). Intervention targeting development of socially synchronous engagement in toddlers with autism spectrum disorder: A randomized controlled trial. *Journal of Child Psychology and Psychiatry, 52*, 13-21.
- *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*, 687-693.
- *Lawton, K., & Kasari, C. (2012). Brief report: Longitudinal improvements in the quality of joint attention in preschool children with autism. *Journal of Autism and Developmental Disorders, 42*, 307-312.
- *Leaf, J. B., Leaf, J. A., Milne, C., Taubman, M., Oppenheim-Leaf, M., Torres, N., ... & Yoder, P. (2017). An evaluation of a behaviorally based social skills group for individuals diagnosed with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 47*, 243-259.
- *Lerna, A., Esposito, D., Conson, M., Russo, L., & Massagli, A. (2012). Social-communicative effects of the Picture Exchange Communication System (PECS) in autism spectrum disorders. *International Journal of Language & Communication Disorders, 47*, 609-617.

- *Lim, H. A. (2010). The effect of “Developmental speech and language training through music” on speech production in children with autism spectrum disorders. *Journal of Music Therapy, 47*, 2-26.
- Lipsey, M. W., & Wilson, D. B. (2001). *Applied social research methods series; Vol. 49. Practical meta-analysis*. Thousand Oaks, CA, US: Sage Publications, Inc.
- Lord, C., Bristol-Power M., Cafiero, J. M., Filipek, P. A., Gallagher, J. J., Harris, S. L., et al. (2001). *Educating Children with autism*. Washington, DC: National Academy Press.
- Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. L. (2012). *Autism Diagnostic Observation Schedule, second edition (ADOS-2) manual (Part I): Modules 1-4*. Los Angeles, CA: Western Psychological Services.
- Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*, 3-9.
- *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*, 803-812.
- *Mahoney, G., & Solomon, R. (2016). Mechanism of developmental change in the PLAY project home consultation program: Evidence from a randomized control trial. *Journal of Autism and Developmental Disorders, 46*, 1860-1871.
- *Malow, B. A., Adkins, K. W., Reynolds, A., Weiss, S. K., Loh, A., Fawkes, D., ... & Clemons, T. (2014). Parent-based sleep education for children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*, 216-228.

- *Mandell, D. S., Stahmer, A. C., Shin, S., Xie, M., Reisinger, E., & Marcus, S. C. (2013). The role of treatment fidelity on outcomes during a randomized field trial of an autism intervention. *Autism, 17*, 281-295.
- *McCalla, M. K. (2014). *Enhancing the social interactions of preschool children with autism: The effectiveness of a peer-mediated* (Unpublished doctoral dissertation). The University of Alabama at Birmingham, Birmingham, AL.
- *McEachin, J. J., Smith, T., & Ivar Lovaas, O. (1993). Long-term outcome for children with autism who received early intensive behavioral treatment. *American Journal of Mental Retardation, 97*, 359-372.
- Mesibov, G. B., Shea, V., & Schopler, E. (2004). *The TEACCH approach to autism spectrum disorders*. New York, NY: Springer Science & Business Media.
- Murad, M. H., Asi, N., Alsawas, M., & Alahdab, F. (2016). New evidence pyramid. *BMJ Evidence-Based Medicine, 21*, 125-127.
- *Murdock, L. C., Dantzler, J. A., Walker, A. N., & Wood, L. B. (2014). The effect of a platform swing on the independent work behaviors of children with autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities, 29*, 50-61.
- National Autism Center (2015). *Findings and conclusions: National standards project, phase 2*. Randolph, MA.
- *Nefdt, N., Koegel, R., Singer, G., & Gerber, M. (2010). The use of a self-directed learning program to provide introductory training in pivotal response treatment to parents of children with autism. *Journal of Positive Behavior Interventions, 12*, 23-32.
- Nuijten, M. B. (2018). Practical tools and strategies for researchers to increase replicability. *Developmental Medicine & Child Neurology*. Advance online publication.

- Odom, S.L., Boyd, B.A., Hall, L.J., Hume K. (2010). Evaluation of comprehensive treatment models for individuals with autism spectrum disorders. *J Autism Dev Disord.*, 40, 437.
- Oliver, P.C., Piachaud, J., Done, J., Regan, A., Cooray, S., Tyrer, P. (2002). Difficulties in conducting a randomized controlled trial of health service interventions in intellectual disability: Implications for evidence-based practice. *Journal of Intellectual Disability Research*, 46, 340-345.
- *Ozonoff, S., & Cathcart, K. (1998). Effectiveness of a home program intervention for young children with autism. *Journal of Autism and Developmental Disorders*, 28, 25-32.
- *Page, C. E. (2012). *The social and emotional benefits of therapeutic riding on children with autism spectrum disorder* (Unpublished doctoral dissertation). Institute of Transpersonal Psychology, Palo Alto, CA.
- *Pajareya, K., & Nopmaneejumrulers, K. (2011). A pilot randomized controlled trial of DIR/Floortime™ parent training intervention for pre-school children with autistic spectrum disorders. *Autism*, 15, 563-577.
- *Pan, C. Y. (2010). Effects of water exercise swimming program on aquatic skills and social behaviors in children with autism spectrum disorders. *Autism*, 14, 9-28.
- *Paul, R., Campbell, D., Gilbert, K., & Tsiouri, I. (2013). Comparing spoken language treatments for minimally verbal preschoolers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 43, 418-431.
- *Paynter, J., & Peterson, C. C. (2013). Further evidence of benefits of thought-bubble training for theory of mind development in children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, 7, 344-348.

- *Peters-Scheffer, N., Didden, R., Mulders, M., & Korzilius, H. (2013). Effectiveness of low intensity behavioral treatment for children with autism spectrum disorder and intellectual disability. *Research in Autism Spectrum Disorders*, 7, 1012-1025.
- *Pickles, A., Le Couteur, A., Leadbitter, K., Salomone, E., Cole-Fletcher, R., Tobin, H., ... & Green, J. (2016). Parent-mediated social communication therapy for young children with autism (PACT): Long-term follow-up of a randomised controlled trial. *The Lancet*, 388, 2501-2509.
- *Piravej, K., Tangtrongchitr, P., Chandarasiri, P., Paothong, L., & Sukprasong, S. (2009). Effects of Thai traditional massage on autistic children's behavior. *The Journal of Alternative and Complementary Medicine*, 15, 1355-1361.
- Provenzano, U., Fusar-Poli, L., Brondino, N., Damiani, S., Vercesi, M., Meyer, N., Rochetti, M., & Politi, P. (2019). What are we targeting when we treat autism spectrum disorder? A systematic review of 406 clinical trials. *Autism*, 1362361319854641.
- *Poslawsky, I. E., Naber, F. B., Bakermans-Kranenburg, M. J., van Daalen, E., van Engeland, H., & Van IJzendoorn, M. H. (2015). Video-feedback Intervention to promote Positive Parenting adapted to autism (VIPP-AUTI): A randomized controlled trial. *Autism*, 19, 588-603.
- *Powell, G., Wass, S. V., Erichsen, J. T., & Leekam, S. R. (2016). First evidence of the feasibility of gaze-contingent attention training for school children with autism. *Autism*, 20, 927-937.
- R Core Team. (2017). R: A language and environment for statistical computing (Version 3.4.1). R Foundation for Statistical Computing: Vienna, Austria. Retrieved from <https://www.R-project.org/>

- Reichow, B., Volkmar, F. R., & Cicchetti, D. V. (2008). Development of the evaluative method for evaluating and determining evidence-based practices in autism. *Journal of autism and developmental disorders*, 38(7), 1311-1319.
- *Reitzel, J., Summers, J., Lory, B., Szatmari, P., Zwaigenbaum, L., Georgiades, S., & Duku, E. (2013). Pilot randomized controlled trial of a Functional Behavior Skills Training program for young children with autism spectrum disorder who have significant early learning skill impairments and their families. *Research in Autism Spectrum Disorders*, 7, 1418-1432.
- *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, 418-438.
- Renty, J. O., & Roeyers, H. (2006). Quality of life in high-functioning adults with autism spectrum disorder: The predictive value of disability and support characteristics. *Autism*, 10(5), 511-524.
- *Rentz, E. A. (2015). *Autism spectrum disorders: Developmental, individual-difference, relationship-based (DIR)/Floortime outcomes* (Unpublished doctoral dissertation). Fordham University, Bronx, NY.
- *Rice, L. M., Wall, C. A., Fogel, A., & Shic, F. (2015). Computer-assisted face processing instruction improves emotion recognition, mentalizing, and social skills in students with ASD. *Journal of Autism and Developmental Disorders*, 45, 2176-2186.
- *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, 308-316.
- *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, 1553-1566.
- Robison, J. E. (2019). Talking about autism—Thoughts for researchers. *Autism Research*, 12, 1004–1006.
- Rogers, S. J., & Dawson, G. (2010). *Early Start Denver Model for young children with autism: Promoting language, learning, and engagement*. New York, NY: The Guilford Press.
- *Rogers, S. J., Hayden, D., Hepburn, S., Charlifue-Smith, R., Hall, T., & Hayes, A. (2006). Teaching young nonverbal children with autism useful speech: A pilot study of the Denver Model and PROMPT interventions. *Journal of Autism and Developmental Disorders*, 36, 1007-1024.
- *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, 1052-1065.
- *Ruble, L. A., Dalrymple, N. J., & McGrew, J. H. (2010). The effects of consultation on individualized education program outcomes for young children with autism: The collaborative model for promoting competence and success. *Journal of Early Intervention*, 32, 286-301.

- Rutter, M., LeCouteur, A., & Lord, C. (2003). *Autism Diagnostic Interview-Revised*. Los Angeles, CA: Western Psychological Services.
- *Sainato, D. M., Morrison, R. S., Jung, S., Axe, J., & Nixon, P. A. (2015). A comprehensive inclusion program for kindergarten children with autism spectrum disorder. *Journal of Early Intervention, 37*, 208-225.
- *Sallows, G. O., & Graupner, T. D. (2005). Intensive behavioral treatment for children with autism: Four-year outcome and predictors. *American Journal on Mental Retardation, 110*, 417-438.
- *Salt, J., Shemilt, J., Sellars, V., Boyd, S., Coulson, T., & McCool, S. (2002). The Scottish Centre for Autism preschool treatment programme. II: The results of a controlled treatment outcome study. *Autism, 6*, 33-46.
- *Sampanthavivat, M., Singkhwa, W., Chaiyakul, T., Karoonyawanich, S., & Ajpru, H. (2012). Hyperbaric oxygen in the treatment of childhood autism: A randomised controlled trial. *Diving and Hyperbaric Medicine, 42*, 128-33.
- *Scarpa, A., & Reyes, N. M. (2011). Improving emotion regulation with CBT in young children with high functioning autism spectrum disorders: A pilot study. *Behavioural and Cognitive Psychotherapy, 39*, 495-500.
- Schaaf, R. C., Benevides, T., Mailloux, Z., Faller, P., Hunt, J., Van Hooydonk, E., ... & Kelly, D. (2014). An intervention for sensory difficulties in children with autism: A randomized trial. *Journal of Autism and Developmental Disorders, 44*(7), 1493-1506.
- *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*, 249-258.

*Schertz, H. H., Odom, S. L., Baggett, K. M., & Sideris, J. H. (2018). Mediating parent learning to promote social communication for toddlers with autism: Effects from a randomized controlled trial. *Journal of Autism and Developmental Disorders*, 48, 853-867.

Schopler E., Reichler R., Rochen Renner B. The childhood autism rating scale. Western Psychological Services; 1988.

*Schreibman, L., & Stahmer, A. C. (2014). A randomized trial comparison of the effects of verbal and pictorial naturalistic communication strategies on spoken language for young children with autism. *Journal of Autism and Developmental Disorders*, 44, 1244-1251.

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, 2411–2428.

*Sheinkopf, S. J., & Siegel, B. (1998). Home-based behavioral treatment of young children with autism. *Journal of Autism and Developmental Disorders*, 28, 15-23.

*Shire, S. Y., Gulsrud, A., & Kasari, C. (2016). Increasing responsive parent–child interactions and joint engagement: Comparing the influence of parent-mediated intervention and parent psychoeducation. *Journal of Autism and Developmental Disorders*, 46, 1737-1747.

*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, 612-622.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological bulletin*, 86, 420.

*Silva, L., & Schalock, M. (2013). Treatment of tactile impairment in young children with autism: Results with qigong massage. *International Journal of Therapeutic Massage & Bodywork*, 6, 12-20.

*Silva, L. M., Schalock, M., Gabrielsen, K. R., Budden, S. S., Buenrostro, M., & Horton, G. (2015). Early intervention with a parent-delivered massage protocol directed at tactile abnormalities decreases severity of autism and improves child-to-parent interactions: A replication study. *Autism Research and Treatment*, 2015.

*Simpson, K., Keen, D., & Lamb, J. (2015). Teaching receptive labelling to children with autism spectrum disorder: A comparative study using infant-directed song and infant-directed speech. *Journal of Intellectual and Developmental Disability*, 40, 126-136.

*Simpson, K., Keen, D., & Lamb, J. (2013). The use of music to engage children with autism in a receptive labelling task. *Research in Autism Spectrum Disorders*, 7, 1489-1496.

* Smith, T., Eikeseth, S., Klevstrand, M., & Lovaas, O. I. (1997). Intensive behavioral treatment for preschoolers with severe mental retardation and pervasive developmental disorder. *American Journal on Mental Retardation*, 102, 238-249.

*Smith, T., Groen, A. D., & Wynn, J. W. (2000). Randomized trial of intensive early intervention for children with pervasive developmental disorder. *American Journal on Mental Retardation*, 105, 269-285.

- *Sokmun, S., Singh, S. J., & Vandort, S. (2018). The impact of Hanen More Than Words programme on parents of children with ASD in Malaysia. *Malaysian Journal of Health Sciences, 15*, 43-51.
- *Solomon, R., Van Egeren, L. A., Mahoney, G., Huber, M. S. Q., & Zimmerman, P. (2014). PLAY Project Home Consultation intervention program for young children with autism spectrum disorders: A randomized controlled trial. *Journal of Developmental and Behavioral Pediatrics, 35*, 475-485.
- *Spjut Jansson, B., Miniscalco, C., Westerlund, J., Kantzer, A. K., Fernell, E., & Gillberg, C. (2016). Children who screen positive for autism at 2.5 years and receive early intervention: A prospective naturalistic 2-year outcome study. *Neuropsychiatric Disease and Treatment, 12*, 2255-2263.
- *Srinivasan, S. M., Eigsti, I. M., Gifford, T., & Bhat, A. N. (2016). The effects of embodied rhythm and robotic interventions on the spontaneous and responsive verbal communication skills of children with Autism Spectrum Disorder (ASD): A further outcome of a pilot randomized controlled trial. *Research in Autism Spectrum Disorders, 27*, 73-87.
- *Srinivasan, S. M., Park, I. K., Neelly, L. B., & Bhat, A. N. (2015). A comparison of the effects of rhythm and robotic interventions on repetitive behaviors and affective states of children with Autism Spectrum Disorder (ASD). *Research in Autism Spectrum Disorders, 18*, 51-63.

- *Srinivasan, S. M., Kaur, M., Park, I. K., Gifford, T. D., Marsh, K. L., & Bhat, A. N. (2015). The effects of rhythm and robotic interventions on the imitation/praxis, interpersonal synchrony, and motor performance of children with Autism Spectrum Disorder (ASD): A pilot randomized controlled trial. *Autism Research and Treatment*, 2015.
- *Stadnick, N. A., Stahmer, A., & Brookman-Frazee, L. (2015). Preliminary effectiveness of Project ImPACT: A parent-mediated intervention for children with autism spectrum disorder delivered in a community program. *Journal of Autism and Developmental Disorders*, 45, 2092-2104.
- 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, 66–79.
- *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, 133-154.
- *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, 688-703.
- *Sullivan, K. A. (2013). *The Early Start Denver Model: Outcomes and moderators of an intervention for toddlers with Autism* (Unpublished doctoral dissertation). University of Washington, Seattle, WA.

- *Suzuki, M., Yamada, A., Watanabe, N., Akechi, T., Katsuki, F., Nishiyama, T., ... & Ota, A. (2014). A failure to confirm the effectiveness of a brief group psychoeducational program for mothers of children with high-functioning pervasive developmental disorders: A randomized controlled pilot trial. *Neuropsychiatric Disease and Treatment*, 10, 1141-1153.
- Tanner-Smith, E. E., Tipton, E., Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology* 2, 85-112.
- *Tellegen, C. L., & Sanders, M. R. (2014). A randomized controlled trial evaluating a brief parenting program with children with autism spectrum disorders. *Journal of Consulting and Clinical Psychology*, 82, 1193-1200.
- *Thompson, G. A., McFerran, K. S., & Gold, C. (2014). Family-centred music therapy to promote social engagement in young children with severe autism spectrum disorder: A randomized controlled study. *Child: Care, Health and Development*, 40, 840-852.
- *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, 166-177.
- *Tsang, S. K., Shek, D. T., Lam, L. L., Tang, F. L., & Cheung, P. M. (2007). Brief report: Application of the TEACCH Program on Chinese pre-school children with autism—Does culture make a difference?. *Journal of Autism and Developmental Disorders*, 37, 390-396.

- *Venker, C. E., McDuffie, A., Ellis Weismer, S., & Abbeduto, L. (2012). Increasing verbal responsiveness in parents of children with autism: A pilot study. *Autism, 16*, 568-585.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*, 1-48. doi:10.18637/jss.v036.i03
- *Vivanti, G., Paynter, J., Duncan, E., Fothergill, H., Dissanayake, C., Rogers, S. J., & 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*, 3140-3153.
- Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Wallace, B., Small, K., Brodley, C.E., Lau, J., & Trikalinos, T.A. (2012). Deploying an interactive machine learning system in an evidence-based practice center: abstrackr. In *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, 819-824.
- Warren, Z., McPheeters, M. L., Sathe, N., Foss-Feig, J. H., Glasser, A., & Veenstra-Vanderweele, J. (2011). A systematic review of early intensive intervention for autism spectrum disorders. *Pediatrics, 127*, e1303-1311.
- *Weiner, R. H., & Greene, R. L. (2014). Intention-based therapy for autism spectrum disorder: Promising results of a wait-list control study in children. *EXPLORE: The Journal of Science and Healing, 10*, 13-23.

- *Welsh, K. (2009). *The use of dogs to impact joint attention in children with Autism Spectrum Disorders* (Unpublished doctoral dissertation). Walden University, Minneapolis, MN.
- *Welterlin, A., Turner-Brown, L. M., Harris, S., Mesibov, G., & Delmolino, L. (2012). The home TEACCHing program for toddlers with autism. *Journal of Autism and Developmental Disorders*, 42, 1827-1835.
- *Wetherby, A. M., & Woods, J. J. (2006). Early social interaction project for children with autism spectrum disorders beginning in the second year of life: A preliminary study. *Topics in Early Childhood Special Education*, 26, 67-82.
- *Whitehouse, A. J., Granich, J., Alvares, G., Busacca, M., Cooper, M. N., Dass, A., ... & Anderson, A. (2017). A randomised controlled trial of an iPad-based application to complement early behavioural intervention in autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, 58, 1042-1052.
- *Whittingham, K., Sofronoff, K., Sheffield, J., & Sanders, M. R. (2009). Stepping Stones Triple P: An RCT of a parenting program with parents of a child diagnosed with an autism spectrum disorder. *Journal of Abnormal Child Psychology*, 37, 469.
- *Wild, D. L. (2012). *The impact of canine assistance for children with autism and the family unit* (Unpublished doctoral dissertation). Walden University, Minneapolis, MN.
- *Williams, B. T., Gray, K. M., & Tonge, B. J. (2012). Teaching emotion recognition skills to young children with autism: A randomised controlled trial of an emotion training programme. *Journal of Child Psychology and Psychiatry*, 53, 1268-1276.

Williams, K., Wray, J. A., & Wheeler, D. M. (2012). Intravenous secretin for autism spectrum disorders (ASD). *Cochrane Database of Systematic Reviews*, 4.

Wolery, M., Busick, M., Reichow, B., & Barton, E. E. (2010). Comparison of overlap methods for quantitatively synthesizing single-subject data. *The Journal of Special Education*, 44, 18-28.

*Wong, V. C., & Kwan, Q. K. (2010). Randomized controlled trial for early intervention for autism: A pilot study of the Autism 1-2-3 Project. *Journal of Autism and Developmental Disorders*, 40, 677-688.

Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., ... & Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of autism and developmental disorders*, 45, 1951-1966.

*Woo, C. C., Donnelly, J. H., Steinberg-Epstein, R., & Leon, M. (2015). Environmental enrichment as a therapy for autism: A clinical trial replication and extension. *Behavioral Neuroscience*, 129, 412-422.

*Woo, C. C., & Leon, M. (2013). Environmental enrichment as an effective treatment for autism: A randomized controlled trial. *Behavioral Neuroscience*, 127, 487-497.

Yoder, P. J., Bottema-Beutel, K., Woynaroski, T., Chandrasekhar, R., & Sandbank, M. (2013). Social communication intervention effects vary by dependent variable type in preschoolers with autism spectrum disorders. *Evidence-based communication assessment and intervention*, 7, 150-174.

- *Young, R. L., & Posselt, M. (2012). Using The Transporters DVD as a learning tool for children with Autism Spectrum Disorders (ASD). *Journal of Autism and Developmental Disorders*, 42, 984-991.
- *Yun, S. S., Choi, J., Park, S. K., Bong, G. Y., & Yoo, H. (2017). Social skills training for children with autism spectrum disorder using a robotic behavioral intervention system. *Autism Research*, 10, 1306-1323.
- *Zachor, D. A., Vardi, S., Baron-Eitan, S., Brodai-Meir, I., Ginossar, N., & Ben-Itzhak, E. (2017). The effectiveness of an outdoor adventure programme for young children with autism spectrum disorder: A controlled study. *Developmental Medicine & Child Neurology*, 59, 550-556.
- Zimmerman, K. N., Pustejovsky, J. E., Ledford, J. R., Barton, E. E., Severini, K. E., & Lloyd, B. P. (2018). Single-case synthesis tools II: Comparing quantitative outcomes measures. *Research in Developmental Disabilities*, 79, 65-76.

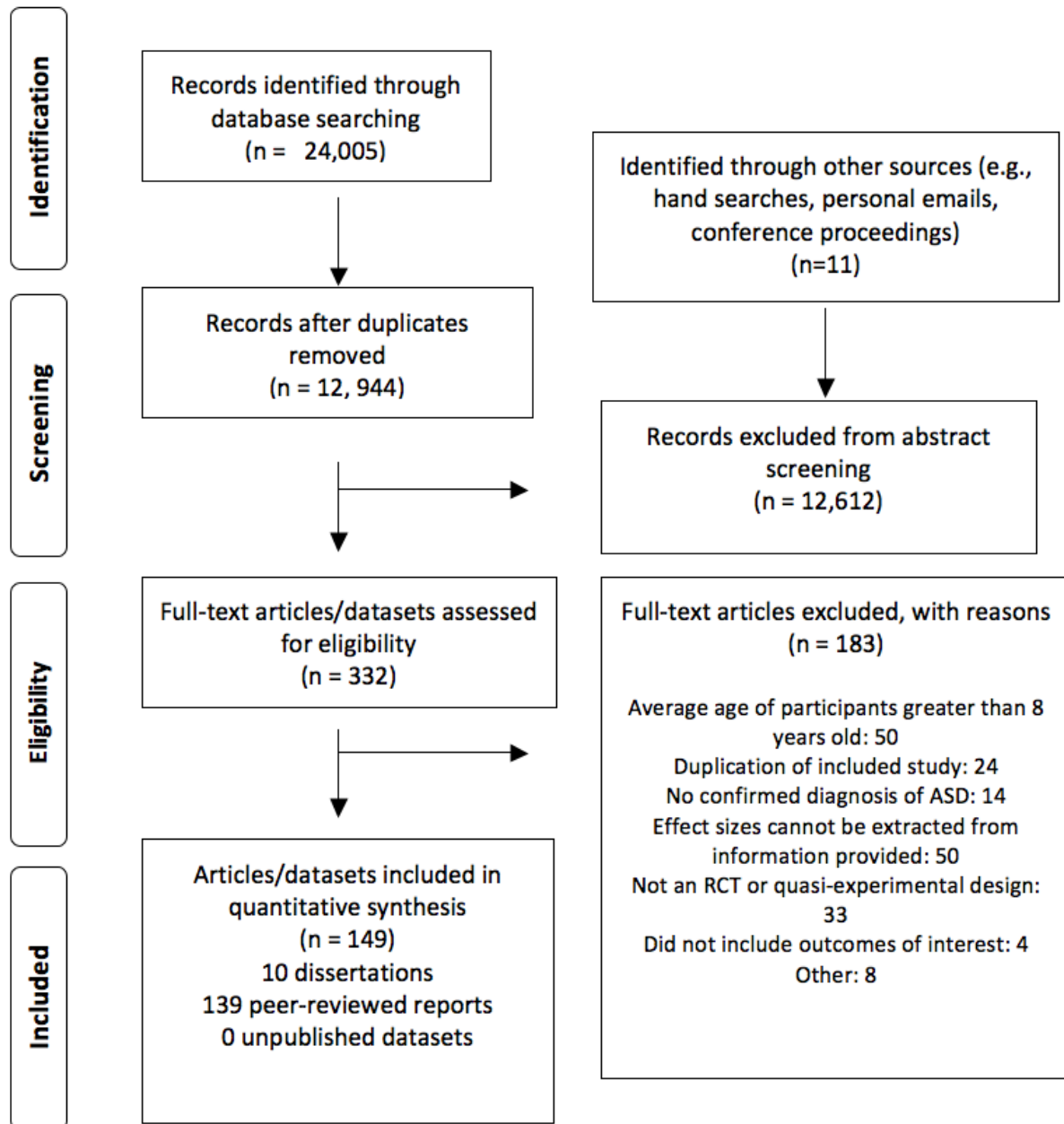


Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram.

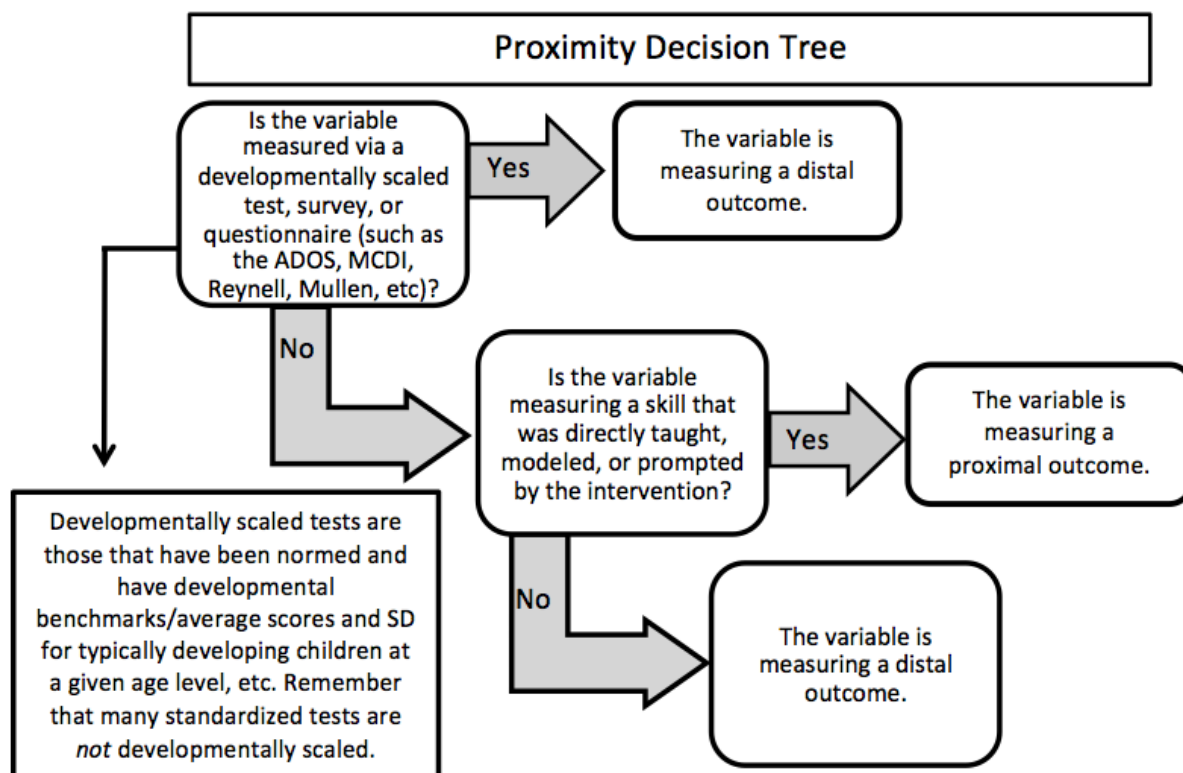


Figure 2. Decision tree (adapted from Yoder et al., 2013) used to code whether a study outcome was proximal or distal to treatment targets.

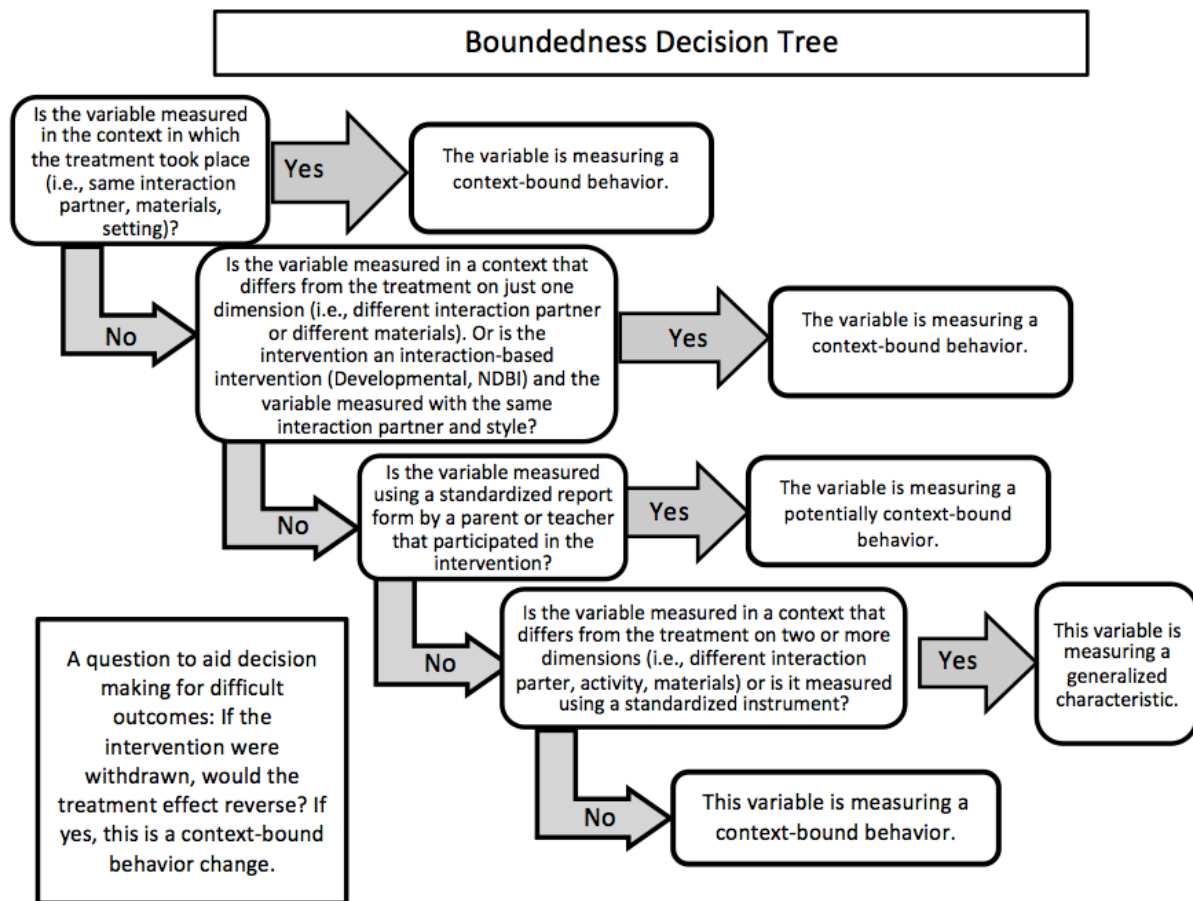


Figure 3. Decision tree (adapted from Yoder et al., 2013) used to code whether a study outcome measured a potentially context-bound or more highly generalized characteristic.

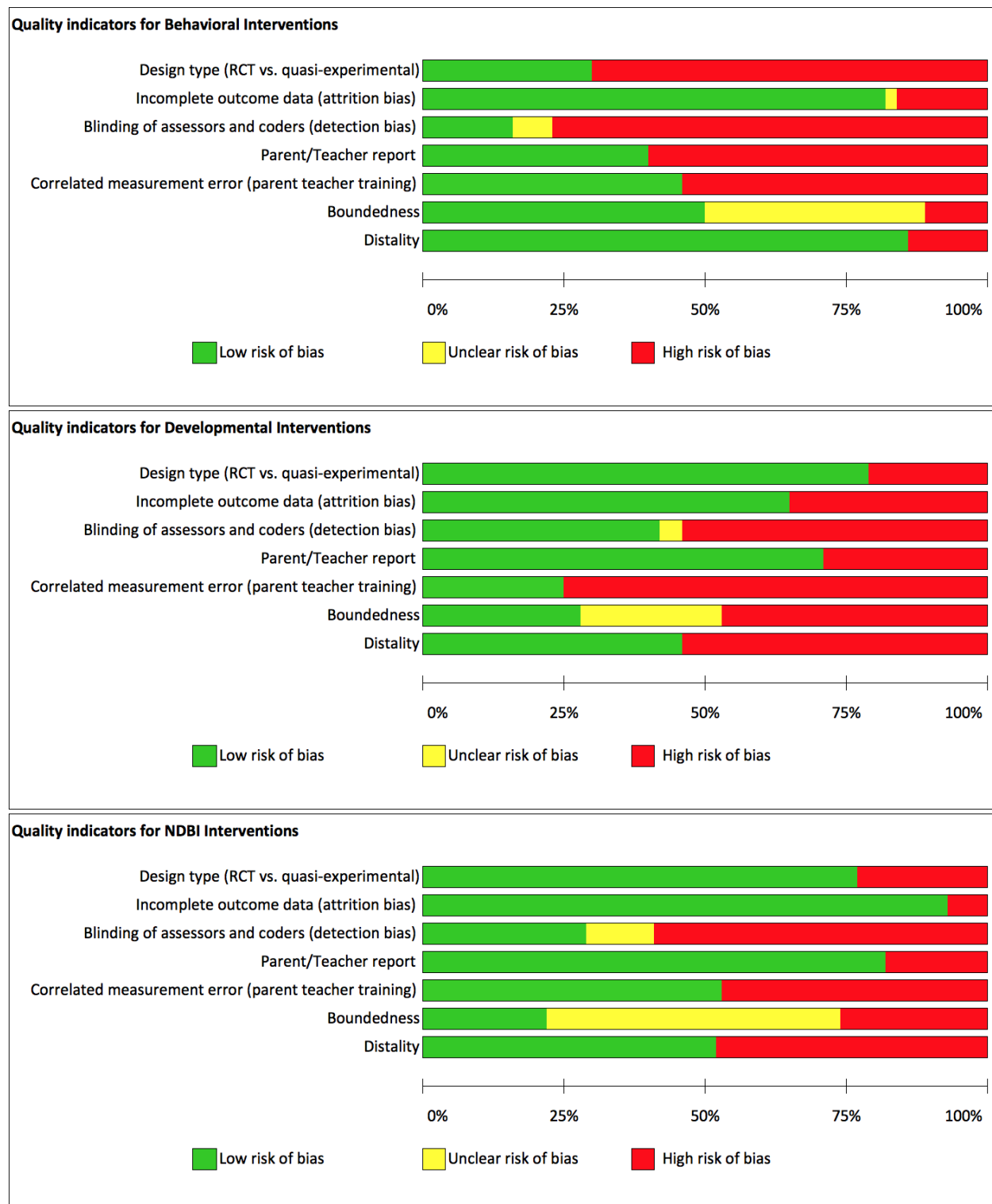


Figure 4. Summary of quality indicator ratings for studies of behavioral, developmental, and naturalistic developmental behavioral intervention (NDBI) types.

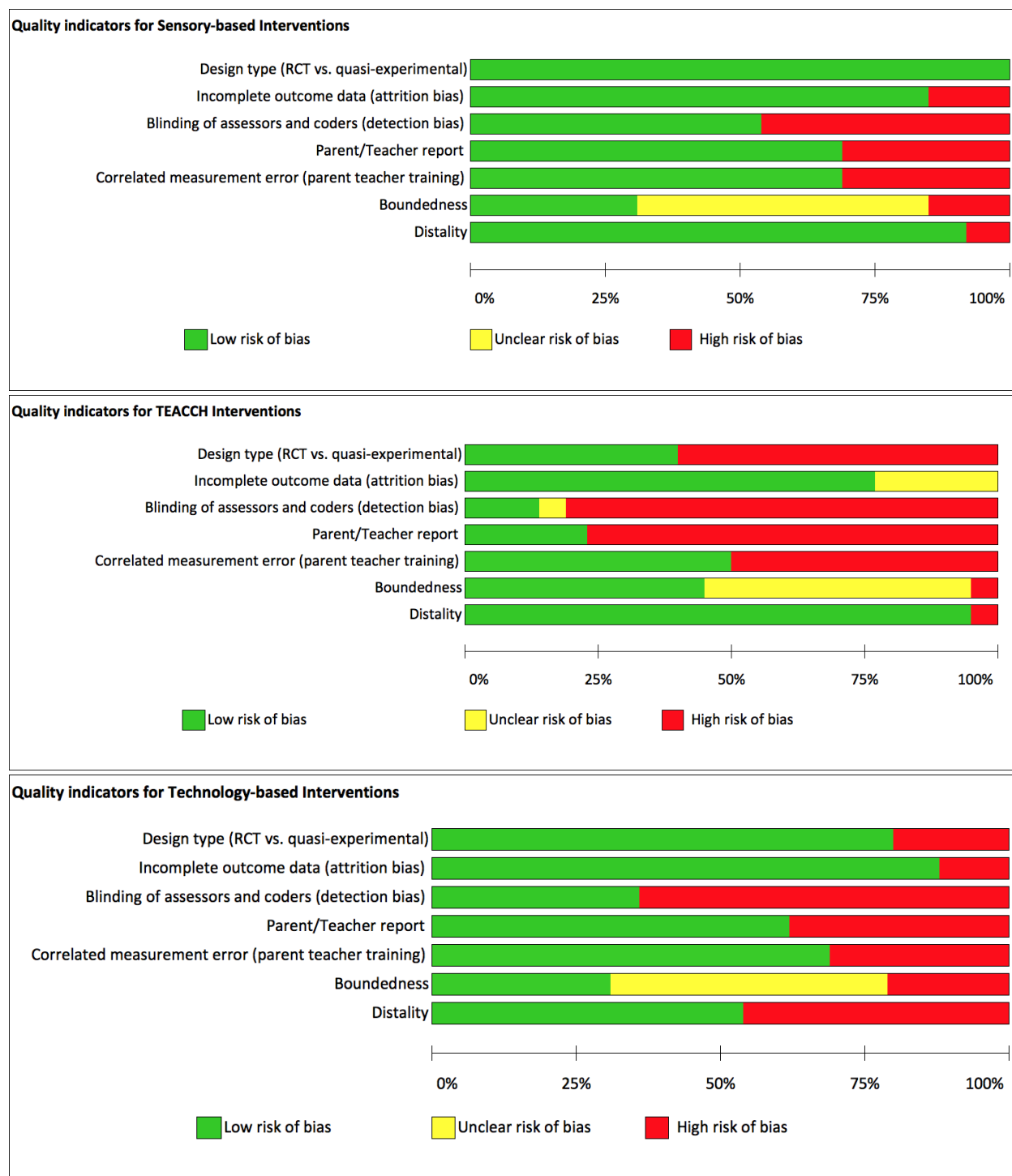


Figure 5. Summary of quality indicator ratings for sensory-based, Treatment and Education of Autistic and related Communication-handicapped Children (TEACCH), and technology-based intervention types.

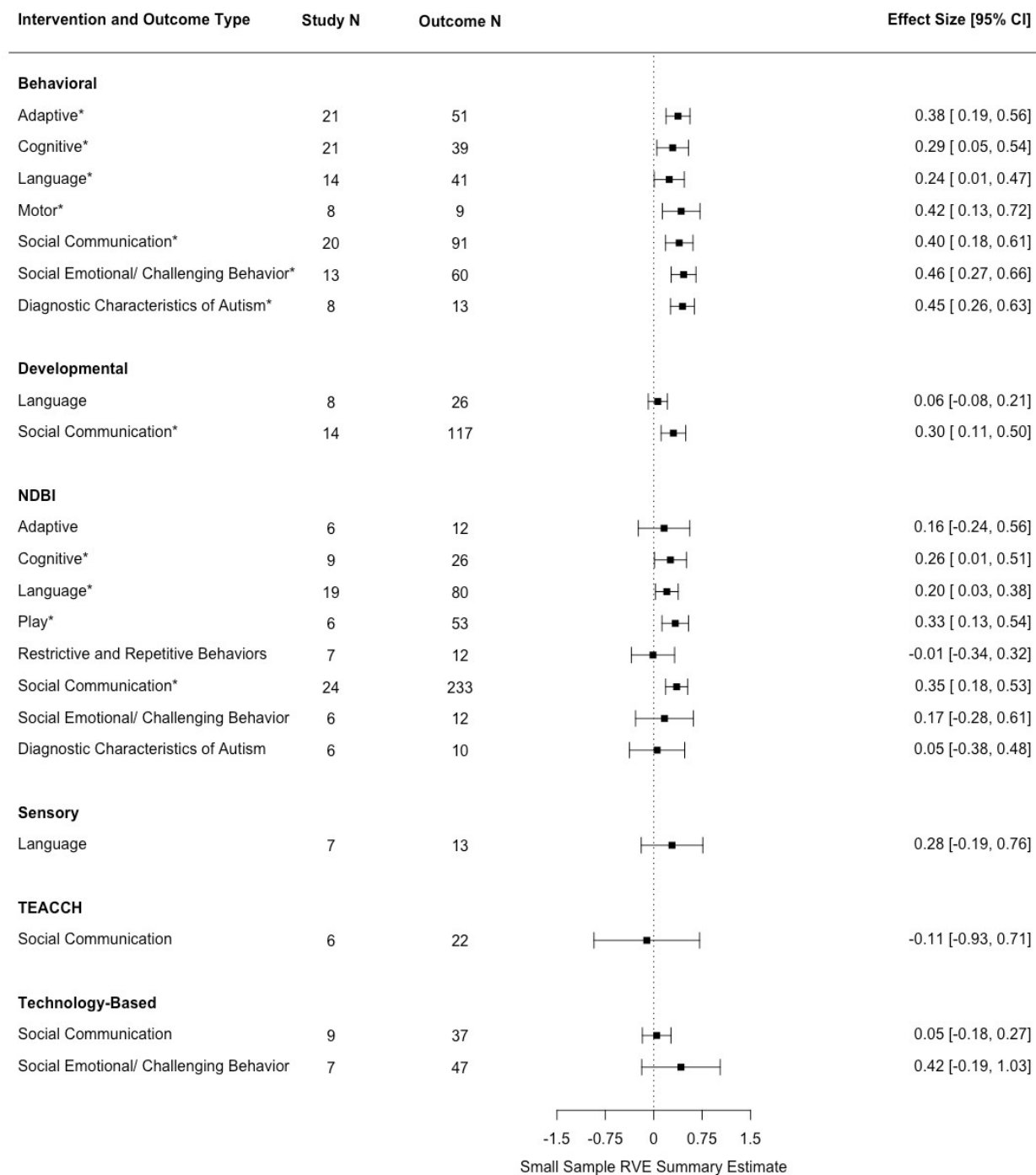


Figure 6. Forest plot of robust variance estimation (RVE) summary estimates with small sample bias correction for each outcome by intervention type, when all outcomes from quasi-

experimental and RCT group design studies are included. * denotes summary effect size estimates with confidence intervals that do not overlap with zero.

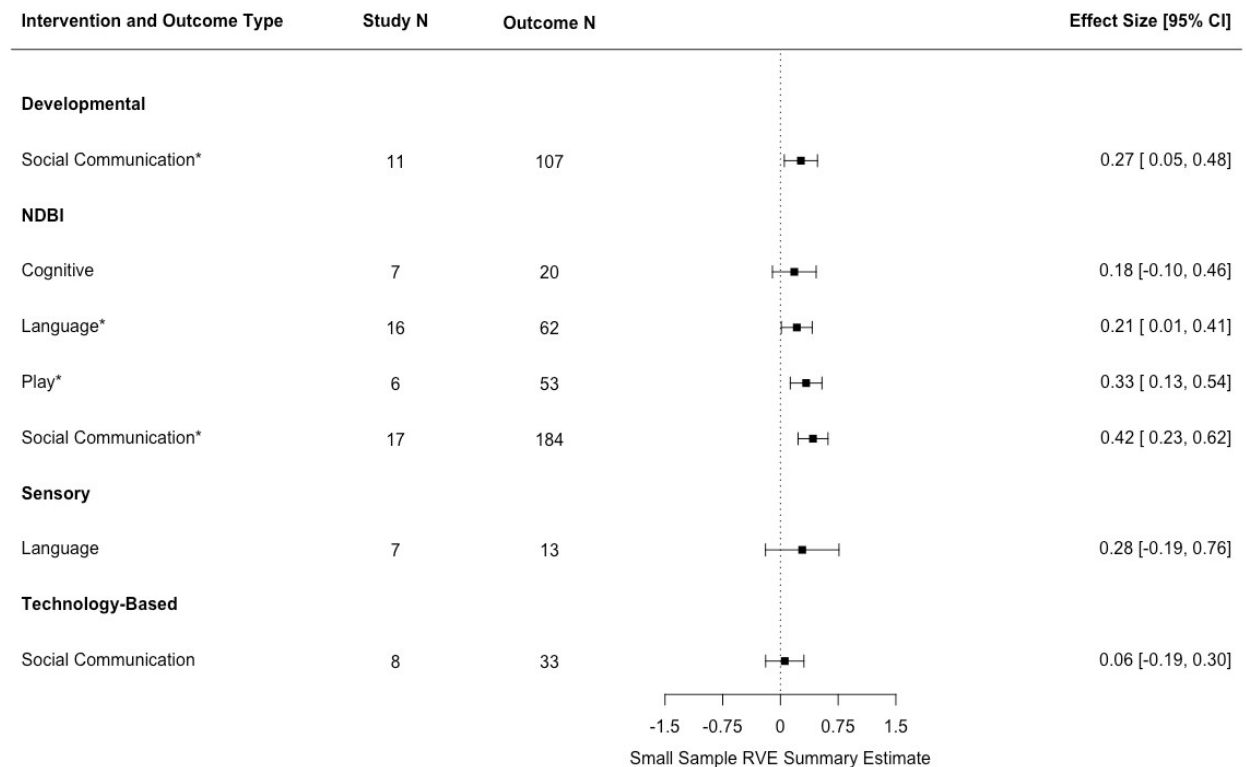


Figure 7. Forest plot of robust variance estimation (RVE) summary estimates with small sample bias correction for each outcome by intervention type, when all outcomes from RCTs are included. * denotes summary effect size estimates with confidence intervals that do not overlap with zero.

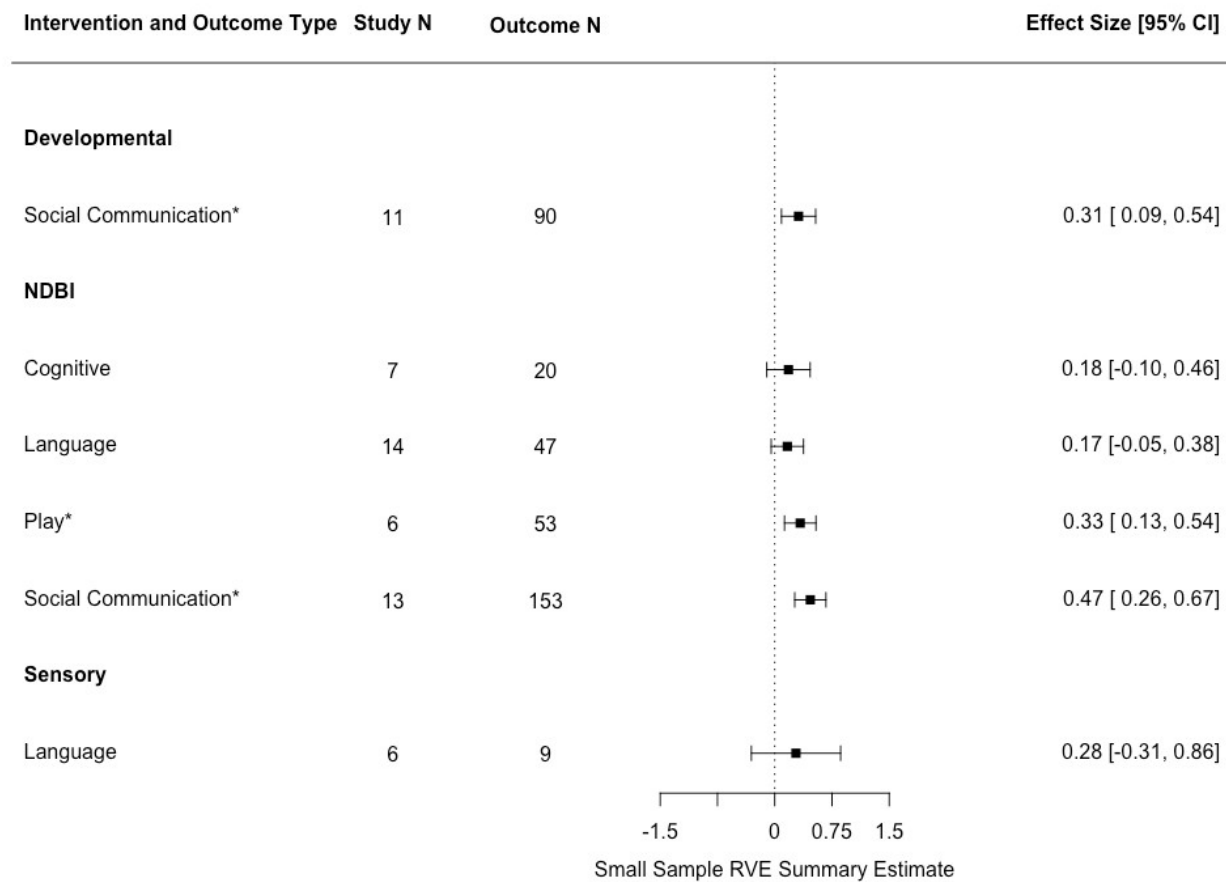


Figure 8. Forest plot of robust variance estimation (RVE) summary estimates with small sample bias correction for each outcome by intervention type, when only non-caregiver report outcomes from RCTs are included. * denotes summary effect size estimates with confidence intervals that do not overlap with zero.

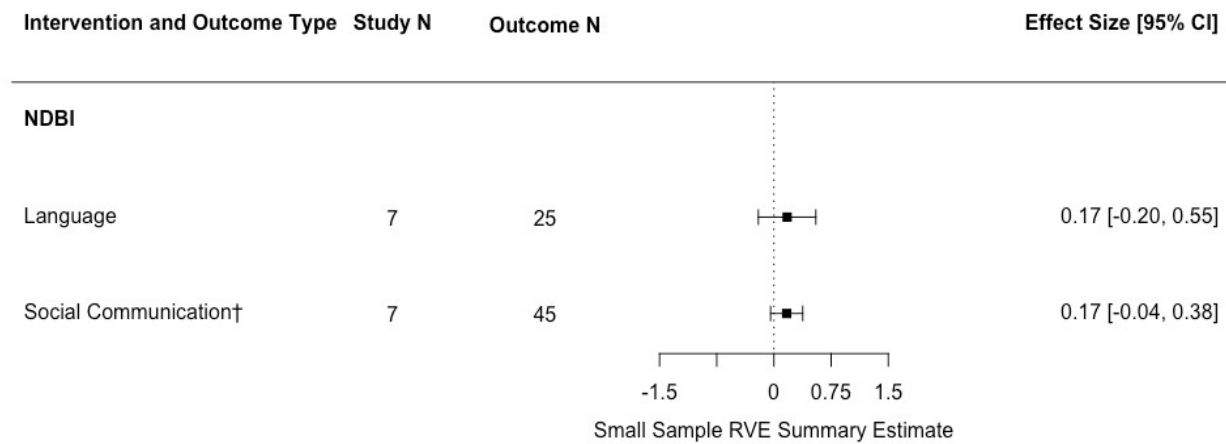


Figure 9. Forest plot of robust variance estimation (RVE) summary estimates with small sample bias correction for each outcome by intervention type, when only outcomes from RCTs that are not threatened by detection bias are included. † denotes summary effect size estimates that have p-values < 0.10.