



# Overview of Meta-Analyses on Naturalistic Developmental Behavioral Interventions for Children with Autism Spectrum Disorder

Jinwei Song<sup>1,2</sup> · Molly Reilly<sup>1,2</sup> · Brian Reichow<sup>1</sup>

Accepted: 17 November 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

## Abstract

We conducted an overview of reviews to determine the effects of naturalistic developmental behavioral interventions (NDBIs; Schreibman et al. (2015) *J Autism Dev Disorders* 45:2411–2428) on children with autism spectrum disorder under 8 years old. We conducted an electronic database search of Academic Search Premier, CINHAL, ERIC, Medline, and APA PsycINFO in October 2022 and August 2023 and utilized snowball methods to locate relevant reviews of NDBI. We included reviews meeting the following inclusion criteria: (1) review included a meta-analytic synthesis for at least one child outcome; (2) primary studies examined a NDBI; (3) primary studies included children with ASD with a mean pre-treatment age under eight years; (4) primary studies were conducted using a two-group comparison design; and (5) review was published in English. We extracted data on characteristics of the review, participant characteristics from the primary studies, intervention characteristics, and assessed the risk of bias of the included reviews. We conducted a narrative synthesis across outcomes reported in the included reviews. We included five reviews from six reports in this overview. Two reviews included studies that examined the Early Start Denver Model, two reviews included studies that examined the Pivotal Response Treatment, and one review included studies examining NDBIs collectively. We found positive effects of NDBIs on child's communication/language, cognition, and adaptive behavior. We found mixed effects for NDBIs on autism symptomatology and restricted and repetitive behaviors. Examination of moderator analyses reported in the included reviews suggested variables influencing the effects of NDBIs included proximity of outcome to intervention, boundedness of outcome to intervention, and study location. As shown in this overview, positive effects of NDBI for young children with ASD are supported by meta-analytic evidence. While the overall findings for NDBI across reviews are positive, the findings on specific outcomes and influential variables moderating the effects of NDBI are inconsistent. Additional evidence from randomized controlled trials and future meta-analyses are needed to strengthen our knowledge of the effects of NDBI for young children with ASD.

**Protocol Registration:** PROSPERO CRD42022353045.

**Keywords** Autism spectrum disorder · Naturalistic developmental behavioral interventions · Overview of reviews · Early Start Denver Model · Pivotal response treatment

## Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder that affects an estimated one in 36 children in the United States (Maenner et al., 2023). ASD is characterized by persistent deficits in social communication and

interaction, as well as repetitive and restricted patterns of behavior, interests, or activities (American Psychiatric Association, 2013). Children with ASD often face challenges in multiple areas of development, including language, communication, social skills, and adaptive functioning (e.g., Zwaigenbaum et al., 2015). Research (e.g., Magiati et al., 2014) has demonstrated the long-term impact of ASD on academic performance, social engagement, and quality of life. Moreover, the increasing economic and societal cost of raising a child with ASD can be a burden to their families (Rogge & Janssen, 2019). Research has shown early intervention for children with ASD can have positive impacts, especially for communication and socialization (e.g., Lord

✉ Brian Reichow  
breichow@uchc.edu

<sup>1</sup> A. J. Pappanikou Center for Excellence in Developmental Disabilities, UConn Health, Farmington, CT, USA

<sup>2</sup> Neag School of Education, University of Connecticut, Storrs, CT, USA

et al., 2022; Zwaigenbaum et al., 2015). However, questions about active elements of early intervention remain, with research showing the effectiveness of intervention can vary regarding to the specific intervention approach and child characteristics (Trembath et al., 2023). Therefore, it is essential to scientifically identify effective early interventions and its active elements for children with ASD.

## Naturalistic Developmental and Behavioral Interventions

In 2015, Schreibman and colleagues proposed a framework to categorize intervention models that integrate behavioral and developmental strategies, which they termed naturalistic developmental and behavioral interventions (NDBIs; Bruinisma et al., 2020; Schreibman et al., 2015). NDBIs integrate behavioral technologies and developmental principles in instructional strategies, such as environmental arrangements, natural reinforcers, prompting and prompt fading strategies, turn-taking, and adult's imitation of child's behavior, and use of the three-term contingency (i.e., antecedent, behavior, consequence). The teaching targets of NDBIs arise from developmental domains, such as language and communication, play, social interaction, cognition, and motor skills. The teaching opportunities occur in child's naturalistic environment, daily routines, or during other highly motivating interactions to promote generalization and maintenance of the new skills in natural settings. Empirical studies have shown positive impacts of NDBIs for children with ASD across areas, such as adaptive skills (e.g., Estes et al., 2015; Ingersoll et al., 2017), cognitive skills (e.g., Kasari et al., 2008; Wetherby & Woods, 2006), social communication (e.g., Brian et al., 2017; Shire et al., 2017) and language (e.g., Chang et al., 2016; Dawson et al., 2010), and play (e.g., Chang et al., 2016; Shire et al., 2017).

While there are a number of interventions that can be considered NDBIs, two of the most researched interventions include the Early Start Denver Model (ESDM) and Pivotal Response Treatment (PRT). ESDM (Rogers & Dawson, 2010) is an evidence-based intervention for young children with ASD and focuses on children's social-emotional, cognitive, and language development. ESDM prioritizes individualized intervention and parental involvement; parents are often trained to implement ESDM during daily interactions to promote generalization. Since the initial clinician led trials of ESDM, further research has shown it to be effective when delivered in groups (e.g., Vinen et al., 2018; Vivanti et al., 2014) and through parent education and training (e.g., Rogers et al., 2012; Vismara et al., 2016). PRT (Koegel & Koegel, 2019) is also an evidence-based intervention for children with ASD that targets pivotal developmental areas, such as motivation, responsivity to multiple cues, and

self-management. PRT utilizes natural reinforcers to encourage engagement and encourages interventionists to respond to children's cues and interests during child-led activities. While PRT is often delivered individually to children, studies have shown it to be effective when delivered in groups (e.g., Hardan et al., 2015) and through parent education and training (e.g., Gengoux et al., 2019; Schreibman & Stahmer, 2014).

Recently, a number of systematic reviews and meta-analyses have been published on NDBIs (e.g., Forbes et al., 2020; Sandbank et al., 2020; Tiede & Walton, 2019). While the reviews have, overall, reported predominantly positive findings, gaps, and inconsistencies in the research on NDBIs for children with ASD remain. For example, there are variations in the way in which NDBIs have been defined and used to include primary studies in reviews of NDBIs. Tiede and Walton (2019) included studies on Learning Experiences and Alternative Program (LEAP; Strain & Bovey, 2011) while others have not. In their review of PRT, Forbes et al. (2020) excluded studies that did not explicitly identify the intervention as PRT by name, possibly to the exclusion of studies with similar intervention techniques. Given the increased adoption of NDBIs in clinical practice (Bruinisma et al., 2020) and divergent findings in NDBIs reviews, we felt an overview of reviews would provide a systematic appraisal of the extant evidence on NDBI from meta-analytic reviews. The purpose of this overview was twofold: 1) What are the overall effects of NDBIs on children with ASD under 8-years-old? and 2) Which variables may influence the effects of NDBIs for children with ASD under 8-years-old?

## Methods

We registered, a priori, a review protocol with the International Prospective Register of Systematic Reviews (PROSPERO; CRD42022353045). We conducted an overview of reviews consistent with the methods outlined by Cochrane (e.g., Pollock et al., 2023) and is reported consistent with the Preferred Reporting Items for Overview of Reviews (Gates et al., 2022).

## Eligibility Criteria

We included systematic reviews meeting the following inclusion criteria:

1. Included at least one meta-analytic synthesis for a child outcome;
2. Primary studies of the review examined a NDBI (as defined by Schreibman et al., 2015);
3. Primary studies of the review included children with ASD who had a mean pre-treatment age less than 8-years-old;

4. Primary studies of the review were conducted using a two-group experimental design comparing NDBIs to a comparator;
5. Review was published in English.

For this overview, we did not place any restrictions on publication status (i.e., we did not exclude grey literature) or publication date. We did not consider systematic reviews with meta-analysis of single case experimental designs or qualitative research given differences in meta-analytic methods for these types of research design.

### Information Sources and Search Strategy

We conducted an electronic database search of Academic Search Premier, Cumulated Index to Nursing and Allied Health Literature (CINHAL), Education Resources Information Center (ERIC), Medline, and APA PsycINFO in EBSCO during October 2022 (updated in August 2023). The search strategy for each database is included in shown in Supplemental Text 1. We also used snowball methods (Greenhalgh & Peacock, 2005) to review the reference lists of all included reviews and Google Scholar to conduct a forward search of articles that cited included reviews.

### Selection Process

We exported the records from the electronic database search into Covidence (Veritas Health Innovation, 2020) for screening and study selection. Two reviewers independently screened records by title and abstract based on eligibility criteria, with disagreements resolved through consensus. The remaining records were then screened at the full-text stage, in which the same two screeners independently screened the full text of each record against the eligibility criteria. Disagreements between reviewers were resolved through discussion with a third party.

### Data Collection Process and Data Items

A data extraction sheet was developed, pilot tested on three randomly selected reviews, and then refined. After finalizing the data extraction form, two reviewers extracted the data independently, with disagreements resolved through discussion and mediation when necessary. We extracted information on research characteristics, participant characteristics, and intervention characteristics. For the research characteristics, we extracted the number of primary studies included, the meta-analytic model, the study selection methods, and the characteristics of the comparator or counterfactual condition. For the characteristics of the participants, we extracted variables such as age and gender. For the characteristics of the intervention, we extracted

the intervention methods/techniques, intervention components, intervention agent (i.e., who implemented the intervention), and intervention density (i.e., duration of intervention and intensity). We also extracted the data, when provided, on moderator and mediator analyses included in the meta-analytic synthesis of the included reviews (i.e., we extracted extant moderator/mediator analyses but did not conduct new moderator/mediator analyses).

Because outcomes were reported differently across reviews, we created outcome categories to help synthesize evidence across reviews. The categories included communication/language, cognition, adaptive behavior, autism symptomatology, and restricted and repetitive behaviors. Given the breadth of studies and outcomes related to communication and language, we created four sub-categories within this outcome category (i.e., generalized language, expressive communication, receptive communication, social communication). Operational definitions of the outcome categories used in this review, with exemplar measures for each category, are shown in Table 1.

### Risk of Bias

We used the Risk of Bias in Systematic Reviews (Whiting et al., 2016) to assess the risk of bias of the included meta-analytic reviews. As with data extraction, two independent reviewers extracted the risk of bias data with disagreements resolved through consensus. For the risk of bias assessment, we judged four domains: study eligibility criteria, identification and selection of studies, data collection and study appraisal, and synthesis findings. After each domain of bias was assessed, the reviewers made a summary-level judgement by collating the concerns of risks of bias identified for each domain. Finally, an overall risk of bias rating (i.e., high, low, or unclear) was made for each review. We did not assess the risk of bias for the primary studies of the included reviews.

### Overlap of Primary Studies

We used the corrected covered area (Pieper et al., 2014) to quantify the degree of primary study overlap across included meta-analyses. Corrected covered area was calculated as  $CCA = (N - u) / (uc - u)$ , where  $N$  was the number of included primary studies (including double counting),  $u$  was the number of primary studies (excluding duplicated reports), and  $c$  was the number of meta-analyses. We used Pieper and colleagues guidelines for quantifying the level of overlap for slight (0–5%), moderate (6–10%), high (11–15%), or very high (15–100%) levels of overlap.

**Table 1** Operational definition of outcome categories

Outcome category	Operational definition	Example of assessment tools used to measure outcome category
Communication and language	Superordinate category related to ways in which individuals actively engages with others or exchanges ideas and information. Subordinate categories include <i>generalized communication</i> (outcomes that integrated expressive communication, receptive communication, social communication, or other communicative and linguistic outcomes), <i>expressive communication</i> (verbal use of words and sentences to communicate with others), <i>receptive communication</i> (process of attending, understanding, and responding appropriately to others), and <i>social communication</i> (processes of communicating or interacting with others within a societal framework)	Generalized communication: <i>Vineland Adaptive Behavior Scales</i> (3rd edition; Sparrow et al., 2016); Communication Domain Score; <i>Griffiths Scales of Child Development</i> (3rd edition; Green et al., 2016); Language and communication scale Expressive communication: <i>Preschool Language Scales</i> (5th edition; Zimmerman et al., 2011); Expressive communication scale; mean length of utterance Receptive communication: <i>Mullen Scales of Early Learning</i> (Mullen, 1995); Receptive language scale Social communication: <i>Vineland Adaptive Behavior Scales</i> (3rd edition; Sparrow et al., 2016); Social domain score; <i>Griffiths Scales of Child Development</i> (3rd edition; Green et al., 2016); Personal-social-emotional scale <i>Mullen Scales of Early Learning</i> (Mullen, 1995); Early learning composite score; <i>Griffiths Scales of Child Development</i> (3rd edition; Green et al., 2016); Total score
Cognition	Processes involving knowledge and awareness, including reasoning and problem solving	<i>Vineland Adaptive Behavior Scales</i> (3rd edition; Sparrow et al., 2016): Adaptive behavior composite score; <i>Vineland Adaptive Behavior Scales</i> (3rd edition): Daily living skills domain score
Adaptive behavior	Skills or behaviors people use to complete activities of everyday life	<i>Vineland Adaptive Behavior Scales</i> (3rd edition; Sparrow et al., 2016): Adaptive behavior composite score; <i>Vineland Adaptive Behavior Scales</i> (3rd edition): Daily living skills domain score
Autism symptomatology	Perceived difficulties in social communication and restricted, repetitive patterns of behavior	<i>Autism Diagnostic Observation Schedule</i> (2nd edition; Lord & Rutter, 2012); Autism severity score; <i>Childhood Autism Rating Scale</i> (2nd edition; Schopler et al., 2010); Total score
Restricted and repetitive behavior	Stereotyped and repetitive activity or speech, insistence on routines, or restricted interests with abnormal intensity and focus, including atypical reactions to sensory input	<i>Repetitive Behavior Scale</i> (Bodfish et al., 1999); <i>Autism Diagnostic Observation Schedule</i> (2nd edition; Lord & Rutter, 2012): Restricted and repetitive behavior score

## Synthesis Methods

We conducted descriptive and narrative syntheses of the outcomes reported in meta-analyses of the included reviews. The intervention effect on each category of outcome was first assessed by examining the estimated magnitude of effects shown by the effect sizes calculated in each review. We extracted the standardized mean difference (SMD) effect size (e.g., Cohen's *d*, Hedges' *g*) or data from which to calculate a SMD for each outcome category or sub-category from each review. When a review reported more than one outcome (i.e., effect estimate) for a single category or sub-category, we selected one dependent measure from each study as a representative estimate based on the following hierarchy: (1) dependent measure collected using direct or standardized assessments, (2) dependent measure with the largest number of primary studies, and (3) dependent measure that included the largest number of child participants. We chose one outcome per outcome category to report for summative analyses, using the decision rules outlined above; we used this to select representative effect size estimates for the Ona et al. (2020) and Uljarević et al. (2022) reviews. To formulate conclusions regarding the evidence of NDBIs on child outcomes across reviews, we created tables to explore patterns of magnitude and statistical significance.

## Deviations from Review Protocol

We made two changes to the intended overview of review methods outlined in our protocol (CRD42011253045). First, we decided to exclude meta-analyses of primary studies that used single-case research designs. This decision was made, in part, due to finding a sufficient number of meta-analyses of group design studies, which met the aims of this overview more closely than single-case research designs. Second, we had planned to include caregiver outcomes in addition to child outcomes, but we were unable to include caregiver outcomes because the reviews that met our inclusion criteria did not contain any caregiver outcomes.

## Results

The electronic database search identified 1,304 records. After removing 398 duplicates, 906 records remained and underwent title and abstract screening. We removed 809 records based on titles and abstracts alone, which left 97 records for full-text screening. After full-text review, five meta-analyses (reported in six articles Fuller et al., 2020; Ona et al., 2020; Sandbank et al., 2020; Uljarević et al., 2022; Wang et al., 2022) met all eligibility criteria and were included in this overview (see Fig. 1 for a flow diagram of review selection). The use of the snowball search methods

identified an additional 1,390 records but yielded no additional reviews meeting all inclusion criteria. All five reviews were published in peer-reviewed journals.

## Review Characteristics

The five reviews that met our inclusion criteria were published between 2020 and 2022. The number of primary studies (*u*)<sup>1</sup> in each review ranged from seven (Ona et al., 2020) to 26 (Sandbank et al., 2020). Across the five reviews, the cumulative number of included studies summed to 66. This figure represents a gross count of primary studies that includes a count including primary studies that were in more than one review. Across reviews, the total number of unique (unduplicated) primary studies was 48 (*u* = 48); 11 primary studies (Dawson et al., 2010; Hardan et al., 2015; Mohammadzaheri et al., 2014, 2015; Nefdt et al., 2010; Rogers et al., 2012, 2019; Schreibman & Stahmer, 2014; Vinen et al., 2018; Vismara et al., 2016; Vivanti et al., 2014) were included in more than one review. The primary study overlap estimated by the corrected covered area (CCA) was approximately 9.38%, indicating a moderate level of overlap.

The characteristics of the five included reviews are shown in Table 2. Most reviews (4 of 5) used multiple search methods for study selection. The most common methods for study selection included electronic database searches (5 of 5) and hand searches of the reference lists of included studies (3 of 5). Three meta-analyses included information on the comparators used within the primary studies, with the most common comparators being treatment as usual (*u* = 9), waitlist control (*u* = 4), and psychoeducation (*u* = 1). Across reviews, different methods of meta-analysis were used to statistically combine effect size estimates. Three reviews (Ona et al., 2020; Uljarević et al., 2022; Wang et al., 2022) used a random-effects model and two reviews (Fuller et al., 2020; Sandbank et al., 2020) used robust variance estimation. Across reviews, we extracted 23 SMD effect size estimates across our five outcome categories and sub-categories.

## Participant Characteristics

The characteristics of the participants of the reviews are shown in Table 3. Across studies, 1,697 child participants were included in the four reviews (data on the total number of participants was not available for the Sandbank et al. review), with a range of 181 (Ona et al., 2020) to 640 (Fuller et al., 2020) children. Across reviews, the mean age of the child participants ranged from 2.5 (Fuller et al., 2020) to 5.6

<sup>1</sup> Consistent with reporting guidelines for overview of reviews, we have used *u* in reference to the number (count) of primary research studies included in the reviews.

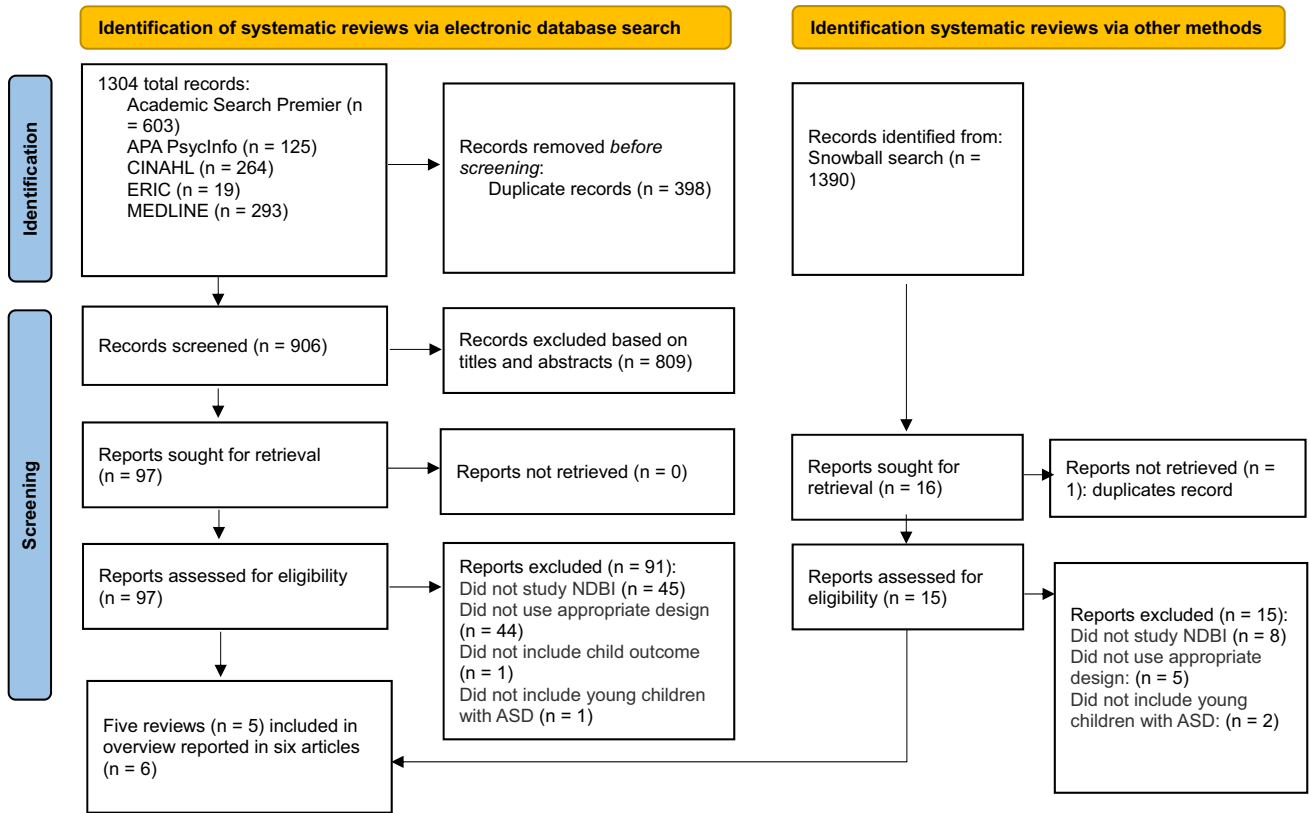


Fig. 1 PRIOR flowchart of initial search

Table 2 Characteristics of included meta-analytic reviews

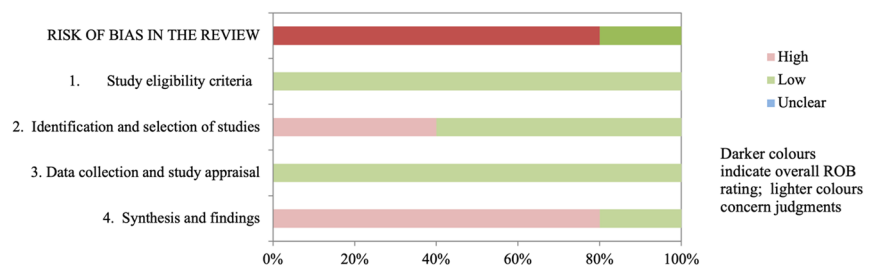
Review (first author and year)	Primary studies (u)	Search methods	Meta-analytic model	Intervention	Comparator
Fuller (2020)	12	Electronic database, hand search of reference lists of included studies	RVE	ESDM	TAU (u=6) WLC (u=1) Other (u=5)
Ona (2020)	7	Electronic database, hand search of reference lists of included studies, examination of extant reviews	RE	PRT	WLC (u=2) PEG (u=1) ABA (u=1) PECS (u=1)
Sandbank (2020)	26	Electronic database, experts contact	RVE	NDBIs	n/r
Uljarević (2022)	10	Electronic database, hand search of reference lists of included studies	RE	PRT	TAU (u=3) WLC (u=3) PEG (u=1) ABA (u=1) PECS (u=1)
Wang (2022)	11	Electronic database	RE	ESDM	n/r

u number of unique primary studies, RVE robust variance estimation, ESDM Early Start Denver Model, TAU treatment as usual, WLC wait list control, RE random-effects model, PRT pivotal response treatment, PEG psychoeducation, ABA = applied behavior analysis, PECS picture exchange communication system, NDBIs naturalistic developmental behavioral interventions; n/r not reported

**Table 3** Characteristics of primary studies included in reviews

Review	Sample size total ( <i>T:C</i> )	Age M (SD; range)	% Male	Intervention agent ( <i>u</i> )			Intervention density	
				CG	INT	COM	Duration (months)	Intensity (hours per week)
Fuller (2020)	640 (286:354)	2.5 (0.89)	81%	5	7	0	1.4–35.9	n/r
Ona (2020)	181 (91:90)	5.3 (2.4–9.2)	n/r	3	4	0	0.2–5.3	n/r
Sandbank (2020)	n/r	3.2 (1.02)	81%	11	9	6	M=0.8	n/r
Uljarević (2022)	252 (130:121)	5.56 (1.16)	60–92%	2	2	6	3–6	n/r
Wang (2022)	624	1.7–4.0	n/r	3	8	0	M=10.8	10.9

*T* treatment, *C* control, *u* number of unique primary studies, *CG* caregivers, *INT* interventionist, *COM* combination of CG and INT, *n/r* not reported

**Fig. 2** Risk of bias assessment

Note. <sup>a</sup>Risk of bias assessed using Whiting, et al. (2016).

(Uljarević et al., 2022) years old. The majority of reviews reported more than 80% of participants were male, which is consistent with the typical gender distribution of males to females for ASD (American Psychiatric Association, 2013).

### Intervention Characteristics

Table 3 also shows the characteristics of the interventions examined in the primary studies included in the reviews. Two reviews (Fuller et al., 2020; Wang et al., 2022) included studies that examined the ESDM, two reviews (Ona et al., 2020; Uljarević et al., 2022) included studies that examined PRT, and one review (Sandbank et al., 2020) included studies examining NDBIs collectively. Across the primary studies included in the reviews, the intervention agents included caregivers ( $u=24$ ), interventionists ( $u=30$ ; i.e., paraprofessionals, therapists, educators, and clinicians), and a combination of caregivers and professionals ( $u=12$ ). Five reviews reported data on intervention density. Two reviews (Sandbank et al., 2020; Wang et al., 2022) reported the mean length of intervention across primary studies, with a range of 0.8 (Sandbank et al., 2020) to 10.8 months (Wang et al., 2022), and the other three reviews reported the range intervention length of primary studies, with a range of 0.2 to 35.9 months. One review (Wang et al., 2022) reported the density of intervention across primary studies in hours per week, with a mean of 10.9 h per week.

### Risk of Bias

A summary of risk of bias across the five included reviews is shown in Fig. 2 and an itemized risk of bias by review and domain is shown in Supplemental Table 1. Overall, 80% of the included reviews were judged to have a high risk of bias. The highest levels of concern were seen in the domains of synthesis and findings (4 of 5 reviews rated as having a high risk of bias) and study selection (2 of 5 reviews rated as having a high risk of bias). Lower risks of bias were seen in the domains of study eligibility criteria and data collection and appraisal, where all reviews were judged to have had a low risk of bias.

### Effects of NDBI on Child Outcomes

Effect size estimates for each outcome category are shown by review in Table 4, which serves an overview of the results of this overview. Four reviews (Fuller et al., 2020; Ona et al., 2020; Sandbank et al., 2020; Wang et al., 2022) meta-analyzed an outcome included in our category of generalized language outcomes. Across meta-analyses, the SMD ranged from 0.20 (95% CI 0.03 to 0.38,  $u=19$ ; Sandbank et al., 2020) to 1.12 (95% CI -0.49 to 2.73,  $u=2$ ; Ona et al., 2020). Two reviews (Ona et al., 2020; Uljarević et al., 2022) meta-analyzed expressive communication outcomes, with a range of SMD 0.48 (95% CI 0.04 to 0.93,  $u=2$ ; Ona et al., 2020)

**Table 4** Summary of effect sizes by outcome categories

Review	Meta-analytic model	Communication/language				Cognition	Adaptive behavior	Autism symptomatology	RRB
		Generalized comm	Expressive comm	Receptive comm	Social comm				
Fuller (2020)	RVE	$g=0.41$ ( $p=0.01$ , $u=11$ )			$g=0.21$ ( $p=0.29$ , $u=8$ )	$g=0.41$ ( $p=0.04$ , $u=9$ )	$g=0.12$ ( $p=0.46$ , $u=6$ )	$g=-0.07$ ( $p=0.62$ , $u=9$ )	$g=0.02$ ( $p=0.88$ , $u=5$ )
Ona (2020)	RE	SMD=1.12, (95% CI -0.49 to 2.73, $u=2$ )	SMD=0.48, (95% CI 0.04 to 0.93, $u=2$ )						
Sandbank (2020)	RVE	$g=0.20$ (95% CI 0.03 to 0.38, $u=19$ )			$g=0.35$ (95% CI 0.18 to 0.53, $u=24$ )	$g=0.26$ (95% CI 0.01 to 0.51, $u=9$ )	$g=0.16$ (95% CI -0.24 to 0.56, $u=6$ )	$g=0.05$ (95% CI -0.38 to 0.48, $u=6$ )	$g=-0.01$ (95% CI -0.34 to 0.32, $u=7$ )
Uljarević (2022)	RE		SMD=1.37 (95% CI -2.53 to 5.27, $u=3$ )	SMD=0.51 (95% CI 0.23 to 0.80, $u=3$ )		SMD=0.15 (95% CI -0.17 to 0.48, $u=3$ )	SMD=0.31 (95% CI -0.03 to 0.65, $u=2$ )	SMD=-6.03 (95% CI -13.45 to 1.40, $u=2$ )	
Wang (2022)	RE	$g=0.28$ (95% CI 0.00 to 0.56, $u=7$ )			$g=0.01$ (95% CI -0.18 to 0.20, $u=7$ )	$g=0.28$ (95% CI 0.11 to 0.45, $u=7$ )		$g=-0.27$ (95% CI -0.53 to -0.02, $u=8$ )	

*generalized comm* generalized communication, *expressive comm* expressive communication, *receptive comm* receptive communication, *social comm* social communication, *RRB* restricted and repetitive behavior, *RVE* robust variance estimation, *g* Hedges' *g*, *u* number of unique primary studies, *RE* random-effects model, *SMD* standardized mean difference effect size, *CI* confidence interval

to 1.37 (95% CI -2.53 to 5.27,  $u=3$ ; Uljarević et al., 2022) and one review meta-analyzed receptive communication (SMD=0.51; 95% CI 0.23 to 0.80,  $u=3$ ; Uljarević et al., 2022). Three reviews (Fuller et al., 2020; Sandbank et al., 2020; Wang et al., 2022) meta-analyzed social communication with the range of SMD from 0.01 (95% CI -0.18 to 0.20,  $u=7$ ; Wang et al., 2022) to 0.35 (95% CI 0.18 to 0.53,  $u=24$ ; Sandbank et al., 2020).

Four reviews (Fuller et al., 2020; Sandbank et al., 2020; Uljarević et al., 2022; Wang et al., 2022) meta-analyzed cognitive outcomes, with SMD ranging across reviews from 0.15 (95% CI -0.17 to 0.48,  $u=3$ ; Uljarević et al., 2022) to 0.41 ( $p=0.04$ ,  $u=9$ ; Fuller et al., 2020). Three reviews (Fuller et al., 2020; Sandbank et al., 2020; Uljarević et al., 2022) meta-analyzed adaptive behavior. Across reviews, the SMD ranged from 0.12 ( $p=0.46$ ,  $u=6$ ; Fuller et al., 2020) to 0.31 (95% CI -0.03 to 0.65,  $u=2$ ; Uljarević et al., 2022). Four reviews (Fuller et al., 2020; Sandbank et al., 2020; Uljarević et al., 2022; Wang et al., 2022) meta-analyzed autism symptomatology, with SMD ranging from -6.03 (95% CI -13.45 to 1.40,  $u=2$ ; Uljarević et al., 2022) to 0.05 (95% CI -0.38 to 0.48,  $u=6$ ; Sandbank et al., 2020). Finally, two reviews (Fuller et al., 2020; Sandbank et al., 2020) meta-analyzed restricted and repetitive behaviors, with SMD

ranging from -0.01 (95% CI -0.34 to 0.32,  $u=7$ ; Sandbank et al., 2020) to 0.02 ( $p=0.88$ ,  $u=5$ ; Fuller et al., 2020).

### Influential Variables

Three reviews (Fuller et al., 2020; Sandbank et al., 2020; Wang et al., 2022) conducted statistical tests (i.e., moderator/mediator analyses) to explore variables that might be related to differences in outcomes reported in their meta-analyses. The variables or factors that were explored included study quality (i.e., correlated measurement error, parent report, independent assessor, and randomization), intervention characteristics (i.e., parent involvement, length of intervention, hours of intervention per week, total number of intervention hours, interventionist, intervention delivery method, and study location), sample characteristics (i.e., mean chronological age, mean language age at study entry, and percentage of males in each sample), and outcome characteristics (i.e., proximity of outcome to intervention, boundedness of outcome to intervention, and measurement proximity). Among these variables, proximity of outcome to intervention (Sandbank et al., 2020), boundedness of outcome to intervention (Sandbank et al., 2020), and study location (Wang et al., 2022) were identified as variables



with a statistically significant association to the effects of NDBI. Outcomes proximal to the intervention had larger effect sizes ( $\beta = 0.25$ ,  $p = 0.041$ ) than distal outcomes in Sandbank et al. (2020). Generalized ( $\beta = -0.40$ ,  $p = 0.003$ ) or potentially context-bounded outcomes ( $\beta = -0.31$ ,  $p = 0.022$ ) had smaller effect sizes than context-bounded outcomes in Sandbank et al. (2020). Additionally, Wang et al. (2022) found that studies conducted in Asia (i.e., China), on average, had larger effect sizes than studies conducted in western countries (i.e., United States, Australia) for both autism symptomatology ( $Q_{\text{between}} = 3.99$ ,  $p = 0.046$ ) and generalized communication/language ( $Q_{\text{between}} = 7.12$ ,  $p = 0.008$ ). However, the moderating effects of proximity of outcome to intervention and boundedness of outcome to intervention did not have a statistically significant association in Fuller et al., 2020 ( $p = 0.20$  to  $0.95$ ).

## Discussion

This overview of reviews synthesizes the meta-analytic evidence on child outcomes associated with NDBIs for young children with ASD. The findings of this overview support the positive findings on the effects of NDBIs on young children's communication and language skills and cognitive development. The findings across reviews for these outcomes were robust, with medium to large effect sizes shown for each outcome in multiple reviews. The largest statistically significant effect sizes for the communication outcomes were shown in Fuller et al. (2020), Ona et al. (2020), and Uljarević et al. (2022) and the largest estimated effects for cognition was shown in Fuller et al. (2020). The effects of NDBIs were not found to have statistically significant differences favoring the treatment group over the control group for adaptive behavior, autism symptomatology, or restrictive and repetitive behaviors. However, the meta-analyses for these outcomes were based on a small sample of primary studies, which can impact statistical power and allow for the presence of Type I error.

While NDBIs were shown to be an effective treatment option for most children with ASD, it is important to consider heterogeneity (between study heterogeneity and between review heterogeneity) when interpreting the findings of this overview. Heterogeneity is most typically used to refer to systematic differences between studies included in a single review or meta-analysis (Borenstein, 2019). For an overview of reviews, heterogeneity can also be used in reference to differences between included reviews (Pollock et al., 2023). Different types of heterogeneity have been suggested including clinical heterogeneity (i.e., differences in interventions, participants, or outcomes; Fletcher, 2007), statistical heterogeneity (i.e., differences in effects or study results; Fletcher, 2007), and methodological heterogeneity (i.e.,

differences in design or study confounds; Deeks et al., 2019). While there are standard methods for dealing with heterogeneity in meta-analyses (e.g., Borenstein, 2019; Deeks et al., 2019), guidelines and standards for dealing with heterogeneity in an overview of reviews has not been established (Gates et al., 2020). Without a standard for detecting and dealing with heterogeneity in overview of reviews, we explore the presence and impact of heterogeneity descriptively for this overview in the following section.

## Between Review Heterogeneity in Overview of NDBI

Across the five reviews included in this overview, two primary intervention methods of NDBI were used—ESDM and PRT. While these interventions have many similarities and are both considered NDBIs, they do have some distinctions, which might introduce clinical heterogeneity. Similarly, the mean age of participants receiving ESDM was lower than the mean age of participants receiving PRT groups. Furthermore, Sandbank et al. (2020) included primary studies of ESDM and PRT with other NDBI interventions, whereas the other meta-analyses included primary studies of either ESDM or PRT alone. Differences in these and other variables (e.g., intervention fidelity) may lead to systematic differences in intervention effects and need further attention in systematic reviews.

We observed variability in the effects (i.e., results) presented across reviews, suggesting the presence of statistical heterogeneity. One possible reason for this heterogeneity could be that different meta-analytic models were used across reviews. In a random-effects model, a true effect size is assumed to be random and is estimated using a weighted mean that takes into account both within-study and across-study variances (Konstantopoulos & Hedges, 2019). When using robust variance estimation, the standard error is estimated by the observed variation in effect sizes and the average effect size does not require the assumptions on normal distribution of effect sizes, correct variances, and inverse variance (Hedges et al., 2010). Given the differences in how data area analyzed in these two meta-analytic models, differences in estimated effects (i.e., mean effect size estimates) would be likely to occur. For instance, Sidik and Jonkman (2006) found a random-effects model produced imprecise results due to errors associated with the estimation of marginal variances not being used. Robust variance estimate may provide a means by which to address some of the issues of multiple measures within a study and robust estimates about the mean effect size; however, the method does not confer robust estimates of the prediction interval, which can serve as a reliable indicator of heterogeneity. Additionally, the amount of within-study correlations between effect sizes (i.e.,  $\rho$ ; Fisher & Tipton, 2015) needs to be justified in a robust variance estimation and is not always readily

known or reported. Finally, between-study variance (i.e., *tau-squared*) and meta-regression coefficients are sensitive to changes in *rho* when between-study covariance is significantly smaller than the within-study covariances. Given these differences and possible influences, evaluation of differences in meta-analytic models using large extant data sets would be helpful.

Analyses of moderator/mediator analyses were done in three reviews (Fuller et al., 2020; Sandbank et al., 2020; Wang et al., 2022) included in this overview. While some variables (i.e., moderators) were found to have a statistically significant relation to the magnitude of effect (e.g., proximity of outcome to intervention, boundedness of outcome to intervention, and study locations) most of the statistical tests that were conducted in the included reviews did not find statistically significant relations between variables and outcomes suggesting the variables were not mediating or moderating the effects of NDBIs. Again, a small number of intervention studies on which the secondary analyses were conducted could limit the power of these tests to show meaningful associations between variables and should be reexamined when more primary studies have been completed.

## Limitations

There are several limitations to this overview that should be taken into consideration when interpreting our findings. First, only one included review, Ona et al. (2020), had a preregistered protocol. Lack of pre-registration has been shown to be associated with greater review-level risks of bias, thus should be considered a limitation of this overview. Differences between the Ona et al. (2020) review protocol and their published review, including changes to the meta-analytic synthesis methods used, were noted, which could increase risks of bias. Second, while a risk of bias of assessment of the primary studies of the reviews included in this overview was not completed, four of five reviews (Ona et al., 2020; Sandbank et al., 2020; Uljarević et al., 2022; Wang et al., 2022) reported concerns with the methodological quality of the primary studies including lack of blinding of outcome assessors, lack of blinding of treatment personnel, and selection bias. Third, we only included meta-analyses of primary studies that were conducted using comparative group experimental designs. Thus, the meta-analysis of single case studies of NDBI by Bozkus-Genc and Yucesoy-Ozkan (2016) was excluded. Restriction on design type for primary studies may limit the generality of our findings given the large number of NDBI primary studies that are conducted using single case experimental designs. Third, the number of primary studies (*u*) included in the meta-analytic syntheses of the reviews was often small, which could lead to underpowered analyses raising a concern of

Type I error. This concern is seen in the wide confidence intervals associated with many of the analyses done (see Table 4). Moreover, the primary studies included in the analyses often contained small sample sizes themselves, which could compound this potential limitation. Future reviews or overviews of NDBI might consider inclusion and synthesis of studies that utilized single case experimental designs to gain further insights into the effects of this category of interventions for children with ASD. A fourth, related limitation is that although three reviews conducted moderator analyses, only a small number of moderators were found to have a statistically significant relation to the effects of NDBI. Given the relatively small sample sizes in the included meta-analyses, replication of these exploratory analyses should be done as the small sample of primary studies could lead to Type I error. A final limitation of this overview is the moderate level of overlap of primary studies across the included reviews (corrected covered area [CCA]=9.38%). Although many studies were included across reviews, methods for dealing with this overlap have yet to be established for overview of reviews.

## Conclusion

As shown in this overview, the positive effects of NDBI for young children with ASD are supported by meta-analytic evidence. While the overall findings for NDBI across reviews are positive, the findings on specific outcomes and influential variables moderating the effects of NDBI are inconsistent. Additional evidence from randomized controlled trials and future meta-analyses are needed to strengthen our knowledge of the effects of NDBI for young children with ASD.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10803-023-06198-x>.

**Funding** The research reported in this overview was supported, in part, by funding from the US Department of Health and Human Services, Administration for Community Living, Administration on Intellectual and Developmental Disabilities (AIDD) (Awards 90DDUC0071; 90DDUC0140) and the US Department of Education Office of Special Education Programs (H325H190004). The opinions expressed, however, are those of the authors and do not necessarily reflect the official position of a Department or Office.

## Declarations

**Competing interest** Brian Reichow was serving as an Associate Editor for the Journal of Autism and Developmental Disorders at the time of submission and publication but did not have any input on the review or decisions of the manuscript. The authors report no other perceived or real conflicts of interest that are directly or indirectly related to the work submitted to publication.

## References

## References marked with an asterisk indicate studies included in this overview.

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). APA.
- Bodfish, J. W., Symons, F. W., & Lewis, M. H. (1999). *The repetitive behavior scale*. Western Carolina Center Research Reports.
- Borenstein, M. (2019). Heterogeneity. In M. Borenstein (Ed.), *Common mistakes in meta-analysis and how to avoid them* (pp. 75–135). Biostat, Inc.
- Bozkus-Genc, G., & Yucesoy-Ozkan, S. (2016). Meta-analysis of pivotal response training for children with autism spectrum disorder. *Education and Training in Autism and Developmental Disabilities, 51*(1), 13–26.
- 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: RCT of caregiver-mediated early ASD intervention. *Autism Research, 10*(10), 1700–1711. <https://doi.org/10.1002/aur.1818>
- Bruinsma, Y., Minjarez, M., Schreibman, L., & Stahmer, A. (Eds.). (2020). *Naturalistic developmental behavioral interventions for autism spectrum disorder*. Brookes Publishing Co.
- 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*(6), 2211–2223. <https://doi.org/10.1007/s10803-016-2752-2>
- Crank, J. E., Sandbank, M., Dunham, K., Crowley, S., Bottema-Beutel, K., Feldman, J., & Woynaroski, T. G. (2021). Understanding the effects of naturalistic developmental behavioral interventions: A project AIM meta-analysis. *Autism Research, 14*(4), 817–834. <https://doi.org/10.1002/aur.2471>
- Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., Donaldson, A., & Varley, J. (2010). Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model. *Pediatrics, 125*(1), e17–e23. <https://doi.org/10.1542/peds.2009-0958>
- Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2019). Analysing data and undertaking meta-analyses. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions* (2nd ed., pp. 241–284). Wiley.
- 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 and Adolescent Psychiatry, 54*(7), 580–587. <https://doi.org/10.1016/j.jaac.2015.04.005>
- Fisher, Z., & Tipton, E. (2015). *robumeta: An R-package for robust variance estimation in meta-analysis*. arXiv. <http://arxiv.org/abs/1503.02220>
- Fletcher, J. (2007). What is heterogeneity and is it important? *British Medical Journal, 334*(7584), 94–96. <https://doi.org/10.1136/bmj.39057.406644.68>
- Forbes, H. J., Travers, J. C., & Johnson, J. V. (2020). A systematic review of linguistic and verbal behavior outcomes of pivotal response treatment. *Journal of Autism and Developmental Disorders, 50*(3), 766–778. <https://doi.org/10.1007/s10803-019-04307-3>
- Fuller, E. A., Oliver, K., Vejnaska, S. F., & Rogers, S. J. (2020). The effects of the Early Start Denver Model for children with autism spectrum disorder: A meta-analysis. *Brain Sciences, 10*(6), 368–368. <https://doi.org/10.3390/brainsci10060368>
- Gates, M., Gates, A., Guitard, S., Pollock, M., & Hartling, L. (2020). Guidance for overviews of reviews continues to accumulate, but important challenges remain: A scoping review. *Systematic Reviews, 9*(1), 254. <https://doi.org/10.1186/s13643-020-01509-0>
- Gates, M., Gates, A., Pieper, D., Fernandes, R. M., Tricco, A. C., Moher, D., Brennan, S. E., Li, T., Pollock, M., Lunny, C., Sepúlveda, D., McKenzie, J. E., Scott, S. D., Robinson, K. A., Matthias, K., Bougioukas, K. I., Fusar-Poli, P., Whiting, P., Moss, S. J., & Hartling, L. (2022). Reporting guideline for overviews of reviews of healthcare interventions: Development of the PRIOR statement. *British Medical Journal, 378*, e070849. <https://doi.org/10.1136/bmj-2022-070849>
- Gengoux, G. W., Abrams, D. A., Schuck, R., Millan, M. E., Libove, R., Ardel, C. M., et al. (2019). A pivotal response treatment package for children with autism spectrum disorder: An RCT. *Pediatrics, 144*, 178–182. <https://doi.org/10.1542/peds.2019-0178>
- Green, E., Stroud, L., Bloomfield, S., Cronje, J., Foxcroft, C., Hurter, K., Lane, H., Marais, R., Marx, C., McAlinden, P., O'Connell, R., Paradiice, R., & Venter, D. (2016). *Griffiths scales of child development* (3rd ed.). Western Psychological Services.
- Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *British Medical Journal, 331*, 1064–1065. <https://doi.org/10.1136/bmj.38636.593461.68>
- Hardan, A. Y., Gengoux, G. W., Berquist, K. L., Libove, R. A., Ardel, C. M., Phillips, J., Frazier, T. W., & 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. <https://doi.org/10.1111/jcpp.12354>
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*(1), 39–65. <https://doi.org/10.1002/jrsm.5>
- Ingersoll, B., Shannon, K., Berger, N., Pickard, K., & Holtz, B. (2017). Self-directed telehealth parent-mediated intervention for children with autism spectrum disorder: Examination of the potential reach and utilization in community settings. *Journal of Medical Internet Research, 19*(7), 410–420. <https://doi.org/10.2196/jmir.7484>
- Kasari, C., Paparella, T., Freeman, S., & Jahromi, L. B. (2008). Language outcome in autism: Randomized comparison of joint attention and play interventions. *Journal of Consulting and Clinical Psychology, 76*(1), 125–137. <https://doi.org/10.1037/0022-006X.76.1.125>
- Koegel, R. L., & Koegel, L. K. (2019). *Pivotal response treatment for autism spectrum disorders* (2nd ed.). Brookes Publishing Co.
- Konstantopoulos, S., & Hedges, L. V. (2019). Statistically analyzing effect sizes: Fixed- and random-effect models. In H. M. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *Handbook of research synthesis and meta-analysis* (3rd ed., pp. 245–278). Russell Sage Foundation.
- Lord, C., Charman, T., Havdahl, A., Carbone, P., Anagnostou, E., Boyd, B., Carr, T., De Vries, P. J., Dissanayake, C., Divan, G., Freitag, C. M., Gotelli, M. M., Kasari, C., Knapp, M., Mundy, P., Plank, A., Scahill, L., Servili, C., Shattuck, P., & McCauley, J. B. (2022). The Lancet Commission on the future of care and clinical research in autism. *The Lancet, 399*, 271–334. [https://doi.org/10.1016/S0140-6736\(21\)01541-5](https://doi.org/10.1016/S0140-6736(21)01541-5)
- Lord, C., & Rutter, M. (2012). *Autism diagnostic observation schedule* (2nd ed.). Western Psychological Services.
- Maenner, M. J., Warren, Z., Williams, A. R., Amoakohene, E., Bakian, A. V., Bilder, D. A., Durkin, M. S., Fitzgerald, R. T., Furnier, S. M., Hughes, M. M., Ladd-Acosta, C. M., McArthur, D., Pas, E. T., Salinas, A., Vehorn, A., Williams, S., Esler, A.,

- Grzybowski, A., Hall-Lande, J., & Shaw, K. A. (2023). Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2020. *Morbidity and Mortality Weekly Report: Surveillance Summaries*, 72(2), 1–14. <https://doi.org/10.15585/mmwr.ss7202a1>
- Magiati, I., Tay, X. W., & Howlin, P. (2014). Cognitive, language, social and behavioural outcomes in adults with autism spectrum disorders: A systematic review of longitudinal follow-up studies in adulthood. *Clinical Psychology Review*, 34(1), 73–86. <https://doi.org/10.1016/j.cpr.2013.11.002>
- Mohammadzaheri, F., Koegel, L. K., Rezaee, M., & Rafiee, S. M. (2014). A randomized clinical trial comparison between pivotal response treatment (PRT) and structured applied behavior analysis (ABA) intervention for children with autism. *Journal of Autism and Developmental Disorders*, 44(11), 2769–2777. <https://doi.org/10.1007/s10803-014-2137-3>
- Mohammadzaheri, F., Koegel, L. K., Rezaei, M., & Bakhshi, E. (2015). A randomized clinical trial comparison between pivotal response treatment (PRT) and adult-driven applied behavior analysis (ABA) intervention on disruptive behaviors in public school children with autism. *Journal of Autism and Developmental Disorders*, 45(9), 2899–2907. <https://doi.org/10.1007/s10803-015-2451-4>
- Mullen, E. M. (1995). *Mullen scales of early learning*. American Guidance Service.
- 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(1), 23–32. <https://doi.org/10.1177/1098300709334796>
- Ona, H. N., Larsen, K., Nordheim, L. V., & Brurberg, K. G. (2020). Effects of pivotal response treatment (PRT) for children with autism spectrum disorders (ASD): A systematic review. *Review Journal of Autism and Developmental Disorders*, 7(1), 78–90. <https://doi.org/10.1007/s40489-019-00180-z>
- Pieper, D., Antoine, S. L., Mathes, T., Neugebauer, E. A. M., & Eikermann, M. (2014). Systematic review finds overlapping reviews were not mentioned in every other overview. *Journal of Clinical Epidemiology*, 67(4), 368–375. <https://doi.org/10.1016/j.jclinepi.2013.11.007>
- Pollock, M., Fernandes, R., Becker, L., Pieper, D., & Hartling, L. (2023). Overviews of Reviews. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, and V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions version 6.4*. Retrieved from <http://www.training.cochrane.org/handbook>
- Rogers, S. J., & Dawson, G. (2010). *Early Start Denver Model for young children with autism: Promoting language, learning, and engagement*. Guilford Press.
- Rogers, S. J., Estes, A., Lord, C., Munson, J., Rocha, M., Winter, J., Greenon, J., Colombi, C., Dawson, G., Vismara, L. A., Sugar, C. A., Helleman, G., Whelan, F., & Talbott, M. (2019). A multisite randomized controlled two-phase trial of the Early Start Denver Model compared to treatment as usual. *Journal of the American Academy of Child and Adolescent Psychiatry*, 58(9), 853–865. <https://doi.org/10.1016/j.jaac.2019.01.004>
- Rogers, S. J., Estes, A., Lord, C., Vismara, L., Winter, J., Fitzpatrick, A., Guo, M., & 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 and Adolescent Psychiatry*, 51(10), 1052–1065. <https://doi.org/10.1016/j.jaac.2012.08.003>
- Rogge, N., & Janssen, J. (2019). The economic costs of autism spectrum disorder: A literature review. *Journal of Autism and Developmental Disorders*, 49(7), 2873–2900. <https://doi.org/10.1007/s10803-019-04014-z>
- Sandbank, M., Bottema-Beutel, K., Crowley, S., Cassidy, M., Dunham, K., Feldman, J. I., Crank, J., Albarran, S. A., Raj, S., Mahub, P., & Woynaroski, T. G. (2020). Project AIM: Autism intervention meta-analysis for studies of young children. *Psychological Bulletin*, 146(1), 1–29. <https://doi.org/10.1037/bul0000215>
- Schopler, E., van Bourgondien, M. E., Wellman, G. J., & Love, S. R. (2010). *Childhood autism rating scale* (2nd ed.). Western Psychological Services.
- Schreibman, L., Dawson, G., Stahmer, A. C., Landa, R., Rogers, S. J., McGee, G. G., Kasari, C., Ingersoll, B., Kaiser, A. P., Bruinsma, Y., McNerney, E., Wetherby, A., & Halladay, A. (2015). Naturalistic developmental behavioral interventions: Empirically validated treatments for autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 45(8), 2411–2428. <https://doi.org/10.1007/s10803-015-2407-8>
- 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(5), 1244–1251. <https://doi.org/10.1007/s10803-013-1972-y>
- Shire, S. Y., Chang, Y. C., Shih, W., Bracaglia, S., Kodjoe, M., & Kasari, C. (2017). Hybrid implementation model of community-partnered early intervention for toddlers with autism: A randomized trial. *Journal of Child Psychology and Psychiatry*, 58(5), 612–622. <https://doi.org/10.1111/jcpp.12672>
- Sidik, K., & Jonkman, J. N. (2006). Robust variance estimation for random effects meta-analysis. *Computational Statistics and Data Analysis*, 50(12), 3681–3701. <https://doi.org/10.1016/j.csda.2005.07.019>
- Sparrow, S. S., Cicchetti, D. V., & Sauliner, C. A. (2016). *The Vineland adaptive behavior scales* (3rd ed.). Pearson.
- Strain, P. S., & Bovey, E. H. (2011). Randomized, controlled trial of the leap model of early intervention for young children with autism spectrum disorders. *Topics in Early Childhood Special Education*, 31(3), 133–154. <https://doi.org/10.1177/0271121411408740>
- Tiede, G., & Walton, K. M. (2019). Meta-analysis of naturalistic developmental behavioral interventions for young children with autism spectrum disorder. *Autism*, 23(8), 2080–2095. <https://doi.org/10.1177/1362361319836371>
- Trembath, D., Varcin, K., Waddington, H., Sulek, R., Bent, C., Ashburner, J., Eapen, V., Goodall, E., Hudry, K., Roberts, J., Silove, N., & Whitehouse, A. (2023). Non-pharmacological interventions for autistic children: An umbrella review. *Autism*, 27(2), 275–295. <https://doi.org/10.1177/13623613221119368>
- Uljarević, M., Billingham, W., Cooper, M. N., Condrón, P., & Hardan, A. Y. (2022). Examining effectiveness and predictors of treatment response of pivotal response treatment in autism: An umbrella review and a meta-analysis. *Frontiers in Psychiatry*, 12, 766150. <https://doi.org/10.3389/fpsy.2021.766150>
- Veritas Health Innovation. (2020). *Covidence systematic review software* [Computer software]. Retrieved from <http://www.covidence.org>
- Vinen, Z., Clark, M., Paynter, J., & Dissanayake, C. (2018). School age outcomes of children with autism spectrum disorder who received community-based early interventions. *Journal of Autism and Developmental Disorders*, 48(5), 1673–1683. <https://doi.org/10.1007/s10803-017-3414-8>
- Vismara, L. A., McCormick, C. E. B., Wagner, A. L., Monlux, K., Nadhan, A., & Young, G. S. (2016). Telehealth parent training in the Early Start Denver Model: Results from a randomized controlled study. *Focus on Autism and Other Developmental Disabilities*, 33(2), 67–79. <https://doi.org/10.1177/1088357616651064>
- Vivanti, G., Paynter, J., Duncan, E., Fothergill, H., Dissanayake, C., & Rogers, S. J. (2014). Effectiveness and feasibility of the Early Start

- Denver Model implemented in a group-based community child-care setting. *Journal of Autism and Developmental Disorders*, 44(12), 3140–3153. <https://doi.org/10.1007/s10803-014-2168-9>
- Wang, Z., Loh, S. C., Tian, J., & Chen, Q. J. (2022). A meta-analysis of the effect of the Early Start Denver Model in children with autism spectrum disorder. *International Journal of Developmental Disabilities*, 68(5), 587–597. <https://doi.org/10.1080/20473869.2020.1870419>
- 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(2), 67–82. <https://doi.org/10.1177/02711214060260020201>
- Whiting, P., Savović, J., Higgins, J. P. T., Caldwell, D. M., Reeves, B. C., Shea, B., Davies, P., Kleijnen, J., & Churchill, R. (2016). ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*, 69, 225–234. <https://doi.org/10.1016/j.jclinepi.2015.06.005>
- Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (2011). *Preschool language scale* (5th ed.). Pearson.
- Zwaigenbaum, L., Bauman, M. L., Choueiri, R., Kasari, C., Carter, A., Granpeesheh, D., Mailloux, Z., Smith Roley, S., Wagner, S., Fein, D., Pierce, K., Buie, T., Davis, P. A., Newschaffer, C., Robins, D., Wetherby, A., Stone, W. L., Yirmiya, N., Estes, A., & Natowicz, M. R. (2015). Early intervention for children with autism spectrum disorder under 3 years of age: Recommendations for practice and research. *Pediatrics*, 136, S60–S81. <https://doi.org/10.1542/peds.2014-3667E>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.