Umbrella Review of Meta-Analyses of Comprehensive Applied Behavior Analytic Interventions for Young Children with Autism Spectrum Disorder

Brian Reichow

A.J. Pappanikou University of Connecticut Center for Excellence in Developmental Disabilities, Education, Research, and Service

UConn Health

Farmington, Connecticut

Erin E. Barton

Nashville, Tennessee

Paper commissioned by the National Academies of Sciences, Engineering, and Medicine Ad Hoc Committee on an Independent Analysis of Department of Defense's Comprehensive Autism Care Demonstration Program whose work was supported by DoD. Opinions and statements included in the paper are solely those of the individual author, and are not necessarily adopted, endorsed, or verified as accurate by the Committee or the National Academies of Sciences, Engineering, and Medicine.

SUMMARY

Background and Objective: Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by challenges in social interaction, communication, and behavior. Current estimates suggest one in 36 children in the United States has ASD (Maenner et al., 2023). Comprehensive early interventions based on the technologies of applied behavior analysis are an evidence-based practice for the treatment of young children with ASD and is a common intervention choice for young children with ASD. Multiple systematic reviews and meta-analyses have shown children receiving early comprehensive applied behavior analytic interventions have better outcomes than children in comparator groups. The objective of this review was to conduct an umbrella review of systematic reviews with meta-analytic syntheses of comprehensive applied behavior analytic interventions for young children with ASD.

Method: We searched six electronic databases and conducted supplemental searches in June of 2024 to identify relevant reviews. We included systematic reviews that included at least one meta-analytic synthesis of a child outcome from a comparative group design study. Additional inclusion criteria for types of participants, intervention characteristics, outcomes and publication status were also included. We examined the meta-analytic syntheses of the included reviews for two primary outcomes (IQ and adaptive behavior) and four secondary outcomes (communication skills, social skills, daily living skills, and autism symptomatology). We also assed risk of bias and primary study overlap for this umbrella review.

Results: Six systematic reviews with meta-analytic syntheses including 36 unique studies (CCA = 20.56%) with 2,241 child participants with ASD were included in this umbrella review. For IQ, the three meta-analyses reporting SMD showed an effect size estimate greater than .50 and the

three meta-analyses reporting the MD effect size showed an estimated effect of 11.98 or greater (on a scale with a SD = 15). For adaptive behavior, the three SMD estimates were all greater than 0.35 and the three estimates for MD were equal to 5.92 or greater. Evidence of heterogeneity was detected in 5 of 6 analyses for IQ and 2 of 6 analyses for adaptive behavior. Effects on secondary outcomes were reported less frequently across reviews with greater variability in findings across meta-analyses, with 17 of 19 (89%) effect size estimates had a SMD or MD effect size greater than 0.20 or 3.0, respectively.

Conclusion: The findings of this umbrella review demonstrate consistent, replicated evidence that comprehensive early interventions based on the technologies of applied behavior analysis were superior to comparison conditions on the primary outcomes (IQ and adaptive behavior) for a majority of the young children with ASD who received the intervention across numerous controlled trials.

Protocol registration: International Prospective Register of Systematic Reviews (PROSPERO; CDR#42024568638).

INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental condition characterized by challenges in social interaction, communication, and behavior (American Psychiatric Association, 2013). Early intervention is widely recognized as critical for optimizing outcomes for young children with ASD. Among the various approaches, comprehensive applied behavior analytic interventions—often delivered through Early Intensive Behavioral Intervention (EIBI) models—aim to improve cognitive and adaptive behaviors by targeting multiple developmental domains. Typically, comprehensive interventions involve at least 10 hours of weekly therapy (often between 20 and 40 hours) over a duration lasting from 24 to 36 months (cf., Odom et al., 2010, 2014). Children with ASD typically begin comprehensive treatment programs when they are between 30 and 60 months of age, as it is thought that optimal outcomes are achieved when treatment is started early (e.g., Fein et al., 2013; Koegel et al., 2014; National Research Council, 2001). Despite their prominence in the field, important questions remain about the overall efficacy of these interventions, the characteristics of effective programs, and how various delivery models impact outcomes.

A central feature of comprehensive applied behavior analytic interventions is that services are often delivered by trained technicians or paraprofessionals, such as registered behavior technicians, under the supervision of clinicians or program managers. This model, though efficient and scalable, introduces variability in implementation fidelity and raises questions about the role of delivery agents in achieving optimal outcomes. Research has not consistently shown if differences in delivery expertise or intervention management consistently effects outcomes. Moreover, the diversity of intervention characteristics, including intensity,

duration, and target outcomes, makes it challenging to draw definitive conclusions about best practices.

Given these complexities, there is a pressing need for a comprehensive synthesis of the evidence on the effectiveness of comprehensive applied behavior analytic interventions for young children with ASD. Systematic reviews and meta-analyses have independently examined these interventions, but the variability in their inclusion criteria, analytic methods, and reporting standards underscores the need for an overarching evaluation. An umbrella review provides a unique opportunity to aggregate findings across systematic reviews, offering a higher-level synthesis that can address broad research questions and identify consistent patterns in the data.

By synthesizing findings from systematic reviews with meta-analytic syntheses, we seek to identify systematic patterns in the evidence of the effects of comprehensive applied behavior analytic interventions for young children with ASD. The findings from this umbrella review are expected to guide practitioners, policymakers, and researchers in making informed decisions regarding the design, implementation, and evaluation of comprehensive applied behavior analytic interventions, ultimately contributing to improved outcomes for children with ASD. Our objective was to conduct an umbrella review (i.e., overview of reviews; Pollock et al., 2023) of systematic reviews with meta-analytic syntheses of comprehensive applied behavior analytic interventions for young children with ASD.

METHOD

Review Protocol

A protocol for this overview was developed and the review was registered a priori at the International Prospective Register of Systematic Reviews (PROSPERO; CDR#42024568638). Deviations from the protocol are shown, with explanation, in Supplemental Table 1. The review is consistent with standards of the Preferred Reporting Items for Overview of Reviews (PRIOR; Gates et al., 2022). Deviations between the review and protocol are shown in Supplemental Table 1.

Study Selection

Types of reviews. We included systematic reviews with at least one meta-analytic synthesis of a child outcome from a comparative group design study (i.e., randomized controlled trial, clinical controlled trial, two group pre-test/post-test comparison). We did not include systematic reviews that included one or more single case experimental design studies or qualitative research design studies. We also amended our protocol to limit our inclusion criteria to systematic reviews containing meta-analytic syntheses of comparative group design studies (i.e., excluding pre/post change designs).

Types of participants. We included reviews in which the pre-treatment average age of the child participants was under 7-years-old or reviews in which greater than 50% of the included primary studies reported a pre-treatment mean age under 7-years-old. We included reviews for which 100% of primary studies included a majority of children with an ASD (American Psychological Association, 2013). Finally, we included reviews based on age and

diagnostic criteria irrelevant of levels of intellectual functioning (i.e., we included reviews in which primary studies include children with and without intellectual disability or IQ < 70).

Types of interventions. We included reviews in which 100% of the primary studies evaluated a comprehensive applied behavior analytic intervention. For this overview, comprehensive applied behavior analytic interventions must have included the following characteristics (adapted from Odom et al. 2010): a) manualized treatment based on the technologies of applied behavior analysis and the science of human behavior; b) high intensity, as defined by 10 or more (on average) hours of treatment per week; c) longevity, as defined by a treatment duration (on average) of at least 6 months; and d) comprehensive, as defined by the treatment of at least two different developmental domains. For this umbrella review, 100% of primary studies included in a review needed to meet all four characteristics of comprehensive applied behavior analytic interventions defined above. Studies in which caregivers were the primary intervention agent (i.e., the individual providing the direct treatment to the child) were excluded from this review.

Types of outcome measures. We included reviews reporting a meta-analytic synthesis of at least one child outcome. The primary outcome measures for children in this umbrella review were IQ and adaptive behavior (i.e., adaptive behavior composite). Secondary outcomes, which were not required for a review to meet inclusion criteria, included communication skills, social skills, daily living skills, and autism symptomatology. Caregiver and family outcomes were not considered when determining the inclusion or exclusion of a review.

Additional inclusion criteria. We included meta-analyses published between 2009 and 2024. We selected this period because 2009 was the year in which the first meta-analyses of

EIBI were published (e.g., Eldevik et al., 2009; Reichow & Wolery, 2009). We also limited the inclusion to articles published in English in peer-reviewed journals.

Search Methods for Identification of Studies

Electronic database searches. We searched Medline, Embase, APA PsycINFO, Academic Search Premier, the Cumulative Index to Nursing and Allied Health (CINAHL), and the Education Resources Information Center (ERIC) for relevant reviews in June 2024. To locate records, we used combinations of terms related to participants (e.g., autism, pervasive developmental disorder) and intervention characteristics (e.g., early intensive behavioral intervention, intensive behavioral intervention, comprehensive model treatment) using both natural language and controlled vocabulary. The electronic search strategies for each database are shown in Supplemental Text 1. Records from the electronic database searches were combined and deduplicated using the Covidence systematic review software (Veritas Health Information).

Searching other resources. We also used "snowball methods" (Greenhalgh & Peacock, 2005) by examining the reference lists of included studies and by searching for studies citing the included studies. We used the Citation Chaser online application (https://estech.shinyapps.io/citationchaser/; Haddaway et al., 2021) to conduct the snowball searches by inputting the DOI of each review and using the Citation Chaser functions to download the references lists from each review (backward citation chasing) and references for citations to each review (forward citation chasing). We then combined and deduplicated the records from Citation Chaser across reviews for screening in Covidence.

Identification and selection of studies. Two authors independently screened all titles and abstracts to exclude clearly irrelevant records. For a record to be moved to the full-text

screening stage, at least one screener had to indicate the full text should be examined against the inclusion criteria. We then obtained full text copies of each record (article) for full-text screening. Two authors assessed each article against the inclusion criteria to determine which reviews met all inclusion criteria. Disagreements were resolved through discussion to reach a consensus decision. Data from multiple reports of the same review were linked together and used to supplement information obtained from the primary report.

Data Extraction

Consistent with methodological standards for overview of reviews (e.g., Gates et al., 2020; Pollock et al., 2023), we extracted data primarily from the data reported in the published articles of the included meta-analyses. To extract data, two authors independently performed data extraction using the Covidence systematic review software, with discrepancies resolved through discussion and consensus. We extracted data on the following variables:

1. Review characteristics: including, but not limited to: publication year; type of review, search date, search methods
2. Population characteristics: including, but not limited to: sample size, diagnostic characteristics, child age, child skill level, gender, practitioner characteristics
3. Intervention characteristics: including, but not limited to: intervention components; intervention delivery methods; training format; intervention location; trainer qualifications; trainer training and supervision; and intervention density, including number of group sessions, number of individual sessions, number of home sessions, duration of each session, and duration, in weeks, of complete program

4. Study results and effects: including, but not limited to: outcome measures used; effect size estimates (i.e., mean difference [MD], standardized mean difference [SMD]), including point estimates and confidence intervals; results of moderator analyses. We extracted results and effects for primary and secondary child outcomes We chose to designate IQ and adaptive behavior composite as primary outcomes because these two outcomes were used as primary outcomes across both primary studies and reviews. All other child outcomes that were reported in at least one review were designated as secondary outcomes, which included generalized communication / language skills, expressive communication, receptive communication, social skills, daily living skills, and autism symptomatology.

Assessment of Risk of Bias

Review-level 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 systematic reviews. As with data extraction, two authors independently assessed the risk of bias of each review with disagreements resolved through discussion to consensus. For the risk of bias assessment, we examined 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, a summary-level judgement was made by collating the concerns of risks of bias identified for each domain. Finally, we rated the overall risk of bias (i.e., high, low, or unclear) for each review. Because the Whiting risk of bias in systematic review tool was designed to assess reviews of medical interventions, we also possible biases in the included reviews using the JBI Critical Appraisal Checklist for Systematic Reviews and Research Syntheses (Aromataris et al., 2015). The JBI

Checklist contains 11 items that helps to assess the methodological rigor of a review and the extent to which the review has potentially addressed possible causes of bias. For this umbrella review, we evaluated the first nine items of the JBI Checklist (we excluded items 10 and 11 related to recommendations and directives of the review). To complete the Checklist, two reviewers independently evaluated the nine items for each review with disagreements resolved through discussion and reaching consensus. Data were analyzed descriptively by creating a summary figure across reviews for each of the nine items on the JBI Checklist.

Risk of bias of primary studies. We did not assess the risk of bias for the primary studies included across the reviews. To examine the risk of bias at the primary study level, we extracted risk of bias or methodological rigor evaluations reported from the included reviews. We narratively summarized review-level bias or methodological rigor based on the evaluations and methods used across the included reviews.

Data Synthesis and Analysis

Primary study overlap. We examined the overlap of primary studies using the corrected covered area (CCA; Pieper et al., 2014), which quantifies the degree of primary study overlap across reviews. We used "primary publications" (i.e., articles; Pieper et al., p. 370) as the unit of analysis/calculation. The CCA was calculated using $CCA = \frac{N-u}{uc-u}$, where *N* is the number of included primary publications (including double counting), *u* is the number of primary publications (excluding duplicated reports), and *c* is the number of systematic reviews. We used Pieper and colleagues' guidelines for quantifying the level of CAA for slight (0 – 5%), moderate (5 – 10%), high (10 – 15%), or very high (> 15%) levels of overlap. We also explored the overlap

of primary studies across reviews using graphical methods (e.g., Bougioukas et al., 2021; Bracchiglione et al., 2022).

Synthesis and analysis across reviews. We used descriptive and narrative syntheses to draw conclusions across the included reviews. We first examined the estimated magnitude of effects shown by the meta-analytic syntheses in each review. All meta-analyses included in this overview provided magnitude estimates using either the mean difference effect size or the standardized mean difference effect size. We then aggregated findings across reviews by constructing tables and harvest plots for each outcome category (McKenzie & Brennan, 2019; Ogilivie et al., 2008). A harvest plot is a method of vote counting using a graphical display of treatment effects, which allows for visual analysis of the strength of treatment effects across variables. As such, the display of the estimates of effects for the meta-analytic syntheses of comprehensive applied behavior analytic treatments are shown based on estimates of magnitude (not statistical significance). For this umbrella review, we constructed two harvest plots, one depicting the primary outcomes (i.e., IQ, adaptive behavior composite) and one depicting the secondary outcomes (i.e., generalized communication skills, expressive communication, receptive communication, socialization, daily living skills, and autism symptomatology). We grouped the effects (depicted using the bars in the harvest plot) using effect size estimates for no effects (SMD < .20 or MD < 2.5), small effects (SMD 0.20 to 0.50 or MD 2.5 to 7.5), medium effects (SMD 0.50 to 0.80 or MD 7.5 to 12.5), and large effects (SMD > 0.80 or MD > 12.5). For our analyses we considered medium and large effects to be clinically significant.

Results

Review Selection

The electronic database search identified 1,097 records, and the snowball search using Citation Chaser identified an additional 1,680 records. After removing 378 duplicates, 2,395 records remained and underwent title and abstract screening. We removed 2,326 records based on titles and abstracts alone, which left 69 records for full-text screening. After full-text review, six systematic reviews with meta-analyses met all inclusion criteria and are included in this review (Eckes et al., 2023; Eldevik et al., 2009; Peters-Scheffer et al. 2011; Reichow et al., 2018; Rodgers et al., 2020; and Shi et al., 2021). A PRIMSA flow diagram of review selection, with reasons for exclusion of records excluded by full-text review, is shown in Figure 1.

Selection of primary review. Two reviews (Reichow et al., 2018 and Rodgers et al., 2020) were reported across multiple publications. Reichow and colleagues published a review protocol with Cochrane in 2011 (Reichow et al., 2011) with a full review published one year later in 2012 (Reichow et al., 2012). The full review was also co-published by the Campbell Collaboration (Reichow et al., 2014). A review update was published in 2018 (Reichow et al., 2018), which was the reference used in this umbrella review because it includes the most recent version and analyses. Rodgers and colleagues led a collaborative review – The Synthesizing Comprehensive Applied Behaviour Analysis Interventions: Research for Children with ASD – the results of which have been published in two peer-reviewed sources (Rodgers et al., 2020; 2021). The data used for the umbrella review were obtained mostly from the 2020 report, thus Rodgers (2020) was used for our purposes.

Selection of meta-analytic synthesis. Two reviews (Rodgers et al., 2020; Shi et al., 2021) included two sets of meta-analytic syntheses. Rodgers et al. included an individual patient data (IPD) meta-analysis with 10 primary studies and a traditional meta-analysis that included these 10 studies plus data from 6 additional studies that were included in their systematic review. This umbrella review includes the findings of the traditional meta-analytic syntheses because the selection of studies (i.e., studies meeting the inclusion criteria of the review rather than reliance of studies that were able and willing to provide raw individual participant data) matched the selection methods of other meta-analyses included in this review. Shi and colleagues (2021) also conducted two sets of meta-analytic syntheses. They conducted one set of meta-analytic syntheses of the difference between treatment and comparison groups and one set of meta-analytic syntheses of pre-intervention to post-intervention change in treatment groups, which included one-group pre/post design studies without a comparator. Because this umbrella review focuses on syntheses of group comparative design studies, we included the analyses from Shi et al. for the comparative design studies and excluded the standardized mean change meta-analyses of treatment change scores. For transparency, the findings of the IPD meta-analysis of Rodgers et al. and the results of the meta-analyses of standardized mean change effect sizes of Shi et al. are shown in Supplemental Table 2.

Review Characteristics

The six reviews that met our inclusion criteria were published between 2009 and 2023. All six reviews retrieved studies using electronic database searches, with five of the six reviews utilizing supplemental search methods such as reference list screening, hand searching, and author contact to locate additional records. All six reviews included studies that used both

randomized control trial designs and quasi-experimental group comparative designs; the range of the percentage of randomized studies was 9% (Eckes et al., 2023) to 37.5% (Shi et al., 2021). Five reviews evaluated the risk of bias or methodological rigor of primary studies included in their reviews. Across reviews, five evaluation tools were used including the Evaluative Method (Reichow et al., 2008; used by Shi et al.), the Quality Index (Downs & Black, 1998; used by Peters-Scheffer et al., 2011), the original Cochrane Risk of Bias Tool (Higgins et al., 2011; used by Reichow et al., 2018 and Eckes et al.), the Cochrane Risk of Bias 2.0 (Cochrane Methods Group, 2019; used for RCTs by Rodgers et al., 2020), and the Risk of Bias in Non-randomized Studies of Interventions tool (Sterne et al., 2016; used by Rodgers et al.). The characteristics of the six included reviews are shown in Table 1.

All six meta-analyses utilized a random effects model when combining the results of the primary studies. Three reviews used the standardized mean difference (SMD) effect size (i.e., Cohen's *d*, Hedges' *g*) and three reviews used the mean difference (MD) effect size estimate. For this overview, we have extracted effect size data using the metric presented in each review (i.e., we did not convert MD to SMD or vice-versa). When interpreting effect sizes, we provide interpretations for magnitude in both SMD units and MD units. Heterogeneity was estimated in all six reviews, with the *Q*-statistic used in three reviews, I^2 used in five reviews, and tau-squared utilized in three reviews. Due to the small number of included primary studies, publication bias was difficult to assess, which was noted by two reviews (Eckes et al., 2023; Reichow et al., 2018). For the reviews that examined publication bias, visual inspection of funnel plots was the most common method used, with Trim and Fill (e.g., Duval & Tweedie, 2000) utilized in one review (Eldevik et al., 2009) and Egger's test (e.g., Egger et al., 1997)

utilized in one review (Shi et al., 2021). The possibility of publication bias was only detected once across all meta-analytic syntheses of the included reviews; Eckes et al. (2023) detected the possibility of publication bias in their synthesis of adaptive behavior composite.

All reviews reported findings for meta-analytic syntheses of the two primary outcomes (IQ and adaptive behavior composite). Five reviews reported the findings of a meta-analysis for at least one secondary outcome; Eldevik et al. (2009) did not report findings from statistical syntheses of a secondary outcome. The range of meta-analytic syntheses conducted within a review for our primary and secondary outcomes had a range of two syntheses (Eldevik et al., 2009) to eight syntheses (Reichow et al., 2018). The review by Reichow and colleagues was the only review to statistically synthesize all of the primary and secondary outcomes of this umbrella review. Different assessment instruments were used for the reported outcomes across primary studies (and by extension reviews). A table of assessment instruments used across primary studies for each primary and secondary outcome in each review is shown in Supplemental Table 3.

Primary Study Overlap

The number of primary studies (*u*) in each review ranged from five (Reichow et al., 2018) to 20 (Rodgers et al., 2020). Across the six reviews, the cumulative number of included primary studies summed to 73. This figure is a gross count of primary studies that includes a count of primary studies included in multiple reviews. Across reviews, the number of primary publications (i.e., unduplicated studies) was 36; characteristics and findings of these studies is shown in Supplemental Table 4. Over 55% of primary studies across the six reviews were included in one review. Two primary studies were included in all six reviews (Howard et al.,

2005; Remington et al., 2007). Fifteen studies were included in multiple reviews including six primary studies that were included in two reviews (Birnbrauer & Leach, 1993; Dawson et al., 2010; Eikeseth et al., 2012; Reed et al., 2007; Smith et al., 1997; Zachor et al., 2010), two primary studies that were included in three reviews (Lovaas, 1987; Sallows & Graupner, 2005), three primary studies that were included in four reviews (Cohen et al., 2006; Eikeseth et al., 2002; Eldevik et al., 2006), and two primary studies that were included in five reviews (Magiati et al, 2007; Smith et al., 2000). The overlap of primary studies across reviews, estimated by the corrected covered area (CCA) was 20.56%, indicating very high overlap. We used the GROOVE tool (Bracchiglione et al., 2022) to create a summary citation matrix showing the percentage of pairwise overlap between reviews, which is shown in Figure 2. As seen in Figure 2, there was very high overlap (≥ 15%) between all pairwise overlap comparisons (nodes) across reviews except for the overlap between Eckes et al. (2023) and Shi et al. (2021), for which had high overlap (10% to 15%).

Participant Characteristics

Participant characteristics are shown in Table 2. As shown in Table 2, there were 2,241 child participants across the primary studies included in meta-analytic syntheses, which is an average of about 44 participants per study. Across the five reviews that reported data on gender, all five reviews contained samples in which the reported percentage of male participants was greater than 80% (range 82.2% to 86.5%). The total number of male participants in the primary studies across reviews was 1,215 (84.8%); 218 (15.2%) participants were identified as female. All six reviews had samples in which the mean age of the children at treatment onset was under the age of 66 months old (range 20.6 to 65.6 months).

Pretreatment cognitive (IQ) estimates across studies had a range of mean IQ scores between 27.6 to 76.5. Collectively, the reviews represent a homogenous sample of young children with autism who have delays in development. Although participant characteristics of primary studies were an inclusion criterion of this overview, this was not a primary reason of exclusion for many reviews that were excluded from this overview.

Intervention Characteristics

Characteristics of the interventions used in the primary studies included in reviews are shown in Table 3. All primary studies across the included reviews examined comprehensive applied behavior analytic interventions. Five of six reviews exclusively examined a format of EIBI to a comparator, with the other review (Rodgers et al., 2020) including two primary studies that utilized a format of the Early Start Denver Model (Dawson et al., 2010; Rogers et al., 2019) that met our criteria of comprehensive applied behavior analytic treatments. One review (Shi et al., 2021) included comprehensive applied behavior analytic interventions categorized as Early Start Denver Model (Estes et al., 2015; Vinen et al., 2018) and other types of comprehensive behavioral interventions (Akshoomoff et al., 2010; Clark et al., 2017; Gabriels et al., 2001; Landa & Kalb, 2012), but the findings from these studies were not included in their meta-analytic syntheses. The type of comparator groups for the included primary studies varied across reviews, with treatment as usual or eclectic approaches being the most common comparator. Consistent with a comprehensive intervention approach, all reviews reported studies with high intervention densities and durations that typically lasted between one and three years.

Risk of Bias

A summary of risk of bias assessed using the Whiting risk of bias tool (Whiting et al., 2016) across the six included reviews is shown in Supplemental Table 5 and Supplemental Figure 1. Overall, five of the six (83%) included reviews (Eckes et al., 2023; Peters-Scheffer et al., 2011; Reichow et al., 2018; Rodgers et al., 2020; Shi et al., 2021) were judged to have a low risk of bias. The highest levels of concern were seen in the domains of synthesis and findings (4 of 6 reviews rated as having a high risk of bias) and data collection and study appraisal (one review rated as having a high risk of bias and one review rated as having an unclear risk of bias). Lower risks of bias was found for the domain of study eligibility criteria, for which all reviews were judged to have had a low risk of bias.

Overall, the results of the evaluation of risk of bias using the JBI Appraisal Checklist (Aromataris et al., 2015) showed low risks of bias in most reviews. As shown in Supplemental Table 6 and Supplemental Figure 2, five reviews (Eckes et al., 2023; Peters-Scheffer et al., 2011; Reichow et al., 2018; Rodgers et al., 2020; Shi et al., 2021) were judged to have low risk of bias on at least 7 of the 9 JBI Checklist items with three reviews having low risk of bias on at least 8 items. Across reviews, the JBI Checklist items for which bias had the highest risk was the minimization of data extraction errors, which was not clearly reported in three of the six reviews. The provision of a clear description of the search strategy and handling of publication bias were both rated as having possible risks of bias in 2 of 6 reviews. No other checklist criteria were rated as a risk of bias for more than one of the six included reviews.

Effects of Comprehensive Applied Behavior Analytic Interventions

Findings for Primary Outcomes. All reviews reported meta-analytic syntheses for the two primary outcomes (IQ and adaptive behavior composite). Figure 1 shows harvest plot depictions for the findings of the meta-analyses for the primary outcomes (IQ and adaptive behavior composite) and details for the findings are shown for IQ and adaptive behavior composite in Tables 4 and 5, respectively. The findings across reviews for IQ were robust; the 3 meta-analyses reporting SMD showed an effect size estimate greater than .50 and the 3 meta-analyses reporting the MD effect size showed an estimated effect of 11.98 or greater. Estimates of heterogeneity were high for five of the six analyses and no analysis reported finding evidence of publication bias. While the findings for adaptive behavior composite are not as robust as IQ, the findings are still noteworthy, with the SMD estimates all greater than 0.35 and the three estimates for MD equal to 5.92 or greater. Heterogeneity was high in two reviews (Rodgers et al., 2020; Eckes et al., 2023), with the possibility of publication bias detected in the Eckes review through the use of Egger's test ($p < .001$).

Findings for Secondary Outcomes. Meta-analytic findings were reported for at least two reviews for the secondary outcomes of generalized language, expressive communication, receptive communication, socialization, daily living skills, and autism symptomatology. Supplemental Figure 1 shows harvest plots for the secondary outcomes and the results by outcome are shown for each review in Supplemental Tables 7-12. Overall, the meta-analytic syntheses for the secondary outcomes were less universal than the analyses for primary outcomes, in which all six reviews conducted analyses of IQ and adaptive behavior composite, with one secondary outcome (generalized communication) analyzed in four reviews and five secondary outcomes analyzed in three reviews. As shown in Supplemental Figure 1, most

outcomes showed positive findings, which were clinically significant for 7 of 19 (37%) analyses across outcomes and reviews (review-level details by outcome are shown in Supplemental Tables 7-12). For the three secondary outcomes related to communication, 2 of 4 effect size estimates were clinically significant for the generalized communication skills outcome and 2 of 3 effect size estimates were clinically significant for expressive and receptive communication apiece. The only other effect size estimate that was clinically significant was one of the three estimates for daily living skills. No effect size estimates were clinically significant for social skills or autism symptomatology, for which both outcomes had three effect size estimates below clinical significance.

Discussion

This umbrella review of comprehensive applied behavior analytic interventions for young children (aged 2- to 9-years-old) with ASD provides a current synthesis of six meta-analyses published since 2009. The six meta-analyses included 37 unique studies with 2,241 young children with autism. The corrected covered area (CCA) for the overview indicated very high overlap of primary studies between reviews (CCA = 20.56%). The children were typically between 30 and 60 months old at the beginning of the intervention studies and received comprehensive behavior analytic interventions that typically had 13 to 40 hours of treatment per week for 24 to 36 months. While there was variability across the studies included in the meta-analytic syntheses with respect to participant ages and intervention density, these average ranges are consistent with the extant literature on early behavior analytic interventions for young children with ASDs. This, coupled with the very high overlap of primary studies suggests the meta-analyses included in the umbrella review examined primary studies of a

similar topic (i.e., comprehensive applied behavior analytic interventions for young children with ASD).

All six meta-analyses showed children receiving the comprehensive applied behavior analytic treatment, on average, had superior outcomes for IQ and adaptive behavior, which are the two most commonly assessed outcomes in the extant studies on comprehensive programs for young children with ASD. The findings for IQ were stronger than the findings for adaptive behavior. All six meta-analyses showed effect size estimates SMD > 0.50/MD > 7.5 for IQ, with three meta-analyses showing SMD > 0.80/MD > 12.0 (Eldevik et al., 2009; Reichow et al., 2018; Rodgers et al., 2020). For adaptive behavior, three meta-analyses showed effect size estimates SMD 0.20 – 0.50/MD 3.0 – 7.5 and three meta-analyses showed effect size estimates for adaptive behavior of SMD 0.50 – 0.80/MD 7.5 to 12.0 (Eldevik et al., 2009; Reichow et al., 2018; Rodgers et al., 2020). Further, the interventions were implemented with fidelity across contexts, suggesting the applied utility of the intervention (cf., Wergeland et al., 2022). Overall, the findings demonstrate consistent, replicated evidence that the intervention was superior to the comparison condition for these two clinically meaningful outcomes for the majority of the children who received the intervention across numerous controlled trials.

While the findings for other outcomes, including communication skills, socialization skills, daily living skills, and autism symptomatology are less robust, 17 of 19 (89%) effect size estimates had a SMD or MD effect size greater than 0.20 or 3.0, respectively (see Supplemental Figure 3). Examination of the effect size estimates for these secondary outcomes suggest the most robust findings were related to communication skills; 2 of 4 meta-analyses examining generalized communication skills had effect size estimates SMD 0.50-0.80/MD 7.5 to 12.0 and 2

of 3 (67%) meta-analyses had effect size estimates SMD > 0.50/MD > 7.5 for both expressive and receptive communication skills. Collectively, the results of this umbrella review show children receiving comprehensive applied behavior analytic interventions, on average, had better outcomes than children receiving the comparison conditions.

This umbrella review focused on findings from meta-analyses conducted using the standardized mean difference or mean difference effect sizes, which compare mean scores between intervention and comparison groups. Two reviews included in this umbrella review (Rodgers et al., 2020 and Shi et al., 2021) also included findings for average change scores for children in the intervention group from pre-intervention to post-intervention (i.e., standardized mean change or mean change effect sizes). Analysis of pre/post change effects in the Shi et al. (2021) and Rodgers et al. (2020) reviews showed similar findings for IQ (SMC = 0.85 [95% CI 0.47, 1.22] and MC = 11.97 [95% CI 6.74, 17.20]), respectively; see Supplemental Table 2). Two contemporary reviews (i.e., Makrygiani et al., 2018; Wergeland et al., 2022) and two seminal reviews (i.e., Reichow & Wolery, 2009; Virues-Ortega, 2010) that relied on a standardized mean change effect size showed similar results for IQ (SMC ES 0.69 to 1.19 for IQ and SMC ES 0.42 to 1.09 for adaptive behavior, see Supplemental Table 13). The findings of this umbrella review are also consistent with the findings of earlier extant umbrella reviews of comprehensive early intervention for young children with ASDs that concluded intensive comprehensive early interventions based on applied behavior analysis showed superior gains in IQ and adaptive behavior compared to treatment as usual, less intensive behavioral interventions, or eclectic approaches (Magiati et al., 2012; Reichow, 2012).

Limitations

Although the use of an umbrella review is indicative of the accumulation of a critical mass of systematic reviews and meta-analyses on a topic, the review method is descriptive in nature and does not allow for examination of possible moderators of best outcome. Thus, we are not able, in this umbrella review, to assess specific characteristics of the intervention or characteristics of children that might be most likely to have a positive outcome after receiving the intervention. The umbrella review is also limited by the relatively small number of reviews, six, meeting our inclusion criteria and the relatively small number of primary studies, 37, included across reviews. As noted earlier in the discussion, this was likely the result of tight inclusion criteria to ensure all primary studies of each included meta-analysis examined a comprehensive applied behavior analytic intervention. While this might be considered a limitation, umbrella reviews that have tight inclusion criteria around intervention types increase the confidence in conclusions about the effectiveness or efficacy of the intervention being reviewed. Examination of other meta-analyses not meeting our inclusion criteria (e.g., Makrygiani et al., 2010, 2018; Reichow & Wolery, 2009; Spreckley & Boyd, 2009; Virues-Ortega, 2010; Wergeland et al., 2022; Yu et al., 2020) showed similar findings to the ones of this umbrella review. Characteristics and findings of these key excluded reviews are shown in Supplemental Table 13, for reference. Additionally, the risks of bias of the included reviews, while on average, was low, did include reviews with higher risks of bias. The risks of bias of the primary studies, as assessed in the included reviews, also indicated risks of bias associated with the lack of randomization and the lack of blind outcome assessors. These risks of bias at both

the primary study and review level should be considered when interpreting the overall findings of this umbrella review.

Conclusion

This umbrella review of six meta-analytic syntheses of comprehensive applied behavior analytic interventions for young children with ASD showed the intervention was superior to comparison in all six statistical syntheses for IQ and adaptive behavior, with positive, but less robust, findings for communication skills. The findings of this umbrella review align with those from other meta-analytic syntheses and umbrella reviews on comprehensive early intervention in autism, reinforcing the robustness and reliability of the conclusions. Collectively, comprehensive applied behavior analytic interventions for young children with ASD should be considered a robust treatment.

References

Meta-analyses included in umbrella review indicated with an *

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders,* 5th ed. Author.

Akshoomoff, N., Stahmer, A. C., Corsello, C., & Mahrer, N. E. (2010). What happens next? Follow-up from the children's toddler school program. *Journal of Positive Behavioral Interventions, 12*(4), 245–53. https://doi.org/10.1177/1098300709343724

Aromataris, E., Fernandez, R., Godfrey, C. M., Holly, C., Khalil, H., & Tungpunkom, P. (2015). Summarizing systematic reviews: Methodological development, conduct and reporting of an umbrella review approach. *International Journal of Evidence-based Healthcare, 13*(3), 132-140. https://doi.org/10.1097/XEB.0000000000000055

Birnbrauer, J. S., & Leach, D. J. (1993). The Murdoch early intervention program after 2 years. *Behaviour Change, 10*(2), 63–74. https://doi.org/10.1017/S0813483900005556

Bougioukas, K. I., Vounzoulaki, E., Mantsiou, C. D., Savvides, E. D., Karakosta, C., Diakonidis, T., Tsapas, A., & Haidich, A. B. (2021). Methods for depicting overlap in overviews of systematic reviews: An introduction to static tabular and graphical displays. *Journal of Clinical Epidemiology, 132,* 34-45. https://doi.org/10.1016/j.jclinepi.2020.12.004

Bracchiglione, J., Meza, N., Bangdiwala, S. I., de Guzman, E. N., Urrutia, G., Bonfill, X., & Madrid, E. (2022). Graphical representation of overlap of overviews: GROOVE tool. *Research Synthesis Methods, 13,* 381-388. https://doi.org/10.1002/jrsm.1557

Clark, M. L., Barbaro, J., & Dissanayake, C. (2017). Continuity and change in cognition and autism severity from toddlerhood to school age. *Journal of Autism and Developmental Disorders, 47*, 328–39. https://doi.org/10.1007/s10803-016-2954-7

Cochrane Methods Group. (2019). 'Risk of bias 2.0' tool. https://methods.cochrane.org/risk-bias-20-tool

Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: Replication of the UCLA model in a community setting. *Journal of Developmental and Behavioral Pediatrics, 27*(2), S145–S155. https://doi.org/10.1097/00004703-200604002-00013

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*, e17–23. https://doi.org/10.1542/peds.2009-0958

Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health, 52*(3), 377-384. https://doi.org/10.1136/jech.52.6.377

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463. https://doi.org/10.1111/j.0006-341x.2000.00455.x

*Eckes, T., Buhlmann, U., Holling H. D., & Mollmann, A. (2023). Comprehensive ABA-based interventions in the treatment of children with autism spectrum disorder: A meta-analysis. *BMC Psychiatry, 23,* 133. https://doi.org/10.1186/s12888-022-04412-1

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997) Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315,* 629–634. https://doi.org/10.1136/bmj.315.7109.629

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(*2), 829-835. https://doi.org/10.1016/j.rasd.2011.09.002

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*(1)*,* 49–68. https://doi.org/10.1177/0145445502026001004

Eldevik, S., Eikeseth, S., Jahr, E., & Smith, T. (2006). Effects of low-intensity behavioral treatment for children with autism and mental retardation. *Journal of Autism and Developmental Disorders, 36*(2), 211–224. https://doi.org/10.1007/s10803-005-0058-x

*Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2009). Meta-analysis of early intensive behavioral intervention for children with autism. *Journal of Clinical Child & Adolescent Psychology, 38*(3), 439-450. https://doi.org/10.1080/15374410902851739

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

Fein, D., Barton, M., Eigsti, I.-M., Kelley, E., Naigles, L., Schultz, R. T., Tyson, K. (2013). Optimal outcome in individuals with a history of autism. *Journal of Child Psychology and Psychiatry, 54*(2), 195–205. https://doi.org/10.1111/jcpp.12037

Gabriels, R.L., Hill, D. E., Pierce, R. A., Rogers, S. J., & Wehner, B. (2001). Predictors of treatment outcome in young children with autism: a retrospective study. *Autism, 5*(4), 407–29. https://doi.org/10.1177/1362361301005004006

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. *BMJ*, *378,* e070849. https://doi.org/10.1136/bmj-2022-070849

Greenhalgh, T., & Peacock, R. (2005). Effectiveness and efficiency of search methods in systematic reviews of complex evidence: Audit of primary sources. *BMJ*, *331*, 1064–1065. https://doi.org/10.1136/bmj.38636.593461.68

Haddaway, N. R., Grainger, M. J., Gray, C. T. (2021). Citationchaser: An R package and Shiny app for forward and backward citations chasing in academic searching. https://doi.org/10.5281/zenodo.4543513

Higgins, J. P. T., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., Savovic, J., Schulz, K. F., Weeks, L., Sterne, J. A. C., & Cochrane Bias Methods Group. (2011). The Cochrane Collaborations tool for assessing risk of bias in randomized trials. *BMJ, 343,* d5928. https://doi.org/10.1136/bmj.d5928

Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatment for young children with autism. *Research in Developmental Disabilities, 26*(12)*,* 359–383. https://doi.org/10.1016/j.ridd.2014.08.021

Koegel, L. K., Koegel, R. L., Ashbaugh, K., & Bradshaw, J. (2014). The importance of early identification and intervention for children with or at risk for autism spectrum disorders. *International Journal of Speech-Language Pathology, 16*(1), 50-56. https://doi.org/10.3109/17549507.2013.861511

Landa, R. J., & Kalb, L. G. (2012). Long-term outcomes of toddlers with autism spectrum disorders exposed to short-term intervention. *Pediatrics, 130*(Suppl. 2), S186–90. https://doi.org/10.1542/peds.2012-0900Q

Lovaas, O. I. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55*(1)*,* 3–9. https://doi.org/10.1037/0022-006x.55.1.3

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 disorder. *Journal of Child Psychology and Psychiatry, 48*(8)*,* 803–812. https://doi.org/10.1111/j.1469-7610.2007.01756.x

Magiati, I., Tay, X. W., & Howlin, P. (2012). Early comprehensive behaviorally based interventions for children with autism spectrum disorders: A summary of findings from recent reviews and meta-analyses. *Neuropsychiatry, 2(*6), 543-570. https://doi.org/10.2217/NPY.12.59

Makrygianni, M. K., Gena, A., Katoudi, S., & Galanis, P. (2018). The effectiveness of applied behavior analytic interventions for children withy autism spectrum disorder: A meta-analytic study. *Research in Autism Spectrum Disorders, 51,* 18-31. https://doi.org/10.1016/j.rasd.2018.03.006

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

McKenzie, J. E., & Brennan, S. E. (2019). Synthesizing and presenting findings using other methods. 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. 321-348). John Wiley & Sons.

National Research Council. (2001). *Educating children with autism.* National Academies Press.

Odom, S. L., Boyd, B. A., Hall, L. J., & Hume, K. (2010). Evaluation of comprehensive treatment models for individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 40*(4), 425-436. https://doi.org/10.1007/s10803-009-0825-1

Odom, S. L., Boyd, B. A., Hall, L. J., & Hume, K. A. (2014). Comprehensive treatment models for children and youth with autism spectrum disorders. In F. Volkmar, S. Rogers, R. Paul, & K. Pelphrey (Eds.), Handbook of autism and pervasive developmental disorders (4th ed., Vol. 2, pp. 770-787). Hoboken, NJ: Wiley.

Ogilivie, D., Fayter, D., Petticrew, M., Sowden, A., Thomas, S., Whitehead, M., & Worthy, G. (2008). The harvest plot: a method for synthesizing evidence about the differential effects of interventions. *BMC Medical Research Methodology, 25,* 8. https://doi.org/10.1186/1471-2288-8-8

*Peters-Scheffer, N., Didden, R., Korzilius, H., & Sturmey, P. (2011). A meta-analytic study on the effectiveness of comprehensive ABA-based early intervention programs for children with autism spectrum disorders. *Research in Autism Spectrum Disorders, 5*(1)*,* 60-69. https://doi.org/10.1016/j.rasd.2010.03.011

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, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions,* version 6.4. www.training.cochrane.org/handbook

Reed, P., Osborne, L. A., & Corness, M. (2007). The real-world effectiveness of early teaching interventions for children with autism spectrum disorder. *Exceptional Children, 73*(4), 417–433. https://doi.org/10.1177/001440290707300402

Reichow, B. (2012). Overview of meta-analyses on early intensive behavioral interventions for children with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 42*(4), 512-520. https://doi.org/10.1007/s10803-011-1218-9

Reichow, B., Barton, E. E., Boyd, B. A., & Hume, K. (2011). Early intensive behavioral intervention for increasing functional behaviors and skills for young children with autism spectrum disorders (ASD). *Cochrane Database of Systematic Reviews, 2011*(8), CD009260. https://doi.org/10.1002/14651858.CD009260.pub1

Reichow, B., Barton, E. E., Boyd, B. A., & Hume, K. (2012). Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD): A systematic review. *Cochrane Database of Systematic Reviews, 2012*(10), CD009260. https://doi.org/10.1002/14651858.CD009260.pub2

Reichow, B., Barton, E. E., Boyd, B. A., & Hume, K. (2014). Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD): A systematic review. *Campbell Systematic Reviews, 10*(1), 9. https://doi.org/10.4073/csr.2014.9

*Reichow, B., Barton, E. E., Boyd, B. A., & Hume, K. (2018). Early intensive behavioral intervention (EIBI) for young children with autism spectrum disorders (ASD). *Cochrane Database of Systematic Reviews 2018*(5), CD009260. https://doi.org/10.1002/14651858.CD009260.pub3

Reichow, B., & Wolery, M. (2009). Comprehensive synthesis of early intensive behavioral interventions for young children with autism based on the UCLA Young Autism Project model. *Journal of Autism and Developmental Disorders, 39*(1)*,* 23-41. https://doi.org/10.1007/s10803-008-0596-0

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. https://doi.org/10.1007/s10803-007-0517-7

Remington, B., Hastings, R. P., Kovshoff, H., Degli Espinosa, F., Jahr, W., Brown, T., Alsford, A., Lemaic, M., Ward, N., & MacLean, M. E. (2007). A field effectiveness study of early intensive behavioral intervention: Outcomes for children with autism and their parents after two years. *American Journal on Mental Retardation, 112*(6)*,* 418–438. https://doi.org/10.1352/0895-8017(2007)112[418:EIBIOF]2.0.CO;2

Rodgers, M., Marshall, D., Simmonds, M., Couteur, A. L., Biswas, M., Wright, K., Rai, D., Palmer, S., Stewart, L., & Hodgson, R. (2020). Interventions based on early intensive applied behaviour analysis for autistic children: A systematic review and cost-effectiveness analysis. *Health Technology Assessment, 24*, 35. https://doi.org/10.3310/hta24350

*Rodgers, M., Simmonds, M., Marshall, D., Hodgson, R., Stewart, L. A., Rai, D., Wright, K., Ben-Itzchak, E., Eikeseth, S., Eldevik, S., Kovshoff, H., Magiati, I., Osborne, L. A., Reed, P., Vivanti, G., Zachor, D., & Couteur, A. L. (2021). Intensive behavioural interventions based on applied behaviour analysis for young children with autism: An international

collaborative individual participant data meta-analysis. *Autism, 25*(4), 1137-1153. https://doi.org/10.1177/1362361320985680

Rogers, S. J., Estes, A., Lord, C., Munson, J., Rocha, M., Winter, J., Greenson, J., Colombi, X., Dawson, G., Vismara, L. A., Sugar, C. A., Hellemann, 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), 853–65. https://doi.org/10.1016/j.jaac.2019.01.004

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*(6), 417–438. https://doi.org/10.1352/0895-8017(2005)110[417:IBTFCW]2.0.CO;2

*Shi, B., Wu, W., Dai, M., Zeng, J., Luo, J., Cai, L., Wan, B., & Jing, J. (2021). Cognitive, language, and behavioral outcomes in children with autism spectrum disorders exposed to early comprehensive treatment models: A meta-analysis and meta-regression. *Frontiers in Psychiatry, 12,* 691148. https://doi.org/10.3389/fpsyt.2021.691148

Smith, T., Eikeseth, S., Klevstrand, M., & Lovaas, O. I. (1997). Intensive behavioral treatment for preschoolers with severed mental retardation and pervasive developmental disorder. *American Journal on Mental Retardation, 102*(3), 238–249. https://doi.org/10.1352/0895-8017(1997)102<0238:IBTFPW>2.0.CO;2

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*(4), 269–285. https://doi.org/10.1352/0895-8017(2000)105<0269:RTOIEI>2.0.CO;2

Spreckley, M., & Boyd, R. (2009). Efficacy of applied behavioral intervention in preschool children with autism for improving cognitive, language, and adaptive behavior: A systematic review and meta-analysis. *Journal of Pediatrics, 154,* 338-344. https://doi.org/10.1016/j.jpeds.2008.09.012

Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A.-W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., ... Higgins, J. P. T. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ, 355,* i4919. https://doi.org/10.1136/bmj.i4919

Veritas Health Innovation. (no date). *Covidence systematic review software* [Computer software].

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,* 1673–83. https://doi.org/10.1007/s10803-017-3414-8

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

Wergeland, G. J. H., Posserud, M-B., Fjermestad, K., Njardvik, U., & Ost, L-G. (2022). Early behavioral interventions for children and adolescents with autism spectrum disorder in routine clinical care: A systematic review and meta-analysis. *Clinical Psychology: Science and Practice, 29*(4), 400-414. https://doi.org/10.1037/cps0000106

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

Yu, Q., Li, E., Li, L., & Liang, W. (2020). Efficacy of interventions based on applied behavior analysis for autism spectrum disorder: A meta-analysis. *Psychiatry Investigation, 17*(5), 432-443. https://doi.org/10.30773/pi.2019.0229

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

Table 1. Characteristics of Included Systematic Reviews

Review	Studies	Study designs	Search	Additional search methods	Rigor assessment method
Eldevik 2009	*u* = 9	RCT (*u* = 1) QEGC (*u* = 8)	March 2008	• Reference list review	not reported
Peters-Scheffer 2011	*u* = 10	RCT (*u* = 1) QEGC (*u* = 9)	2009	• Hand search • Reference tracking	Quality Index (Downs & Black, 1998)
Reichow 2018	*u* = 5	RCT (*u* = 1) QEGC (*u* = 4)	August 2017	• Reference list review (included studies) • Contacted authors	Cochrane Risk of Bias (Higgins et al. 2011)
Rodgers 2020	*u* = 20[a]	RCT (*u* = 4) QEGC (*u* = 16)	June 2019	• Reference list review (included studies) • Contacted authors	Cochrane Risk of Bias 2.0 (Cochrane Methods Group, 2019) ROBINS-I (Sterne et al. 2016)
Shi 2021	*u* = 18[b]	RCT (*u* = 3) QEGC (*u* = 5) QEPP (*u* = 10)	June 2019	• Reference list review (included studies and seminal reviews)	Evaluative Method (Reichow et al. 2008)
Eckes 2023	*u* = 11	RCT (*u* = 1) QEGC (*u* = 10)	March 2020	• Google Scholar • Reference list review (extant meta-analyses)	Cochrane Risk of Bias (Higgins et al., 2011)

Note: [a] = 10 studies included in meta-analytic syntheses (2 RCT and 8 QEGC); b = 6 studies included in meta-analytic syntheses of group design studies (2 RCT and 4 QEGC); *u* = number of primary studies; RCT = randomized controlled trial; QEGC = quasi-experimental two group comparison; QEPP = quasi-experimental (one group) pre/post; ROBINS-I = Risk of Bias in Non-randomized Studies - of Interventions

Table 2. Pre-treatment Child Participant Characteristics for Studies Included in Meta-Analytic Syntheses

Review	Number of children	Gender M:F	Age	IQ
Eldevik 2009	*N* = 297 (*u* = 9)	not reported	M range = 32.1 to 65.6 months (*u* = 9)	M range = 27.6 to 63.6 (*u* = 9)
Peters-Scheffer 2011	*N* = 344 (*u* = 10)	197:33 (*u* = 7)	M range = 33.6 to 65.6 months (*u* = 10)	M range = 27.6 to 76.5 (*u* = 10)
Reichow 2018	*N* = 219 (*u* = 5)	154:24 (*u* = 4)	M range = 30.2 to 42.5 months (*u* = 5)	M range = 50.6 to 76.5 (*u* = 5)
Rodgers 2020	*N* = 513 (*u* = 10)[a]	359:72 (*u* = 8)	M range = 20.6 to 65.6 months (*u* = 9)	M range = 51.6 to 76.5 (*u* = 9)
Shi 2021	*N* = 236 (*u* = 6)[b]	157:34 (*u* = 5)	M range = 33.6 to 37.8 months (*u* = 5)	M range = 50.6 to 61.9 (*u* = 5)
Eckes 2023	*N* = 632 (*u* = 11)	348:55 (*u* = 9)	M range = 25.6 to 65.6 months (*u* = 11)	not reported

Key: [a] = 10 studies included in meta-analyses of comparative design studies (i.e., early intensive ABA-based interventions including studies not providing individual patient data); [b] = 6 studies included in meta-analyses of comparative design studies

Table 3. Intervention Characteristics for Studies Included in Meta-Analytic Syntheses

Review	Intervention Conditions	Comparison Conditions	Intervention Density (mean range in hours/week)	Intervention Duration (mean range in months)
Eldevik 2009	EIBI (*u* = 9)	not reported	13.0 to 40.0	12 to 24-36
Peters-Scheffer 2011	EIBI (*u* = 11)	TAU (*u* = 7) Eclectic (*u* = 3) Low-intensity ABA (*u* = 1)	12.5 to 38.6	10 to 24+
Reichow 2018	EIBI (*u* = 5)	TAU (*u* = 4) Other (*u* = 1)	24.0+	24 to 36
Rodgers 2020	EIBI (*u* = 8) ESDM (*u* = 2)	TAU (*u* = 3) Eclectic (*u* = 5) Other (*u* = 2)	13.6 to 40.0	14 to 36
Shi 2021	EIBI (*u* = 6)	TAU (*u* = 2) Eclectic (*u* = 1) Low-intensity ABA (*u* = 2) Other (*u* = 1)	24.5 to 40.0	24 to 60
Eckes 2023	EIBI (*u* = 11)	TAU (*u* = 8) WLC (*u* = 1) Eclectic (*u* = 1) Low-intensity ABA (*u* = 1)	13.6 to 32.4	6 to 25

Key: [a] = 10 studies included in meta-analyses of comparative design studies (i.e., early intensive ABA-based interventions including studies not providing individual patient data); [b] = 6 studies included in meta-analyses of comparative design studies; *u* = number of primary studies; EIBI = early intensive behavioral intervention; TAU = Treatment as Usual; ABA = Applied Behavior Analytic; ESDM = Early Start Denver Model; WLC = Waitlist Control

Table 4. Effects of Comprehensive Applied Behavior Analytic Interventions on IQ

Review	Effect Size Estimate	Heterogeneity	Publication Bias
Eldevik 2009	SMD = 1.10 (95% CI 0.87, 1.34; *k* = 10[a])	*Q*(9) = 10.07, *p* = 0.35; I^2 = 10.66%	Visual inspection and Trim and Fill – not detected
Peters-Scheffer 2011	MD = 11.98 (95% CI 6.73, 17.23; *k* = 10)	Tau^2 = 56.03	not reported
Reichow 2018	MD = 15.44 (95% CI 9.29, 21.59; *k* = 5)	*Q*(4) = 1.16, *p* = 0.88; I^2 =0%	not evaluated[b]
Rodgers 2020	MD = 14.10 (95% CI 8.92, 19.29; *k* = 8)	I^2 = 0%; Tau^2 = 0.0	not reported
Shi 2021	SMD = 0.53 (95% CI 0.16 to 0.90; *k* = 6)	I^2 = 41.3%	Egger's Test – not detected (Egger's Test, *p* = -.90)
Eckes 2023	SMD = 0.51 (95% CI 0.09, 0.92; *k* = 8)	*Q*(7) = 17.87, *p* = 0.013; I^2 = 63.59%; Tau^2 = 0.22	Visual inspection – not detected

Notes: [a] = 10 pairwise comparisons from 9 studies; [b] = authors reported small number of studies did not allow inspection of publication bias; SMD = standardized mean difference effect size; 95% CI = 95th percentile confidence interval; *k* = number of comparisons included in meta-analysis; MD = mean difference effect size

Table 5. Effects of Comprehensive Applied Behavior Analytic Interventions on Adaptive Behavior

Review	Effect Size Estimate	Heterogeneity	Publication Bias
Eldevik 2009	SMD = 0.66 (95% CI 0.41, 0.90; *k* = 8[a])	$Q(7) = 8.50$, $p = 0.29$; $I^2 = 17.65\%$	Visual inspection and Trim and Fill – not detected
Peters-Scheffer 2011	MD = 5.92 (95% CI 2.72, 9.13; *k* = 7)	$Tau^2 = 0.14$	not reported
Reichow 2018	MD = 9.58 (95% CI 5.57, 13.60; *k* = 5)	$Q(4) = 2.43$, $p = 0.66$; $I^2 = 0.0\%$; $Tau^2 = 0.0$	not evaluated[b]
Rodgers 2020	MD = 9.38 (95% CI 4.41, 14.36; *k* = 10)	$I^2 = 72\%$; $Tau^2 = 43.94$	not reported
Shi 2021	SMD = 0.47 (95% CI 0.11, 0.83; *k* = 5)	$I^2 = 24.2\%$	Egger's Test – not detected ($p = 0.74$)
Eckes 2023	SMD = 0.37 (95% CI 0.03, 0.70; *k* = 28[c])	$Q(27) = 106.18$, $p < 0.001$; $I^2 = 72.95\%$; $Tau^2 = 0.24$	Visual inspection and Egger's Test – Detected ($z = 3.36$, $p < 0.001$)

Notes: [a] = 8 pairwise comparisons from 7 studies (Howard et al. treatment group compared separately for control group and comparison group); [b] = authors reported small number of studies did not allow inspection of publication bias; [c] = 28 pairwise comparisons from 9 studies; SMD = standardized mean difference effect size; 95% CI = 95th percentile confidence interval; *k* = number of comparisons included in meta-analysis; MD = mean difference effect size

Figure 1. Flow diagram of review selection

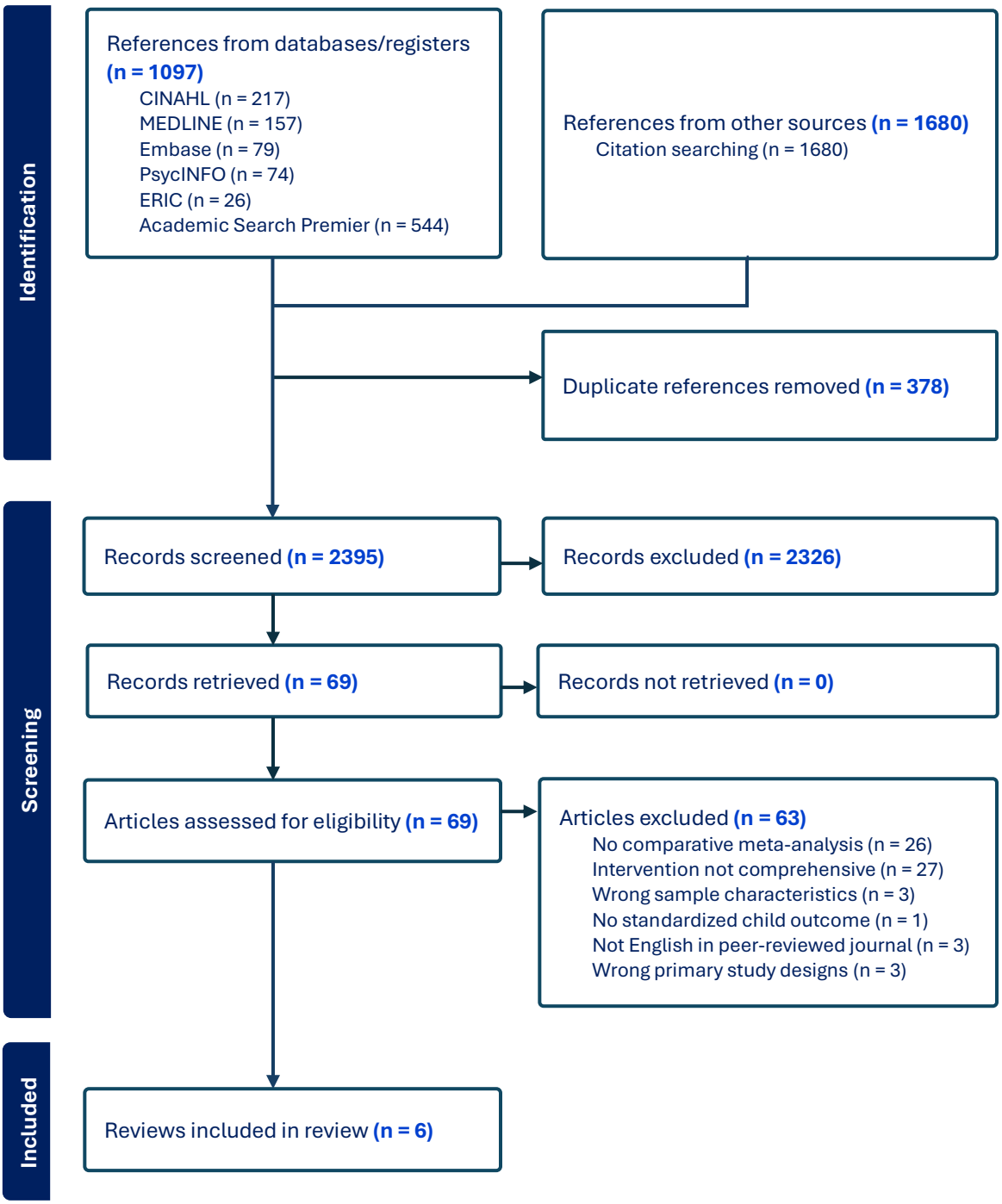
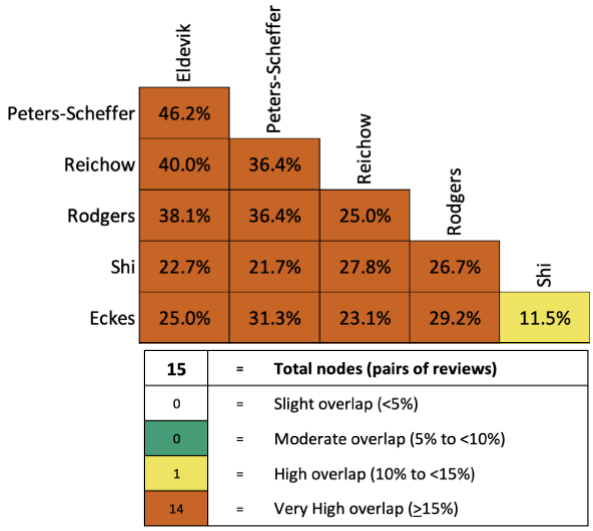


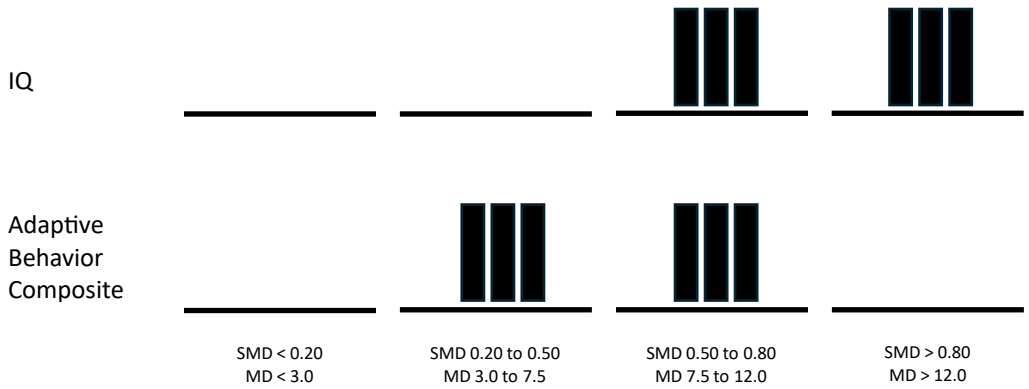
Figure 2: Visual Display of Primary Study Overlap

Overall results		
Number of columns (number of reviews)	c	6
Number of rows (number of index publications)	r	36
Number of included primary studies (including double counting)	N	73
Covered area	N/(rc)	33.80%
Corrected covered area	(N-r)/(rc-r)	20.56%
Interpretation of overlap	Very High overlap	
Structural Zeros	X	0
Corrected covered area (adjusting by structural zeros)	(N-r)/(rc-r-X)	20.56%
N° of non-overlapped primary studies	In 1 SR	20
	In 2 SRs	6
	In 3 SRs	2
	In 4 SRs	3
	In 5 SRs	2
	In 6 SRs	2



- Visual display created using Graphical Representation of Overlap for OVerviews (GROOVE; Bracchiglione et al., 2022)

Figure 3. Harvest Plot Depiction of Magnitude of Effect for Primary Outcomes



Umbrella Review of Meta-Analyses of Comprehensive Applied Behavior Analytic Interventions for Young Children with Autism Spectrum Disorder

SUPPLEMENTAL MATERIALS

Supplemental Text 1. Search terms for Academic Search Premier (searched July 12, 2024 using EBSCO)

1. TI “systematic review”
2. TI “systematic literature”
3. TI “meta analysis”
4. TI “meta analyses”
5. TI “metaanalyses”
6. TI “metaanalysis”
7. TI “meta synthesis”
8. TI “meta syntheses”
9. TI “comprehensive literature”
10. TI “comprehensive review”
11. AB “systematic review”
12. AB “systematic literature”
13. AB “meta analysis”
14. AB “meta analyses”
15. AB “metaanalyses”
16. AB “metaanalysis”
17. AB “comprehensive literature”
18. AB “comprehensive review”
19. TX “intensive behavioral intervention”
20. TX “intensive behavioural intervention”
21. TX “early intensive behavioral intervention”
22. TX “early intensive behavioural intervention”
23. TX IBI
24. TX EIBI
25. TX (intensive N3 (intervention or therapy or treatment or program))
26. TX (comprehensive behavioral N3 (intervention or therapy or treatment or program))
27. TX (comprehensive behavioural N3 (intervention or therapy or treatment or program))
28. TX comprehensive ABA
29. TX (comprehensive N3 (intervention or therapy or treatment or program))
30. TX autis*
31. TX ASD
32. TX Asperger*
33. TX PDD*
34. TX pervasive developmental disorder
35. DE “AUTISM” OR DE “ASPERGER’S syndrome” OR DE “AUTISM in adolescence” OR DE “AUTISM in adults” OR DE “AUTISM in children”
36. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18
37. 19 OR 20 OR 21 OR 23 OR 24 OR 25 OR 26 OR 27 OR 28 OR 29
38. 30 OR 31 OR 32 OR 33 OR 34 OR 35
39. 36 AND 37 AND 38
40. Limiters – Publication Years: 20090101-20241231

Supplemental Text 2. Search terms for APA PsycINFO (searched July 12, 2024 using EBSCO)

1. TI “systematic review”
2. TI “systematic literature”
3. TI “meta analysis”
4. TI “meta analyses”
5. TI “metaanalyses”
6. TI “metaanalysis”
7. TI “meta synthesis”
8. TI “meta syntheses”
9. TI “comprehensive literature”
10. TI “comprehensive review”
11. AB “systematic review”
12. AB “systematic literature”
13. AB “meta analysis”
14. AB “meta analyses”
15. AB “metaanalyses”
16. AB “metaanalysis”
17. AB “comprehensive literature”
18. AB “comprehensive review”
19. TX “intensive behavioral intervention”
20. TX “intensive behavioural intervention”
21. TX “early intensive behavioral intervention”
22. TX “early intensive behavioural intervention”
23. TX IBI
24. TX EIBI
25. TX (intensive N3 (intervention or therapy or treatment or program))
26. TX (comprehensive behavioral N3 (intervention or therapy or treatment or program))
27. TX (comprehensive behavioural N3 (intervention or therapy or treatment or program))
28. TX comprehensive ABA
29. TX (comprehensive N3 (intervention or therapy or treatment or program))
30. TX autis*
31. TX ASD
32. TX Asperger*
33. TX PDD*
34. TX pervasive developmental disorder
35. MM “Autism Spectrum Disorders” OR MM “Autistic Traits”
36. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18
37. 19 OR 20 OR 21 OR 23 OR 24 OR 25 OR 26 OR 27 OR 28 OR 29
38. 30 OR 31 OR 32 OR 33 OR 34 OR 35
39. 36 AND 37 AND 38
40. Limiters – Publication Years: 2009-2024

Supplemental Text 3. Search terms in CINAHL Plus with Full Text

1. TI “systematic review”
2. TI “systematic literature”
3. TI “meta analysis”
4. TI “meta analyses”
5. TI “metaanalyses”
6. TI “metaanalysis”
7. TI “meta synthesis”
8. TI “meta syntheses”
9. TI “comprehensive literature”
10. TI “comprehensive review”
11. AB “systematic review”
12. AB “systematic literature”
13. AB “meta analysis”
14. AB “meta analyses”
15. AB “metaanalyses”
16. AB “metaanalysis”
17. AB “comprehensive literature”
18. AB “comprehensive review”
19. TX “intensive behavioral intervention”
20. TX “intensive behavioural intervention”
21. TX “early intensive behavioral intervention”
22. TX “early intensive behavioural intervention”
23. TX IBI
24. TX EIBI
25. TX (intensive N3 (intervention or therapy or treatment or program))
26. TX (comprehensive behavioral N3 (intervention or therapy or treatment or program))
27. TX (comprehensive behavioural N3 (intervention or therapy or treatment or program))
28. TX comprehensive ABA
29. TX (comprehensive N3 (intervention or therapy or treatment or program))
30. TX autis*
31. TX ASD
32. TX Asperger*
33. TX PDD*
34. TX pervasive developmental disorder
35. (MM “Autistic Disorders”) OR (MM “Autism Spectrum Disorder”)
36. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18
37. 19 OR 20 OR 21 OR 23 OR 24 OR 25 OR 26 OR 27 OR 28 OR 29
38. 30 OR 31 OR 32 OR 33 OR 34 OR 35
39. 36 AND 37 AND 38
40. Limiters – Publication Years: 20090101-20241231

Supplemental Text 4. Search Strategy for Embase (searched July 12, 2024 using Elsevier)

1. 'systematic review':ti
2. 'systematic literature':ti
3. 'meta analysis':ti
4. 'meta analyses':ti
5. 'metaanalysis':ti
6. 'metaanalyses':ti
7. 'meta synthesis':ti
8. 'meta syntheses':ti
9. 'comprehensive review':ti
10. 'comprehensive literature':ti
11. 'systematic review':ab
12. 'systematic literature':ab
13. 'meta analysis':ab
14. 'meta analyses':ab
15. 'metaanalysis':ab
16. 'metaanalyses':ab
17. 'meta synthesis':ti
18. 'meta syntheses':ab
19. 'comprehensive review':ab
20. 'comprehensive literature':ab
21. 'intensive behavioral intervention'
22. 'intensive behavioural intervention'
23. 'early intensive behavioral intervention'
24. 'early intensive behavioural intervention'
25. ibi
26. eibi
27. intensive NEAR/3 (intervention OR therapy OR treatment OR program)
28. comprehensive AND (behavioral NEAR/3 (intervention OR therapy OR treatment OR program))
29. comprehensive AND (behavioural NEAR/3 (intervention OR therapy OR treatment OR program))
30. comprehensive NEAR/3 (intervention OR therapy OR treatment OR program)
31. comprehensive AND aba
32. pervasive developmental AND disorder
33. pdd
34. asperger*
35. autis*
36. asd
37. #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20
38. #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31
39. #32 OR #33 OR #34 OR #35 OR #36
40. #37 AND #38 AND #39
41. Limiters – Publication Year: 2009-2024

Supplemental Text 5. Search terms for Education Resources Information Center (ERIC; searched July 12, 2024 using EBSCO)

1. TI “systematic review”
2. TI “systematic literature”
3. TI “meta analysis”
4. TI “meta analyses”
5. TI “metaanalyses”
6. TI “metaanalysis”
7. TI “meta synthesis”
8. TI “meta syntheses”
9. TI “comprehensive literature”
10. TI “comprehensive review”
11. AB “systematic review”
12. AB “systematic literature”
13. AB “meta analysis”
14. AB “meta analyses”
15. AB “metaanalyses”
16. AB “metaanalysis”
17. AB “comprehensive literature”
18. AB “comprehensive review”
19. TX “intensive behavioral intervention”
20. TX “intensive behavioural intervention”
21. TX “early intensive behavioral intervention”
22. TX “early intensive behavioural intervention”
23. TX IBI
24. TX EIBI
25. TX (intensive N3 (intervention or therapy or treatment or program))
26. TX (comprehensive behavioral N3 (intervention or therapy or treatment or program))
27. TX (comprehensive behavioural N3 (intervention or therapy or treatment or program))
28. TX comprehensive ABA
29. TX (comprehensive N3 (intervention or therapy or treatment or program))
30. TX autis*
31. TX ASD
32. TX Asperger*
33. TX PDD*
34. TX pervasive developmental disorder
35. DE “Autism Spectrum Disorders”
36. S30 OR S31 OR S32 OR S33 OR S34 OR S35
37. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18
38. 19 OR 20 OR 21 OR 23 OR 24 OR 25 OR 26 OR 27 OR 28 OR 29
39. 36 AND 37 AND 38
40. Limiters – Publication Year: 2009-2024

Supplemental Text 6. Search terms for Medline (searched July 12, 2024 using EBSCO)

1. TI “systematic review”
2. TI “systematic literature”
3. TI “meta analysis”
4. TI “meta analyses”
5. TI “metaanalyses”
6. TI “metaanalysis”
7. TI “meta synthesis”
8. TI “meta syntheses”
9. TI “comprehensive literature”
10. TI “comprehensive review”
11. AB “systematic review”
12. AB “systematic literature”
13. AB “meta analysis”
14. AB “meta analyses”
15. AB “metaanalyses”
16. AB “metaanalysis”
17. AB “comprehensive literature”
18. AB “comprehensive review”
19. TX “intensive behavioral intervention”
20. TX “intensive behavioural intervention”
21. TX “early intensive behavioral intervention”
22. TX “early intensive behavioural intervention”
23. TX IBI
24. TX EIBI
25. TX (intensive N3 (intervention or therapy or treatment or program))
26. TX (comprehensive behavioral N3 (intervention or therapy or treatment or program))
27. TX (comprehensive behavioural N3 (intervention or therapy or treatment or program))
28. TX comprehensive ABA
29. TX (comprehensive N3 (intervention or therapy or treatment or program))
30. TX autis*
31. TX ASD
32. TX Asperger*
33. TX PDD*
34. TX pervasive developmental disorder
35. (MM “Autistic Disorders”) OR (MM “Autism Spectrum Disorder”)
36. 1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7 OR 8 OR 9 OR 10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16 OR 17 OR 18
37. 19 OR 20 OR 21 OR 23 OR 24 OR 25 OR 26 OR 27 OR 28 OR 29
38. 30 OR 31 OR 32 OR 33 OR 34 OR 35
39. 36 AND 37 AND 38
40. Limiters – Publication Years: 20090101-20241231

Supplemental Table 1. Differences between Protocol and Review (with rationale)

Area	Protocol (planned)	Review (actual)	Rationale
Review inclusion criteria	Systematic reviews with meta-analyses of group research design studies (i.e., randomized controlled trial, clinical controlled trial, two group pre-test/post-test comparison, pre-post design)	Systematic reviews with at least one meta-analytic synthesis of a child outcome from a comparative group design study (i.e., randomized controlled trial, clinical controlled trial, two group pre-test/post-test comparison)	This change was made to create a group of reviews with more comparable methods and resulted in the exclusion of one review (Reichow & Wolery, 2009) that met all inclusion criteria and would have otherwise been included in this overview.
Review selection	Use of "snowball methods" (Greenhalgh & Peacock, 2005) by examining the reference lists of included studies and by searching for studies citing the included studies.	We used the Citation Chaser online application to conduct the snowball searches by inputting the DOI of each review and using the Citation Chaser functions to download the references lists from each review (backward citation chasing) and references for citations to each review (forward citation chasing).	We did not alter the use of snowball methods, but we did choose to use the Citation Chaser app, which was not specified in the protocol.
Synthesis and analyses across reviews	Aggregate findings across reviews by creating tables and other graphic displays of data.	Aggregated findings across reviews by constructing tables and harvest plots for each outcome category	Visual data representation recommended for reviews without synthesis (McKenzie & Brennan, 2019)

Supplemental Table 2. Findings for Rodgers Individual Patient Data Meta-Analyses and Shi Pre/Post Within Group Change Effect Meta-Analyses

Review	Studies	Study designs	Sample size	Meta-analytic findings
Rodgers 2020	*u* = 10	QEGC (*u* = 10)	*N* = 491	IQ: MC = 11.97 (95% CI 6.74, 17.20; *u* = 6) AB: MC = 7.74 (95% CI 1.87, 13.61; *u* = 7)
Shi 2021	*u* = 18	RCT (*u* = 3) QEGC (*u* = 5) QEPP (*u* = 10)	*N* = 495	IQ: SMC = 0.85 (95% CI 0.47, 1.22; *u* = 15) AB: SMC = 0.15 (95% CI -0.28, 0.57; *u* = 11) GC: SMC = 0.75 (95% CI 0.47, 1.02; *u* = 8) EC: SMC = 1.12 (95% CI 0.70, 1.53; *u* = 5) RC: SMC = 1.11 (95% CI 0.82, 1.41; *u* = 5) SS: SMC = 0.55 (95% CI 0.17, 0.92; *u* = 7) DL: SMC = -0.05 (95% CI -0.49, 0.39; *u* = 7) AS: SMC = 0.68 (95% CI; 0.12, 1.24; *u* = 7)

Key: *u* = number of primary studies; QEGC = quasi-experimental group comparison; RCT = randomized controlled trial; QEPP = quasi-experimental pre/post design; *N* = number of child participants; MC = mean change; SMC = standardized mean change; CI = confidence interval; AB = adaptive behavior composite; GC = generalized communication/language skills; EC = expressive communication; RC = receptive communication; SS = social skills; DL = daily living skills; AS = autism symptomatology

Supplemental Table 3. Assessment Instruments Used Across Outcomes in Primary Studies as Reported by Included Systematic Reviews

Review	Outcome Category	Measures Used Across Primary Studies
Eldevik 2009	IQ	BSID, BSID-II, SB-4, WISC-III, WISC-R, WPPSI-R
	Adaptive Behavior Composite	VABS-ABC
Peters-Scheffer 2011	IQ	Specific measures not provided
	Adaptive Behavior Composite	VABS-ABC
	Generalized Communication Skills	VABS-C
	Social Skills	VABS-S
	Daily Living Skills	VABS-DLS
Reichow 2018	IQ	BSID-II; BSID-R; DAS; DAYC; DP-II; PEP-R; SB-4; WPPSI-R
	Adaptive Behavior Composite	VABS-ABC, DDST, DP-II, RIDES
	Generalized Communication Skills	VABS-C
	Expressive Communication	EOWPVT, EOWPVT-R, EVT, ITDA, ITLS, PLS-3, RDLS, REEL-R
	Receptive Communication	BPVS-II, ITDA, ITLS, PLS-3, PPVT-III, RDLS, REEL-R, ROWPVT
	Social Skills	VABS-S
	Daily Living Skills	VABS-DLS
	Autism Symptom Severity	ASQ, ADI-R
Rodgers 2020	IQ	BAS-II, BSID-II, BSID-III, BSID-IV, BSID-R, CIIS, GIDS, LIPS, MPS, MSEL, PEP-R, PPVT, SB-4, SB-5, WISC-III, WISC-IV, WISC-R, WPPSI, WPPSI-III, WPPSI-R
	Adaptive Behavior Composite	VABS-ABC
	Autism Symptom Severity	ADOS
Shi 2021	IQ	BSID, BSID-R, BSID-II, DAS, DP-II, LIPS-R, MPS, MSEL, WASI, WISC-II, WISC-III, WISC-R, WPPSI, WPPSI-R
	Adaptive Behavior Composite	VABS-ABC
	Generalized Communication Skills	VABS-C
	Expressive Communication	BPVS-2, RDLS, EOWPVT, SICD-R
	Receptive Communication	BPVS-2, RDLS, ROWPVT, SICD-R
	Social Skills	VABS-S
	Daily Living Skills	VABS-DLS
Eckes 2023	IQ	BSID, BSID-II, GMDS-ER, MSEL, SB-4, WPPSI-R, WISC-R, WISC-III
	Adaptive Behavior Composite	VABS-ABC
	Generalized Communication Skills	BPVS-II, CDI, EOWPVT-R, MSEL, RDLS, RDLS-3
	Autism Symptom Severity	ADOS, ADI-R

Key: -- = not reported in systematic review; ADI-R = Autism Diagnostic Interview – Revised; ADOS = Autism Diagnostic Observation Schedule; ASQ = Autism Spectrum Quotient; BAS-II = British Ability Scales, 2nd ed.; BPVS-II = British Picture Vocabulary Scale, 2nd ed.; BSID = Baily Scales of Infant Development; BSID-II = Bayley Scales of Infant Development, 2nd ed.; BSID-III = Bayley Scales of Infant Development, 3rd ed.; BSID-IV = Bayley Scales of Infant Development, 4th ed.; BSID-R = Bayley Scales of Infant Development – Revised; CDI – MacArthur-Bates Communication Development Inventories; CIIS = Cattell Infant Intelligence Scale; DAS = Differential Abilities Scale; DAYC = Developmental Assessment of Young Children; DDST = Denver Developmental Screening Test; DP-II = Developmental Profile – 2nd ed; EOWPVT = Expressive One-Word Picture

Vocabulary Test; EOWPVT = Expressive One Word Picture Vocabulary Test – Revised; EVT = Expressive Vocabulary Test; GIDS = Gesell Infant Development Scale; GMDS-ER = Griffith Mental Development Scales – Extended Revised; ITDA = Infant Toddler Developmental Assessment; ITLS = Infant-Toddler Language Scale; LIPS = Leiter International Performance Scale; LIPS-R = Leiter-International Performance Scale - Revised; MPS = Merrill Palmer Scale of Mental Tests; MESL = Mullen Scales of Early Learning; PEP-R = Psychoeducational Profile – Revised; PLS-3 = Preschool Language Scale – 3rd ed.; PPVT = Peabody Picture Vocabulary Test; PPVT-III = Peabody Picture Vocabulary Test, 3rd ed.; REEL-R = Receptive Expressive Emergent Language Scales – Revised; RDLS = Reynell Developmental Language Scales; RDLS-3 = Reynell Developmental Language Scales, 3rd ed.; RIDES = Rockford Infant Developmental Evaluation Scale – Revised; ROWPVT = Receptive One Word Picture Vocabulary Test; SICD-R = Sequenced Inventory of Communication Development – Revised; SB-4 = Stanford-Binet Intelligence Scale, 4th ed.; SB-5 = Stanford-Binet Intelligence Scale, 5th ed.; VABS = Vineland Adaptive Behavior Scales; VABS-ABC = Vineland Adaptive Behavior Scales – Adaptive Behavior Composite; VABS-C = Vineland Adaptive Behavior Scales – Communication [subscale]; VABS-DLS = Vineland Adaptive Behavior Scales – Daily Living Skills [subscale]; VABS-S = Vineland Adaptive Behavior Scales – Socialization [subscale]; WASI = Wechsler Abbreviated Scale of Intelligence; WISC-II = Wechsler Intelligence Scale for Children, 2nd ed.; WISC-III = Wechsler Intelligence Scale for Children, 3rd ed.; WISC-IV = Wechsler Intelligence Scale for Children, 4th ed.; WISC-R = Wechsler Intelligence Scale for Children – Revised; WPPSI = Wechsler Preschool and Primary Scale of Intelligence; WPPSI-III = Wechsler Preschool and Primary Scale of Intelligence, 3rd ed.; WPPSI-R = Wechsler Preschool and Primary Scale of Intelligence - Revised

Supplemental Table 4. Characteristics and Findings of Primary Studies Included across Reviews

					Participant Characteristics by Group						Treatment Characteristics		
Study	Year	Review	Design	Sample	Group	*n*	age	M,F	IQ	VABS	Model	hr/wk	M of Tx
Lovaas	1987	2, 5	QEGC	59	TX	19	34.6	--	62.7	--	EIBI	40	24+
					C	19	40.9	--	57.0	--	LI-ABA	10	24+
					C	21	<42	--	60.0	--	TAU	--	24+
Birnbrauer	1993	2, 5	QEGC	14	TX	9	38.1	5,4	51.3	46.1	EIBI	18.7	21.6
					C	5	33.2	5,0	54.5	51.5	NTC	--	24
Smith	1997	2, 3	QEGC	21	TX	11	36	11,0	28	50.3	EIBI	30	35
					C	10	38	8,2	27	--	LI-ABA	10	26
Sheinkopf	1998	3	QEGC	22	TX	11	33.8	--	62.8	--	EIBI	27.0	15.7
					C	11	35.3	--	61.7	--	TAU	11.1	18
Weiss	1999	6	QEPP	20	TX	20	41.5	19,1	--	49.9	EIBI	40	24
Harris	2000	6	QEPP	27	TX	27	49.0	--	59.3	--	EIBI	35-40	93
Smith	2000	2, 3, 4, 5, 6	RCT	28	TX	15	36.1	12,3	50.5	63.4	EIBI	24.5	33.4
					C	13	35.8	11,2	50.7	65.2	LI-ABA	15-20	24
Bibby	2001	6	QEPP	66	TX	66	45.0	55,11	50.8	54.5	EIBI	30.3	32.8
Gabriels	2001	6	QEPP	17	TX	17	30.6	12,5	57.8	--	C-ABA	23	36
Eikeseth	2002	1, 2, 3, 5	QEGC	25	TX	13	66.3	8,5	61.9	55.8	EIBI	28.0	12.2
					C	12	65.0	11,1	65.2	60.0	TAU	29.1	13.6
Farrell	2005	5	QEGC	17	TX	8	<60	6,2	--	--	EIBI	30	26
					C	9	<60	7,2	--	--	TAU	--	20
Howard	2005	2, 3, 4, 5, 6	QEGC	62	TX	29	30.9	25,4	58.5	70.5	EIBI	25-40	14.2
					C	16	37.4	13,3	53.7	69.8	TAU	25-30	13.3
					C	16	34.6	16,0	59.9	71.6	TAU	15	14.8
Sallows	2005	3, 5, 6	RCT	23	TX	13	35.0	11,2	50.9	59.5	EIBI	37.6	48
					TX	10	37.1	8,2	52.1	60.9	EIBI	31.3	48
Cohen	2006	2, 4, 5, 6	QEGC	42	TX	21	30.2	18,3	61.6	69.8	EIBI	35-40	36
					C	21	33.2	17,4	59.4	70.6	TAU	--	36
Eldevik	2006	1, 5	QEGC	28	TX	13	53.0	10,3	41.0	52.5	EIBI	12.5	20.3
					C	15	49.0	14,1	47.2	52.5	TAU	12.0	21.4

Magiati	2007	3, 4, 5, 6	QEGC	44	TX	28	38.0	27,1	83.0	59.6	EIBI	32.4	25.5
					C	16	42.5	12,4	65.2	55.4	TAU	25.6	26.0
Reed	2007	3, 5	QEGC	48	TX	12	40	11,1	56.8	58.2	EIBI	30.4	9
					C	20	43	18,2	57.8	53.0	TAU	12.7	9
					C	16	38	--	53.4	58.6	Portage	8.5	9
Remington	2007	2, 3, 4, 5	QEGC	44	TX	23	38.4	--	61.4	--	EIBI	25.6	24
					C	21	35.7	--	62.3	--	TAU	15.3	24
Zachor	2007	5	QEGC	39	TX	20	27.7	19,1	--	--	EIBI	35	12
					C	19	28.8	18,1	--	--	TAU	4	12
Hayward	2009	5	QEPP	44	TX	44	35.1	34,10	53.8	63.7	EIBI	35.8	12
Akshoomoff	2010	6	QEGC	20	TX	20	28.9	18,2	--	--	C-ABA	31	7.7
Dawson	2010	5, 6	RCT	48	TX	24	23.9	--	61	69.5	ESDM	31.5	24
					C	24	23.1	18,6	59.4	69.9	TAU	18.4	24
Zachor	2010	5	QEGC	78	TX	45	25.1	--	72.2	66.2	EIBI	20	12
					C	33	26	--	73.3	68.6	TAU	19	12
Fava	2011	1	QEGC	22	TX	12	52	10,2	62.1	63.3	EIBI	14	6.4
					C	10	43.7	9,1	69.8	44.3	TAU	12	7.2
Fernell	2011	1	QEGC	192	TX	91	37.6	--	--	67.9	EIBI	15-40	25
					C	101	43.5	--	--	70.4	LI-ABA	--	25
Landa	2012	6	QEPP	48	TX	48	27.2	39,11	60.1	69.7	ABA	10+	6
Eikeseth	2012	1, 5	QEGC	43	TX	31	25.1	25,6	51.6	62.5	EIBI	13.6	25.1
					C	12	24.6	8,4	51.7	58.9	TAU	5	24.6
Stock	2013	5	QEGC	28	TX	14	46.7	12,2	--	--	EIBI	15	12
					C	14	46	12,2	--	--	GB-ABA	15	12
Vivanti	2014	5	QEGC	57	TX	27	40.3	23,4	53.5	68.7	ESDM	15-25	11.9
					C	30	42	27,3	49.0	68.5	TAU	15	11.8
Smith	2015	6	QEPP	64	TX	64	39.1	54,10	58.8	62.7	EIBI	16.7	12
Shawler	2016	1	QEGC	51	TX	32	28	28,4	70.2	--	EIBI	22	13.2
					C	19	27.2	16,3	62.6	--	TAU	1-8	13.1
Clark	2017	6	QEPP	48	TX	48	25.5	36,12	65.7	--	C-ABA	10+	24
Hagulund	2017	5	QEGC	94	TX	67	<72	--	--	--	C-ABA	15-25	24

					C	27	<72	--	--	--	TAU	--	24
Molnar	2017	[1]	QEGC	20	TX	13	56.4	13,0	--	--	EIBI	17.5	10.7
					C	7	56.4	7,0	--	--	WLC	--	6
Viven	2018	[6]	QEGC	59	TX	31	39.2	27,4	55.4	--	ESDM	16.7	22.4
					C	28	35.5	25,3	58.5	--	TAU	15-25	36
Perry	2019	[6]	QEPP	21	TX	21	40.9	19,2	--	--	EIBI	20-40	26.8

Key: age – average age by group in months; M – male; F – female; IQ – intelligence quotient; VABS – Vineland Adaptive Behavior Scales composite; hr/wk – average number of hours per week of treatment; M of Tx – average number of months of treatment; QEGC – quasi-experimental group comparison; TX – treatment group; C – control / comparison group; -- not reported; EIBI – early intensive behavioral intervention; LI-ABA – low intensity applied behavior analysis; TAU – treatment as usual (including eclectic models); NTC – no treatment control; QEPP – quasi-experimental pre/post; RCT – randomized controlled trial; C-ABA – community-based applied behavior analysis; ESDM – Early Start Denver Model; GB-ABA – group-based applied behavior analysis; WLC – waitlist control

Review: [1] – Eckes et al., 2023; [2] – Eldevik et al., 2009; [3] – Peters-Sheffer et al., 2011; [4] – Reichow et al., 2018; [5] – Rodgers et al., 2020; [6] – Shi et al., 2021

Supplemental Table 5. Study-level Risk of Bias in Systematic Review (ROBIS) Ratings

Review	Domain 1: Study eligibility criteria	Domain 2: Identification and selection of studies	Domain 3: Data collection and study appraisal	Domain 4: Synthesis and findings	Overall risk of bias rating
Eldevik 2009	Low	Low	High	High	High
Peters-Scheffer 2011	Low	Low	Low	High	Low
Reichow 2018	Low	Low	Low	Low	Low
Rodgers 2020	Low	Low	Low	Low	Low
Shi 2021	Low	Low	Low	Low	Low
Eckes 2023	Low	High	Unclear	Low	Low

Supplemental Table 6. Study-level JBI Appraisal Checklist for Systematic Reviews and Research Syntheses Ratings

Review	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9
Eldevik 2009	No	Yes	Unclear	Yes	No	N/A	Unclear	Yes	Yes
Peters-Scheffer 2011	Yes	Yes	No	Yes	Yes	Yes	Unclear	Yes	Yes
Reichow 2018	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Unclear
Rodgers 2020	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Shi 2021	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Eckes 2023	Yes	Yes	Yes	Yes	Yes	Unclear	Unclear	Yes	Yes

Supplemental Table 7. Effects of Comprehensive Applied Behavior Analytic Interventions on Generalized Communication Skills

Review	Effect Size Estimate	Heterogeneity	Publication Bias
Peters-Scheffer 2011	MD = 10.44 (95% CI 5.03, 15.85; *k* = 7)	Tau^2 = 35.98	not reported
Reichow 2018	MD = 11.22 (95% CI 5.39, 17.04; *k* = 5)	*Q*(4) = 1.86, *p* = 0.76; I^2 = 0%; Tau^2 = 0.0	not evaluated[a]
Shi 2021	SMD = 0.38 (95% CI 0.03, 0.73; *k* = 5)	I^2 = 21.0%	Egger's Test – not detected (*p* = 0.41)
Eckes 2023	SMD = 0.30 (95% CI -0.13, 0.72; *k* = 9[b])	*Q*(8) = 16.81, *p* = 0.03; I^2 = 61.94%; Tau^2 = 0.17	Visual inspection – not detected

Notes: [a] = authors reported small number of studies did not allow inspection of publication bias; [b] = 9 pairwise comparisons from 5 studies; SMD = standardized mean difference effect size; 95% CI = 95th percentile confidence interval; *k* = number of comparisons included in meta-analysis; MD = mean difference effect size

Supplemental Table 8. Effects of Comprehensive Applied Behavior Analytic Interventions on Expressive Language

Review	Effect Size Estimate	Heterogeneity	Publication Bias
Peters-Scheffer 2011	MD = 15.21 (95% CI 8.43, 21.99; *k* = 5)	Tau^2 = 27.23	not reported
Reichow 2018	SMD = 0.51 (95% CI 0.12, 0.90; *k* = 4)	*Q*(3) = 4.46, *p* = 0.22; I^2 = 32.77%; Tau^2 = 0.05	not evaluated[a]
Shi 2021	SMD = 0.46 (95% CI -0.08, 1.00; *k* = 4)	I^2 = 56.1%	Egger's Test – not detected (*p* = 0.54)

Notes: [a] = authors reported small number of studies did not allow inspection of publication bias; SMD = standardized mean difference effect size; 95% CI = 95th percentile confidence interval; *k* = number of comparisons included in meta-analysis; MD = mean difference effect size

Supplemental Table 9. Effects of Comprehensive Applied Behavior Analytic Interventions on Receptive Language

Review	Effect Size Estimate	Heterogeneity	Publication Bias
Peters-Scheffer 2011	MD = 13.94 (95% CI 10.14, 17.75; *k* = 5)	Tau^2 = 0.75	not reported
Reichow 2018	SMD = 0.55 (95% CI 0.23, 0.87; *k* = 4)	*Q*(3) = 1.52, *p* = 0.368; I^2 = 0.0%; Tau^2 = 0.0	not evaluated[a]
Shi 2021	SMD = 0.42 (95% CI -0.06, 0.91; *k* = 4)	I^2 = 45.5%	Egger's Test – not detected (*p* = 0.45)

Notes: [a] = authors reported small number of studies did not allow inspection of publication bias; SMD = standardized mean difference effect size; 95% CI = 95th percentile confidence interval; *k* = number of comparisons included in meta-analysis; MD = mean difference effect size

Supplemental Table 10. Effects of Comprehensive Applied Behavior Analytic Interventions on Social Skills

Review	Effect Size Estimate	Heterogeneity	Publication Bias
Peters-Scheffer 2011	MD = 4.96 (95% CI 0.18, 9.75; *k* = 10)	Tau^2 = 0.75	not reported
Reichow 2018	MD =6.56 (95% CI 1.52, 11.61; *k* = 5)	*Q*(4) = 5.25, *p* = 0.26; I^2 = 23.9%; Tau^2 = 7.94	not evaluated[a]
Shi 2021	SMD = 0.38 (95% CI 0.03, 0.73; *k* = 5)	I^2 = 20.8%	Egger's Test – not detected (*p* = 0.89)

Notes: [a] = authors reported small number of studies did not allow inspection of publication bias; SMD = standardized mean difference effect size; 95% CI = 95th percentile confidence interval; *k* = number of comparisons included in meta-analysis; MD = mean difference effect size

Supplemental Table 11. Effects of Comprehensive Applied Behavior Analytic Interventions on Daily Living Skills

Review	Effect Size Estimate	Heterogeneity	Publication Bias
Peters-Scheffer 2011	MD = 5.49 (95% CI 3.36, 7.61; *k* = 10)	Tau^2 = 1.30	not reported
Reichow 2018	MD = 7.77 (95% CI 3.75, 11.79; *k* = 5)	*Q*(4) = 1.73, *p* = 0.79; I^2 = 0%; Tau^2 = 0.0	not evaluated[a]
Shi 2021	SMD = 0.183 (95% CI -0.16, 0.53; *k* = 4)	I^2 = 0%	Egger's Test – not detected (*p* = 0.45)

Notes: [a] = authors reported small number of studies did not allow inspection of publication bias; SMD = standardized mean difference effect size; 95% CI = 95th percentile confidence interval; *k* = number of comparisons included in meta-analysis; MD = mean difference effect size

Supplemental Table 12. Effects of Comprehensive Applied Behavior Analytic Interventions on Symptom Severity

Review	Effect Size Estimate	Heterogeneity	Publication Bias
Reichow 2018	SMD = -0.34 (95% CI -0.79, 0.11; *k* = 2)	$Q(1) = 0.23$, $p = 0.63$; $I^2 = 0\%$; $Tau^2 = 0.0$	not evaluated[a]
Rodgers 2020	MD = 0.27 (95% CI -0.19, 0.73; *k* = 2)	$I^2 = 0\%$; $Tau^2 = 0.0$	not reported
Eckes 2023	SMD = -0.26 (95% CI -0.60, 0.07; *k* = 4[b])	$Q(3) = 2.39$, $p = 0.50$; $I^2 = 0\%$; $Tau^2 = 0.0$	Visual inspection – not detected[c]

Notes: [a] = authors reported small number of studies did not allow inspection of publication bias; [b] = 4 pairwise comparisons from 3 studies; [c] = authors reported low number of studies limited a proper inspection of funnel plot; SMD = standardized mean difference effect size; 95% CI = 95th percentile confidence interval; *k* = number of comparisons included in meta-analysis; MD = mean difference effect size

Supplemental Table 13. Summary Characteristics of Meta-Analytic Syntheses Early Intervention Programs for Young Children with Autism Spectrum Disorder using Pre/Post Within Group Change Effect Size Analysis

Review	Studies	Study designs	Sample size for treatment group	Intervention type(s)	Intervention density and duration	Meta-analytic findings
Reichow 2009	*u* = 13	RCT (*u* = 2) QEGC (*u* = 9) QEPP (*u* = 2)	*N* = 251	EIBI (*u* = 13)	18.7 to 40.0 hours per week for 12 to 48 months	IQ: SMC = 0.69 (95% CI n/r; $p < 0.001$; *u* = 11)
Virues-Ortega 2010	*u* = 21	RCT (*u* = 2) QEGC (*u* = 12) QEPP (*u* = 7)	*N* = 381	EIBI (*u* = 13) ABA (*u* = 9)	5.9 to 45.0 hours per week for 2.8 to 94.7 months	IQ: SMC = 1.19 (95% CI 0.91, 1.47; *u* = 20) AB: SMC = 1.09 (95% CI 0.70, 1.47; *u* = 15)
Makrygiani 2018	*u* = 29	RCT (*u* = 2) QEGC (*u* = 20) QEPP (*u* = 7)	*N* = 831	ABA (*u* = 29)	5.0 to 40.0 hours per week for 8 to 93 months	IQ: SMC = 0.74 (95% CI 0.55, 0.93; *u* = 21) AB: SMC = 0.42 (95% CI 0.18, 0.66; *u* = 17)
Wergeland 2022	*u* = 29	RCT (*u* = 4) QE (*u* = 25)[a]	*N* = 1,422	EIBI (*u* = 18) ABA (*u* = 6) IBT (*u* = 2) Other (*u* = 3)	7.3 to 37.6 hours per week for 3 to 48 months	IQ: SMC = 0.76 (95% CI 0.57, 0.96; *u* = 24) AB: SMC = 0.81 (95% CI 0.50, 1.12; *u* = 21)

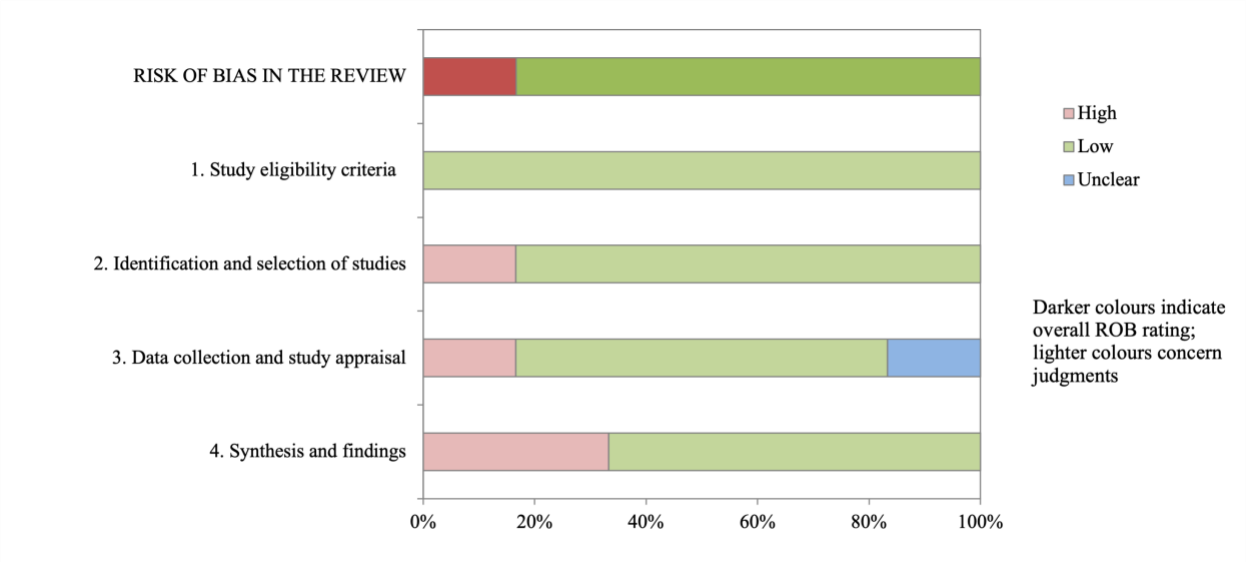
Key: [a] = non-randomized trials in Wegeland et al. review were classified as open trials and included quasi-experimental design studies with and without comparison groups; RCT = randomized controlled trial; QEGC = quasi-experimental group comparison; QEPP = quasi-experimental pre/post; EIBI = early intensive behavioral intervention; SMC = standardized mean change effect size; ABA = applied behavior analytic; AB = adaptive behavior composite; QE = quasi-experimental; IBT = intensive behavioral intervention

Supplemental Table 14. Characteristics and Findings of Other Key Excluded Meta-Analyses of Early Intervention Programs for Young Children with Autism Spectrum Disorder

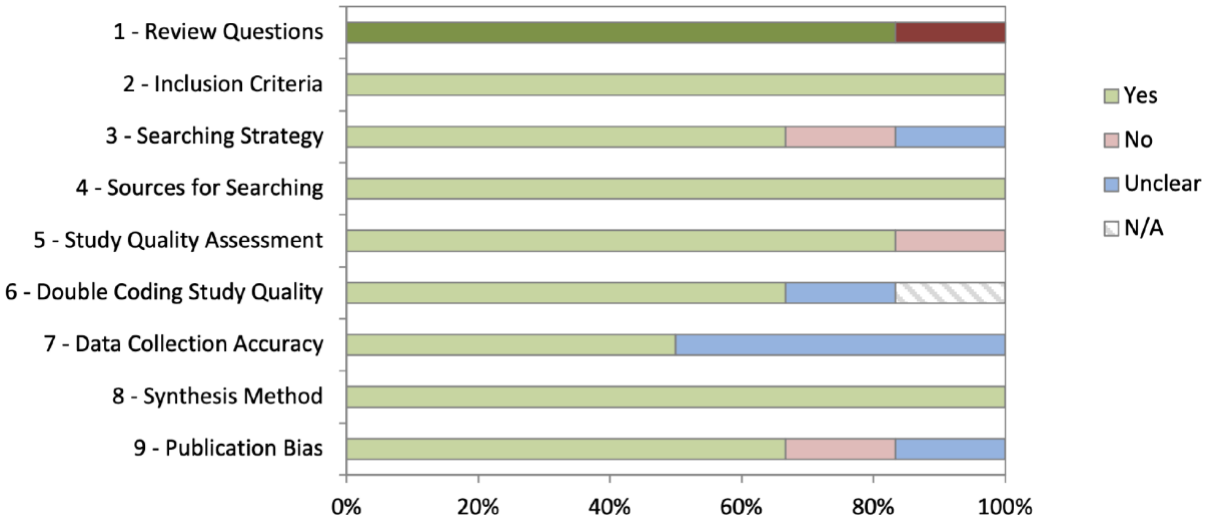
Review	Studies	Study designs	Sample Size (Participants)	Intervention Type(s)	Intervention Density	Comparators reported in review	Effect size	Findings for Meta-analysis of Primary Outcomes
Reichow 2009	*u* = 13	RCT (*u* = 2) QEGC (*u* = 9) QEPP (*u* = 2)	*N* = 373 T = 251 C = 122	EIBI based on UCLA YAP	19 to 40 HPW 12 to 48 M	Low intensity ABA, TAU, eclectic	SMC	**IQ**: SMC = 0.69 (*p* < 0.001; *u* = 11)
Spreckley 2009	*u* = 4	RCT (*u* = 3) QEGC (*u* = 1)	*N* = 101	ABA based on UCLA YAP	18 to 39 HPW 12 to 48 M	Parent training, parent directed EIBI, eclectic	SMD	**IQ:** SMD = 0.38 (95% CI -0.09, 0.84; *p* = 0.10) **AB**: SMD = 0.30 (95% CI -0.16, 0.77; *p* = 0.20)
Makrygiani 2010	*u* = 14	RCT (*u* = 2) QEGC (*u* = 8) QEPP (*u* = 4)	*N* = 387 T = 257 C = 130	Behavioual early interventions	10 to 40 HPW 8 to 58 M	Eclectic	SMC	**IQ:** SMC = 0.95 (SE = 0.13; *u* =5; HQ); SMC = 0.09 (SE = 0.08; *u* = 11; LQ) **AB**: SMC = 0.42 (SE = 0.15; *u* = 4; HQ); SMC = 0.47 (SE = 0.11; *u* = 7; LQ)
Virues-Ortega 2010	*u* = 22	RCT (*u* = 5) QEGC (*u* = 10) QEPP (*u* = 7)	*N* = 836 T = 638 C = 198	ABA interventions	6 to 45 HPW 12 to 407 M	Eclectic, TAU, low intensity ABA	SMC	**IQ**: SMC = 1.19 (95% CI 0.91, 1.47; *p* < 0.001; *u* = 18) **AB**: SMC = 1.09, 95% CI 0.70, 1.47; *p* < 0.001; *u* = 15)
Makrygiani 2018	*u* = 29	RCT (*u* = 2) QEGC (*u* = 20) QEPP (*u* = 7)	*N* = not reported T = 831	ABA interventions	4 to 40 HPW 8 to 93 M	Eclectic, TAU	SMC	**IQ**: SMC = 0.74 (95% CI 0.55, 0.93; *u* = 21) **AB**: SMC = 0.42 (95% CI 0.18, 0.66; *u* = 17)
Yu 2020	*u* = 14	RCT (*u* = 14)	*N* = 555 T = 278 C = 277	ABA, DTT, ESDM, PECS	2 to 40 HPW 3 to 36 M	TAU	SMD	**AB DLS**: SMD = 0.31 (95% CI -0.22, 0.84; *p* = 0.26)
Wergeland 2022	*u* = 29	RCT (*u* = 4) QEGC or QEPP (*u* = 25)	*N* = not reported T = 1422	ABA, EIBI, JAML	7 to 38 HPW 3 to 48 M	Not reported	SMC	**IQ**: SMC = 0.76 (95% CI 0.57, 0.96; *u* = 24) **AB**: SMC = 0.81 (95% CI 0.50, 1.12; *u* = 21)

Key: u = number of studies; RCT = randomized controlled trial; QEGC = quasi-experimental group comparison; QEPP = quasi-experimental pre/post; *N* = total sample; T = treatment; C = control; EIBI = early intensive behavioral intervention; UCLA YAP – University of California at Los Angeles Young Autism Project; HPW = hours per week; M = months; TAU = treatment as usual; SMC = standardized mean change; CI = confidence interval; SMD = standardized mean difference; SE = standard error; AB = adaptive behavior; HQ = high quality; LQ = low quality; ABA = applied behavior analysis/analytic; DTT = discrete trial teaching; ESDM = Early Start Denver Model; PECS = picture exchange communication system; DLS = daily living skills; joint attention mediated learning

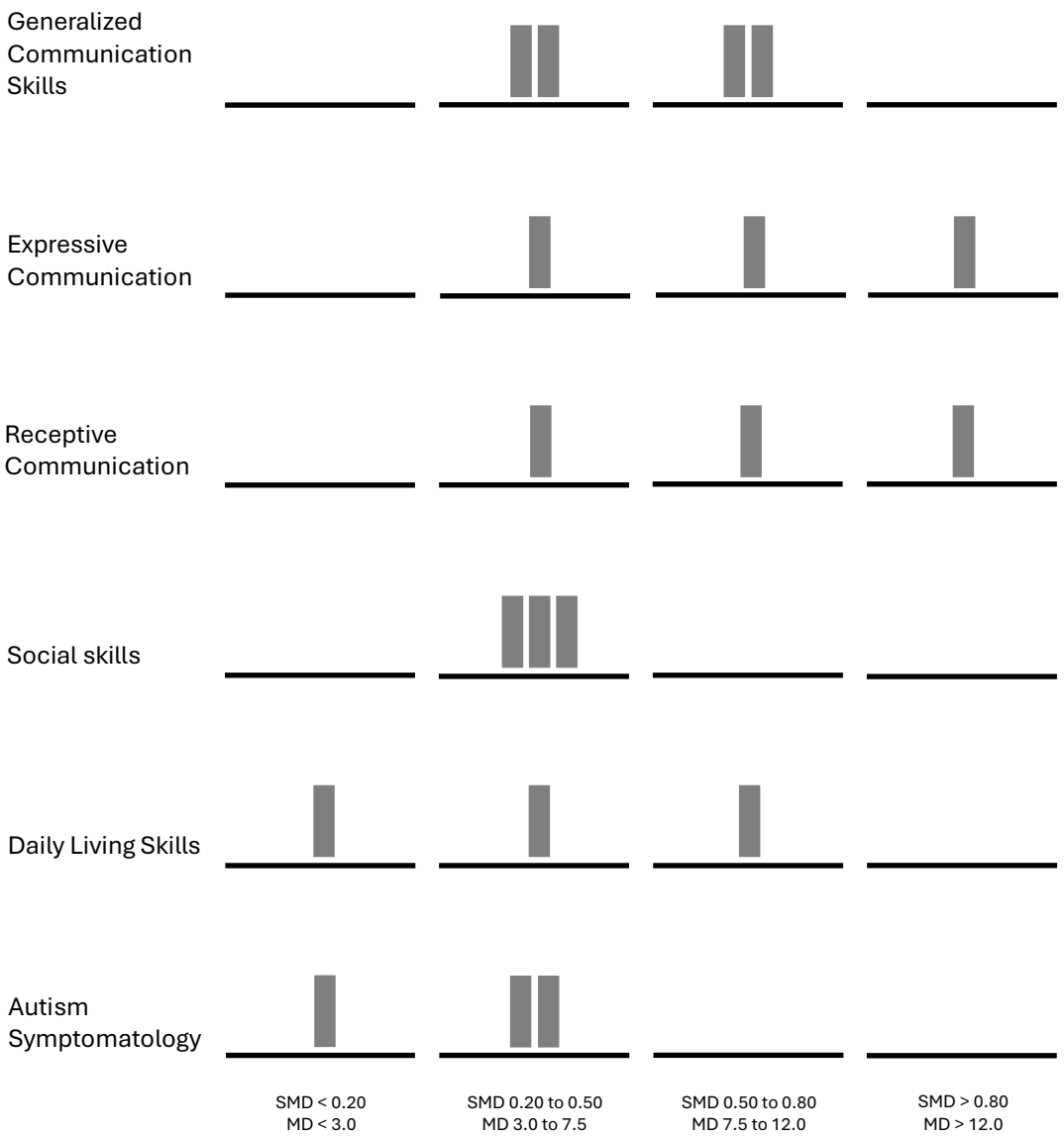
Supplemental Figure 1. Risk of Bias in Systematic Review (ROBIS) Ratings across Included Reviews



Supplemental Figure 2. JBI Appraisal Checklist for Systematic Reviews and Research Syntheses Ratings across Included Reviews



Supplemental Figure 3. Harvest Plot Depiction of Magnitude of Effect for Secondary Outcomes.



References for Supplemental Materials

Alpern, G., Boll, T., & Shearer, M. (1986). *Developmental profile II: Manual*. Western Psychological Services.

Akshoomoff, N., Stahmer, A. C., Corsello, C., & Mahrer, N. E. (2010). What happens next? Follow-up from the children's toddler school program. *Journal of Positive Behavioral Interventions, 12*(4), 245–53. https://doi.org/10.1177/1098300709343724

Bayley, N. (1969). *Bayley scales of infant development*. Psychological Corporation.

Bayley, N. (1993). *Bayley scales of infant development* (2nd ed.). Psychological Corporation.

Bayley, N. (2006). *Bayley scales of infant and toddler development* (3rd ed.). Harcourt Assessment.

Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. *British Journal of Psychiatry*, *175*, 444–451. https://doi.org/10.1192/bjp.175.5.444

Bibby, P., Eikeseth, S., Martin, N. T., Mudford, O. C., & Reeves D. (2002). Progress and outcomes for children with autism receiving parent-managed intensive interventions. *Research in Developmental Disabilities, 23*(1), 81–104. https://doi.org/10.1016/S0891-4222(02)00095-1

Birnbrauer, J. S., & Leach, D. J. (1993). The Murdoch early intervention program after 2 years. *Behaviour Change, 10*(2), 63–74. https://doi.org/10.1017/S0813483900005556

Brownell, R. (2000a). *Expressive one-word picture vocabulary test: Manual*. Academic Therapy Publications.

Brownell, R. (2000b). *Receptive one-word picture vocabulary test: Manual*. Academic Therapy Publications.

Bzoch, K. R., & League, R. (1991). *Receptive-expressive emergent language scale* (2nd ed.). PRO-ED.

Cattell, P. (1960). *The measurement of intelligence of infants and young children*. Psychological Corporation.

Children's Development Center, & Project PRHISE. (1979). *RIDES (Rockford infant development evaluation scales) checklist*. Scholastic Testing Services.

Clark, M. L., Barbaro, J., & Dissanayake, C. (2017). Continuity and change in cognition and autism severity from toddlerhood to school age. *Journal of Autism and Developmental Disorders, 47,* 328–39. https://doi.org/10.1007/s10803-016-2954-7

Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early intensive behavioral treatment: Replication of the UCLA model in a community setting. *Journal of Developmental and Behavioral Pediatrics, 27*(2), S145–S155. https://doi.org/10.1097/00004703-200604002-00013

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*, e17–23. https://doi.org/10.1542/peds.2009-0958

Dunn, L. M., & Dunn, L. M. (1997). *Peabody picture vocabulary test – III* (3rd ed.). American Guidance Service.

Dunn, L. M., Whetton, C., & Burley, J. (1997). *The British picture vocabulary scale* (2nd ed.). NFER-Nelson.

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(*2), 829-835. https://doi.org/10.1016/j.rasd.2011.09.002

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*(1)*,* 49–68. https://doi.org/10.1177/0145445502026001004

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*(3), 264–278. https://doi.org/10.1177/0145445506291396

Eldevik, S., Eikeseth, S., Jahr, E., & Smith, T. (2006). Effects of low-intensity behavioral treatment for children with autism and mental retardation. *Journal of Autism and Developmental Disorders, 36*(2), 211–224. https://doi.org/10.1007/s10803-005-0058-x

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*, 210–20. https://doi.org/10.1007/s10803-011-1234-9

Elliot, C. D. (1990). *Differential ability scales: Introductory and technical handbooks*. Psychological Corporation.

Elliot, C. D., Smith, P., & McCulloch, K. (1996). *British ability scales II*. NFER-Nelson.

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

Farrell, P., Trigonaki, N., & Webster, D. (2005). An exploratory evaluation of two early intervention programmes for young children with autism. *Educational and Child Psychology, 22*(4), 29–40. https://doi.org/10.53841/bpsecp.2005.22.4.29

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*(4), 1479–1492. https://doi.org/10.1016/j.rasd.2011.02.009

Fenson, L., Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. *Applied Psycholinguistics*, *21*(1), 95–116. https://doi.org/10.1017/S0142716400001053

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

Frankenburg, W. K., Dodds, J., Archer, P., Shapiro, H., & Bresnick, B. (1992). The Denver II: a major revision and restandardization of the Denver developmental screening test. *Pediatrics*, *89*(1), 91–97. https://doi.org/10.1542/peds.89.1.91

Gabriels, R.L., Hill, D. E., Pierce, R. A., Rogers, S. J., & Wehner, B. (2001). Predictors of treatment outcome in young children with autism: A retrospective study. *Autism, 5*(4), 407-429. https://doi.org/10.1177/1362361301005004006

Gardner, M. F. (1990). *Expressive one-word picture vocabulary test, revised*. Academic Therapy Publications.

Gesell, A. (1949). *Gesell developmental schedules*. Psychological Corporation.

Haglund, N., Dahlgren, S., Rastam, M., Gustafsson, P., & Kallen, K. (2017). Improvement of autism symptoms after comprehensive intensive early interventions in a clinical setting. *European Psychiatry, 41*, S129. https://doi.org/10.1016/j.eurpsy.2017.01.1940

Harris, S. L., & Handleman, J. S. (2000). Age and IQ at intake as predictors of placement for young children with autism: A four-to six-year follow-up. *Journal of Autism and Developmental Disorders, 30*(2), 137-142. https://doi.org/10.162-3257/00/0400-0137$18.00/0

Hayward, D., Eikeseth, S., Gale, C., & Morgan, S. (2013). Assessing progress during treatment for young children with autism receiving intensive behavioural interventions. *Autism, 13*, 613–33. https://doi.org/10.1177/1362361309340029

Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatment for young children with autism. *Research in Developmental Disabilities, 26*(12)*,* 359–383. https://doi.org/10.1016/j.ridd.2014.08.021

Landa, R. J., & Kalb, L. G. (2012). Long-term outcomes of toddlers with autism spectrum disorders exposed to short-term intervention. *Pediatrics, 130*(Suppl. 2), S186–90. https://doi.org/10.1542/peds.2012-0900Q

Leiter, R. G. (1959). Part 1 of the manual for the 1948 revision of the Leiter International Performance Scale. *Psychological Service Center Journal, 11,* 1–72.

Lichtenberger, E. O. (2014). Sequenced inventory of communication development revised. In C. R. Reynolds, K. J. Vannest, & E. Fletcher-Janzen (Eds.), *Encyclopedia of special education* (1st ed.). Wiley. https://doi.org/10.1002/9781118660584.ese2182

Liuz, D., Barnard, A., Knosen, N., Kotras, N., Horrocks, S., McAlinden, P., et al., (2006). *Griffith mental developmental scales – extended revised: 2 to 8 years*. The Test Agency

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

Lord, C., Rutter, M., DiLavore, P. C., & Risi, S. *Autism diagnostic observation schedule*. 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*(1)*,* 3-9. https://doi.org/10.1037/0022-006x.55.1.3

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 disorder. *Journal of Child Psychology and Psychiatry, 48*(8)*,* 803-812. https://doi.org/10.1111/j.1469-7610.2007.01756.x

Molnar, C., & Eldevik, S. (2017). Verhaltenstherapeutische intervention für vorschulkinder mit autismus [Behavioral intervention for preschool children with autism]. *Zeitschrift für Kinder- und Jugendpsychiatrie und Psychotherapie, 45*(3), 181-191. https://doi.org/10.1024/1422-4917/a000469

Mullen, E. M. (1995). *Mullen scales of early learning*. AGS Publishing.

Perry, A., Koudys, J., Prichard, A., & Ho, H. (2017). Follow-up study of youth who received EIBI as young children. *Behavior Modification, 43*(2), 181–201. https://doi.org/10.1177/0145445517746916

Provence, S., Eriksen, J., Vater, S., & Palmeri, S. (1985). *Infant toddler developmental assessment*. Riverside Publishing.

Reed, P., Osborne, L. A., & Corness, M. (2007). The real-world effectiveness of early teaching interventions for children with autism spectrum disorder. *Exceptional Children, 73*(4), 417–433. https://doi.org/10.1177/001440290707300402

Remington, B., Hastings, R. P., Kovshoff, H., Degli Espinosa, F., Jahr, W., Brown, T., Alsford, A., Lemaic, M., Ward, N., & MacLean, M. E. (2007). A field effectiveness study of early intensive behavioral intervention: Outcomes for children with autism and their parents after two years. *American Journal on Mental Retardation, 112*(6), 418–438. https://doi.org/10.1352/0895-8017(2007)112[418:EIBIOF]2.0.CO;2

Reynell, J. K., & Gruber, C. P. (1990). *Reynell developmental language scales*. Western Psychological Services.

Roid, G. H., & Pomplun, M. (2012). *The Stanford-Binet intelligence scales*. Guilford Press.

Roid, G. H., & Samper, J. L. (2004). *Merrill-Palmer-revised scales of development*. Stoelting.

Rosetti, L. M. (1990). *The Rosetti infant-toddler language scale: A measure of communication and interaction*. LinguiSystems.

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*(6), 417–438. https://doi.org/10.1352/0895-8017(2005)110[417:IBTFCW]2.0.CO;2

Schopler, E., Reichler, R. J., Bashford, A., Lansing, M. D., & Marcus, L. M. (1990). *Psychoeducational profile – revised.* PRO-ED.

Sheinkopf, S. J., & Siegel, B. (1998). Home based behavioral treatment of young children with autism. *Journal of Autism and Developmental Disorders, 28*(1), 15–23. https://doi.org/10.1023/a:1026054701472

Shawler, P. (2016). *Does early intervention reduce the risk of future emotional and behavioral problems in children with autism spectrum disorder.* (Publication No. 10190272) [Doctoral dissertation, Oklahoma State University]. ProQuest Dissertations and Theses Global.

Smith, T., Eikeseth, S., Klevstrand, M., & Lovaas, O. I. (1997). Intensive behavioral treatment for preschoolers with severed mental retardation and pervasive developmental disorder. *American Journal on Mental Retardation, 102*(3), 238–249. https://doi.org/10.1352/0895-8017(1997)102<0238:IBTFPW>2.0.CO;2

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*(4), 269–285. https://doi.org/10.1352/0895-8017(2000)105<0269:RTOIEI>2.0.CO;2

Smith, T., Klorman, R., & Mruzek, D. W. (2015). Predicting outcome of community-based early intensive behavioral intervention for children with autism. *Journal of Abnormal Child Psychology, 43,* 1271–82. https://doi.org/10.1007/s10802-015-0002-2

Sparrow, S. S., Balla, D. A., & Cicchetti, D. V. (1984). *Vineland adaptive behavior scales*. American Guidance Service.

Stock, R., Mirenda, P., & Smith, I. M. (2013). Comparison of community-based verbal behavior and pivotal response treatment programs for young children with autism spectrum disorder. *Research in Autism SpectrumDisorders,7*, 1168–81. https://doi.org/10.1016/j.rasd.2013.06.002

Thorndike, R. L., Hagen, E. R., & Sattler, J. M. (1986). *The Stanford-Binet intelligence scale* (4th ed.). Riverside.

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,* 1673–83. https://doi.org/10.1007/s10803-017-3414-8

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,* 3140–53. https://doi.org/10.1007/s10803-014-2168-9

Voress, J. K., & Maddox, T. (1998). *Developmental assessment of young children*. PRO-ED.

Wechsler, D. (1974). *The Wechsler intelligence test for children - revised*. Psychological Corporation.

Wechsler, D. (1989). *Wechsler preschool and primary scale of intelligence – revised.* Psychological Corporation.

Wechsler, D. (1990). *The Wechsler preschool and primary scale of intelligence revised*. Psychological Corporation.

Wechsler, D. (1993). *Wechsler intelligence test for children* (3rd ed.). Psychological Corporation.

Wechsler, D. (2003). *The Wechsler preschool and primary scale of intelligence administration and scoring manual* (3rd ed.). Psychological Corporation.

Wechsler, D. (2004). *The Wechsler intelligence scale for children* (4th UK ed.). Harcourt Assessment, Psychological Corporation.

Weiss, M. J., & Delmolino, L. (2006). The relationship between early learning rates and treatment outcome for children with autism receiving intensive homebased applied behavior analysis. *Behavior Analysis Today, 7,* 96–110. https://doi.org/10.1037/h0100140

Williams, K. T. (1997). *Expressive vocabulary test*. American Guidance Service.

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

Zachor, D. A., Ben-Itzchak, E., Rabinovich, A., & Lahat, E. (2007). Change in autism core symptoms with intervention. *Research in Autism Spectrum Disorders, 1*(4), 304–317. https://doi.org/10.1016/j.rasd.2006.12.001

Zimmerman, I. L., Steiner, V. G., & Pond, R. E. (1992). *Preschool language scale – 3*. Psychological Corporation.