ASHA American Speech-Language-Hearing Association

JSLHR

Review Article

Long-Term Effects of Early Communication Interventions: A Systematic Review and Meta-Analysis

Natalie S. Pak,[a] Jason C. Chow,[b] Kelsey M. Dillehay,[a] and Ann P. Kaiser[a]

[a] Department of Special Education, Peabody College, Vanderbilt University, Nashville, TN [b] University of Maryland at College Park

ARTICLE INFO

Article History:
Received December 13, 2022
Revision received March 16, 2023
Accepted April 14, 2023

Editor: Mary Alt

https://doi.org/10.1044/2023_JSLHR-22-00711

ABSTRACT

Purpose: Early language and communication interventions for children with language impairments have been shown to be effective in assessments administered immediately after treatment. The purpose of the current systematic review and meta-analysis was to assess the overall durability of those effects over time and whether durability was related to outcome type, etiology of child language impairments, implementer of intervention, magnitude of posttest effects, time between intervention and follow-up, and study risk of bias.

Method: We conducted a systematic search of online databases and reference lists to identify experimental and quasi-experimental group design studies. All studies tested the effects of early communication interventions at least 3 months post-intervention. Participants were children 0–5 years old with language impairments. Coders identified study features and rated methodological quality indicators for all studies. Effect sizes at long-term timepoints and associations with potential moderators were estimated using multilevel meta-analysis with robust variance estimation.

Results: Twenty studies with 129 long-term outcome effect sizes met inclusion criteria. Studies included children with developmental language disorders or language impairment associated with autism. The overall average effect size was small and significant (g = .22, p = .002). Effect size estimates were larger for prelinguistic outcomes (g = .36, p < .001) than for linguistic outcomes (g = .14, p = .101). Significant factors were the posttest effect sizes, the risk of bias for randomized trials, and etiology of language impairment for linguistic outcomes. Time post-intervention did not significantly predict long-term effect sizes.

Conclusions: Outcomes of early language and communication interventions appear to persist for at least several months post-intervention. More research is needed with collection and evaluation of long-term outcomes, a focus on measurement, and consistency of primary study reporting.

Supplemental Material: https://doi.org/10.23641/asha.23589648

As many as 10% of children will have developmental language disorders (DLDs) as indicated by marked difficulties in language development that persist into the school years (Bishop et al., 2016; Norbury et al., 2016; Tomblin et al., 1997). Many other children experience language impairments associated with global developmental delays, autism, or Down syndrome (J. E. Roberts et al., 2007; Tager-Flusberg et al., 2011). Given that achieving early language milestones provides important foundations for later learning (Adamson et al., 2020; Chow et al., 2018), effective early communication interventions for toddlers and preschool-aged children with developmental language delays or disorders may be critical for later linguistic, academic, and social success for these children.

Correspondence to Natalie S. Pak: natalie.s.pak@vanderbilt.edu. ***Disclosure:*** *Ann P. Kaiser was an author on studies that were reviewed in this article. The other authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

Effectiveness of Early Languages Interventions

Literature reviews and meta-analyses have demonstrated positive effects of early language interventions for young children with language impairments. Meta-analyses have identified significant positive effects of parent-implemented language interventions on expressive language outcomes for children younger than 8 years old (Heidlage et al., 2020) and on both language and social communication outcomes for children younger than 6 years old (M. Y. Roberts et al., 2019). In both meta-analyses, effects were larger for children with or at risk for primary language impairments such as DLD than for children with autism or other developmental disabilities. Meta-analyses of early interventions for autistic children have also indicated small to moderate significant positive effects on social communication and spoken language outcomes (Fuller & Kaiser, 2020; Hampton & Kaiser, 2016). In general, early language interventions appear to improve language and social communication outcomes; however, the etiology of children's language impairments may be associated with differential effect sizes.

It is unclear exactly how and to what degree other variables may influence the effectiveness of early language interventions. Relevant variables might include who implements the intervention and how outcomes are measured. Common implementers of early language interventions are caregiving family members (e.g., parents and grandparents) or clinicians. Beginning in kindergarten, more children may attend school full time and receive school-based services. Hampton and Kaiser (2016) concluded that a clinician plus parent model of implementation resulted in better spoken language outcomes than parent-only or clinician-only implementation for autistic children younger than 8 years old. Fuller and Kaiser (2020) found that the implementer of the intervention was not associated with effect sizes for social communication outcomes except in cases in which the outcomes were context bound (Sandbank et al., 2021). In other words, effect sizes were significantly larger when the communication partner, materials, and setting were the same across intervention and outcome measurement contexts. Significant effects for context-bound outcomes are often important but should be interpreted carefully as they do not necessarily indicate generalized change in children's development (Sandbank et al., 2021).

In summary, there are mixed findings regarding whether the implementer of early language and communication interventions is related to outcomes; however, there is evidence of larger intervention effects for context-bound outcomes than generalized outcomes (Chow et al., 2023; Fuller & Kaiser, 2020; Hampton & Kaiser, 2016; Sandbank et al., 2020).

Educational Interventions and Fade-Out

Recent educational research findings have suggested that positive effects of early childhood interventions may not always persist over time. Research on the long-term effects of young children's participation in experimental preschool programs has revealed evidence of fade-out of effects in targeted academic and social skills as early as the end of kindergarten (Abenavoli, 2019; Burchinal et al., 2022; Lipsey et al., 2018). *Fade-out* occurs when the differences in outcomes between participants in the intervention and comparison conditions decrease over time after intervention is removed (D. H. Bailey et al., 2020). Fade-out may not indicate a loss of skills but rather a slowing down in the rate of development in the intervention group (not adding new skills at the same rate) or an acceleration of development in the control group (adding new skills more quickly than children in the intervention group) during the follow-up period. The factors contributing to fade-out are complex but indicate that the observed effects of early intervention may be transitory.

Early Language Interventions and Fade-Out

Previous meta-analyses of the effects of language and communication interventions with children with language impairments have not included long-term effects or addressed the possibility that effects fade out over time. In individual studies, language and social communication outcomes typically are measured soon after the end of the intervention period, with a few exceptions. Hampton et al. (2017) conducted a follow-up study of children with DLD who participated in a randomized controlled trial of the effects of Enhanced Milieu Teaching (M. Y. Roberts & Kaiser, 2015). Significant group differences in receptive language scores immediately post-intervention were nonsignificant by 6-month and 12-month follow-up timepoints, thus demonstrating fade-out of effects. In contrast, in a randomized controlled trial investigating Joint Attention Mediated Learning with autistic children and their caregivers, Schertz et al. (2018) reported large and significant effects at both posttest and 6-month follow-up for social outcomes such as turn-taking and responding to joint attention. These individual studies are informative; however, the extent to which early language intervention effects persist over time across studies remains unknown. In addition, to our knowledge, no meta-analyses have explored associations between long-term outcomes and child etiology, intervention agent, or measurements. Understanding the extent to which the fade-out from targeted interventions for children with language impairments may occur is essential to helping researchers and clinicians to design and implement language interventions that continue to support and potentially accelerate children's language development.

Current Review

In the current systematic review and meta-analysis, we estimated the average long-term effects of early communication interventions on communication outcomes for young children with language impairments. Based on theory and previous research, we selected several potential factors that may influence the persistence or fade-out of early intervention effects. The first factor was the type of communication outcome. Certain communication outcomes may not increase in a linear fashion after a certain developmental stage. Similarly, different intervention targets may follow different trajectories during the period following an intervention. For example, early skills such as babbling and joint attention are foundational for later linguistic skills (e.g., vocabulary and syntax) but do not continue to grow over time (Pence Turnbull & Justice, 2017; Romano et al., 2019). Thus, patterns of fade-out for intervention targets such as rate of vocalizations may differ from fade-out for skills such as receptive or expressive vocabulary, which are expected to show continued growth throughout early development.

The second factor was the etiology of children's language impairments. Studies have shown that the age at which different types of communication behaviors emerge and how long the children rely on them to communicate are related to children's overall language learning abilities (Adamson et al., 2009; Romano et al., 2019). Children with DLD may continue to exhibit improvements in vocabulary or early syntax after an intervention ends, whereas children with global developmental delays who struggle in multiple developmental areas may need sustained support to show continued growth in these areas. In this case, children with DLD could evidence less fade-out than children with other etiologies of language impairments such as autism. The inverse could also be true. Fade-out of intervention effects might be more likely for children with DLD if the children in the comparison group demonstrate relatively greater developmental growth in the later preschool period and catch up with children who had intervention in toddlerhood. This may be especially likely if children in both groups access community-based language-focused interventions. Autistic children participate in interventions with a variety approaches and target outcomes (Sandbank et al., 2020); therefore, access to community-based interventions in the follow-up period may not necessarily lead to catch up for the control group, because the focus of intervention could be distributed across multiple domains.

Third, the implementer(s) of the intervention could moderate persistence or fade-out of early intervention effects. Although meta-analyses have yielded mixed findings regarding the most effective implementer at the immediate posttest for children with autism (Fuller & Kaiser, 2020; Hampton & Kaiser, 2016), findings could be different at later or long-term assessments. When interventions are implemented by only clinicians or researchers, children may lose access to effective language facilitation strategies when the experimental intervention ends. When at-home caregivers (e.g., parents and other family members) or group caregivers (e.g., teachers and daycare providers) are taught to provide intervention to children, they might continue to use those strategies beyond the end of the experimental intervention period. This could create a "sustaining environment" (D. Bailey et al., 2017) that helps children in the intervention group maintain newly learned skills or learn new skills more rapidly than children in the control group.

The fourth and fifth predictors are the magnitude of effects immediately following intervention and the amount of time between posttest and follow-up assessments. When the effect size is large at the end of intervention, there may be proportionally less fade-out, because the control group must make more progress to reach the level achieved by the intervention group. Finally, more time from the end of intervention to follow-up assessment could be associated with greater fade-out of treatment effects. This has been found to be true for interventions targeting increases in IQ (Protzko, 2015) but not for interventions targeting executive functioning skills (Burchinal et al., 2022; Takacs & Kassai, 2019).

For the current systematic review and meta-analysis, we identified group experimental or quasi-experimental studies with outcomes measured at least 3 months after the end of intervention. A time frame of 3 months represents a meaningful length of time for demonstrating sustained effects and would likely be long enough for fade-out of effects to become evident. The primary research questions were as follows: (a) Do language and communication interventions have significant long-term effects on communication and language skills for young children (ages 0 to 5;11 [years;months]) with language impairments? and (b) Do long-term outcomes differ depending on outcome, child, intervention, and design characteristics? We proposed analyses with the following factors determined a priori: (a) type of outcome assessed, (b) etiology of the children's language impairments, (c) intervention implementer, (d) posttest effect sizes, (e) time between posttest and follow-up, and (f) risk of bias in the methods of the individual studies. A Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist (Page et al., 2021) is available in Supplemental Material S1.

Method

Search and Study Selection

Inclusion criteria for selecting studies for review and meta-analysis are summarized in Table 1. Included studies

Table 1. Inclusion criteria and search terms.

Variable	Inclusion criteria	Search terms
Participants	Ages 0–5;11 (years;months; mean of entire sample) at the beginning of intervention, any diagnosis, and any language of intervention	Toddler* OR preschool* OR child* OR infant*
Intervention	Language-focused or communication-focused intervention involving communication/interaction with a partner. Primary intervention targets may include vocabulary, grammar/syntax, narrative production, prelinguistic skills (e.g., joint attention), specific communicative functions, etc. Technology-based interventions excluded.	(Intervention OR treat* OR therapy OR teach* OR support* OR facilitat*) AND (language OR gramma* OR vocabulary OR synta* OR morpholog* OR morphosynta* OR verbal OR linguistic OR communicat*)
Comparison	"Business-as-usual" (BAU) or a similar control condition. This could be another intervention if it is unrelated to the primary outcome measures.	
Outcomes	At least one outcome related to child language or communication measured at least 3 months after the end of intervention with no more than one session in the postintervention follow-up period	
Study design	Experimental or quasi-experimental group designs	"randomized control trial" OR "randomized controlled trial" OR RCT OR "group design" OR "quasi-experimental" OR experimental
Other	Published in English	

(a) were published in English, (b) used experimental or quasi-experimental group designs, (c) included child participants younger than 6 years old (i.e., infants, toddlers, and preschoolers) with evidence of language impairments, (d) compared a language- or communication-focused intervention condition to a business-as-usual (BAU) control condition, and (e) included at least one language or communication outcome measured during an assessment that occurred least 3 months after the end of the intervention. We excluded studies with more than one interventionist contact (e.g., booster session) between the end of intervention and the follow-up period. After beginning the search, we added two inclusion criteria: (a) interventions were interaction-based, rather than delivered primarily via computer, and (b) study reports included sufficient data for effect size calculation.

We searched the PubMed, PsycInfo, ERIC, and Linguistics and Language Behavior Abstracts online databases on March 1, 2021, using the search terms in Table 1. Duplicate records were removed using Zotero (Version 5.0). We screened titles and abstracts of all studies identified from online databases using Rayyan online software (Ouzzani et al., 2016). The first author and a trained undergraduate research assistant each completed title and abstract screening for all identified articles. Articles that were excluded by *both* screeners were eliminated from the search, and we retrieved the remaining articles for full text review (k = 229). The first and third authors, both doctoral students in Early Childhood Special Education, conducted full-text reviews of retrieved articles. Reviewers discussed disagreements and determined final inclusion by consensus. The first author also searched the reference lists of studies identified for inclusion and of relevant recent systematic reviews and meta-analyses (Fuller & Kaiser, 2020; Hampton & Kaiser, 2016; Heidlage et al., 2020; M. Y. Roberts et al., 2019).

Data Collection

Participant and Study Characteristics

We coded relevant features of the study design, participant samples, and experimental and control conditions in a REDCap database (Harris et al., 2019). Design features coded were as follows: (a) publication years of all records for each study, (b) type of research design, (c) type of control group, (d) number and timing of outcome measurement timepoints, (e) number of conditions, and (f) sample size in each group at each timepoint. In addition to the comparison of interest, some studies included comparisons that were not relevant to this review (e.g., symbolic play intervention). In these cases, we only coded the features and outcomes of the target groups. For child participants, coded features included (a) etiology of language impairment (i.e., primary language impairment, autism, Down syndrome, cleft lip and/or palate, hearing loss, or other), (b) age, (c) gender/sex, (d) race/ethnicity, (e) languages spoken, and (f) socioeconomic status indicators. For studies in which caregivers participated in the intervention, we coded the same features (b)–(e) and the caregiver's relation to the child. Coded features of the intervention and control conditions included (a) name, (b) targets, (c) instructional strategies, (d) implementer and training, (e) dosage and duration, and (f) setting.

Outcomes

We coded outcomes related to language and/or communication at long-term (3+ months post-intervention) measurement points in the REDCap database. For each outcome variable, coders identified (a) whether the outcome was primary, secondary, or exploratory; (b) construct (i.e., overall language, overall receptive language, receptive vocabulary, grammatical comprehension, overall expressive language, expressive vocabulary, expressive grammar, or prelinguistic social communication); (c) assessment context (e.g., standardized assessment, structured language sample, caregiver–child interaction, and caregiver report); (d) metric (e.g., frequency, rate, and percentage); (e) name of the assessment tool, if applicable; and (f) any effect size data reported. Coders also extracted data for calculating effect sizes: (a) number of participants per group, (b) mean scores, and (c) standard deviations of scores. Effect size data could include multiple outcome variables for each group at multiple timepoints (e.g., pretest, posttest, and long term), meaning there were several effect sizes for each study and participant sample.

The first author coded all studies. We randomly selected seven studies (35%) for reliability coding by the third author and calculated point-by-point agreement on coded features of the studies and the outcomes. The mean agreement between coders was 88.0% (79.1%–93.5%) for study features (participants, conditions, and design) and 96.6% (91.2%–100.0%) for outcomes. Coders discussed disagreements and arrived at consensus codes for use in final analyses.

Risk of Bias

Coders rated the risk of bias for each long-term outcome at each timepoint based on the effect of assignment to condition (i.e., intention-to-treat). We used the Revised Cochrane Risk-of-Bias Tool for Randomized Trials (RoB 2; Sterne et al., 2019) to evaluate risk of bias of outcomes in randomized designs. The RoB 2 guides reviewers in rating evidence across five domains: (a) randomization process, (b) deviations from intended interventions, (c) missing outcome data, (d) measurement of the outcome, and (e) selection of the reported results. The risk of bias ratings for each RoB 2 domain is *high risk*, *some concerns*, or *low risk*. We used the Risk of Bias in Non-Randomized Studies of Interventions (ROBINS-I; Sterne et al., 2016) for evaluating nonrandomized trials. The ROBINS-I risk of bias domains address: (a) confounding, (b) selection of participants into the study, (c) classification of interventions, (d) deviations from intended interventions, (e) missing data, (f) measurement of outcomes, and (g) selection of the reported result. Risk of bias ratings for each ROBINS-I domain are *critical*, *serious*, *moderate*, *low*, or *no information*. The first and third authors independently completed the RoB 2 and ROBINS-I for all long-term outcome data (i.e., all outcomes at all timepoints). Coders discussed discrepancies and determined consensus codes for use in the final analyses.

Methods of Synthesis

All analyses were conducted using RStudio Version 1.2.5033 (R Core Team, 2020). We calculated Hedges' g standardized mean differences (Hedges, 1981) for each selected outcome at each timepoint using the escalc function in the metafor package (Viechtbauer, 2010). Findings were synthesized using random-effects meta-analyses with robust variance estimation (RVE). RVE allows for meta-analysis of clusters of related effect sizes without knowing exact correlations (Hedges et al., 2010). This allowed for analysis of multiple effect sizes per study or participant sample. The correlated and hierarchical effects (CHE) working model accounted for between-study and within-study heterogeneity as well as dependence among effect sizes from the same study (Pustejovsky & Tipton, 2022). The assumed correlation among effect sizes from each study (rho) was held constant across studies at 0.6. Sensitivity analyses at various values of rho (i.e., .2, .4, and .8) were conducted to confirm that findings were robust to varying levels of correlated dependent effects.

First, we estimated the overall mean effect size at long-term timepoints using an intercept-only model. We then added predictor variables one at a time to the model. The predictors corresponded to our factors determined a priori: (a) type of outcome, (b) etiology of the children's language impairments, (c) intervention implementer, (d) posttest effect sizes, (e) time between end of intervention and follow-up assessments, and (f) risk of bias in the methods of the studies. For each factor, we first examined the distribution of the levels of the variable across studies to determine whether there was enough variability to proceed with the analysis. We terminated the analysis when categorical variables were highly imbalanced, which can lead to variation in Type I error rates (Tipton, 2015). We also terminated the analysis when there were few (< 5) studies or effect sizes in one or more categories of the variable, and the categories could not be aggregated logically.

Publication Bias and Selective Reporting

We sought evidence of publication bias and selective reporting in multiple ways. The first indicator was the selective reporting domain rating of the RoB 2 and the ROBINS-I. Coders rated this domain using evidence from clinical trial registries, trial protocols, or other sources, when available. Second, funnel plots were visually inspected for asymmetry. Third, meta-regression analyses were conducted to assess potential funnel plot asymmetry while accounting for dependence among effect sizes (Rodgers & Pustejovsky, 2021).

Results

Study Selection

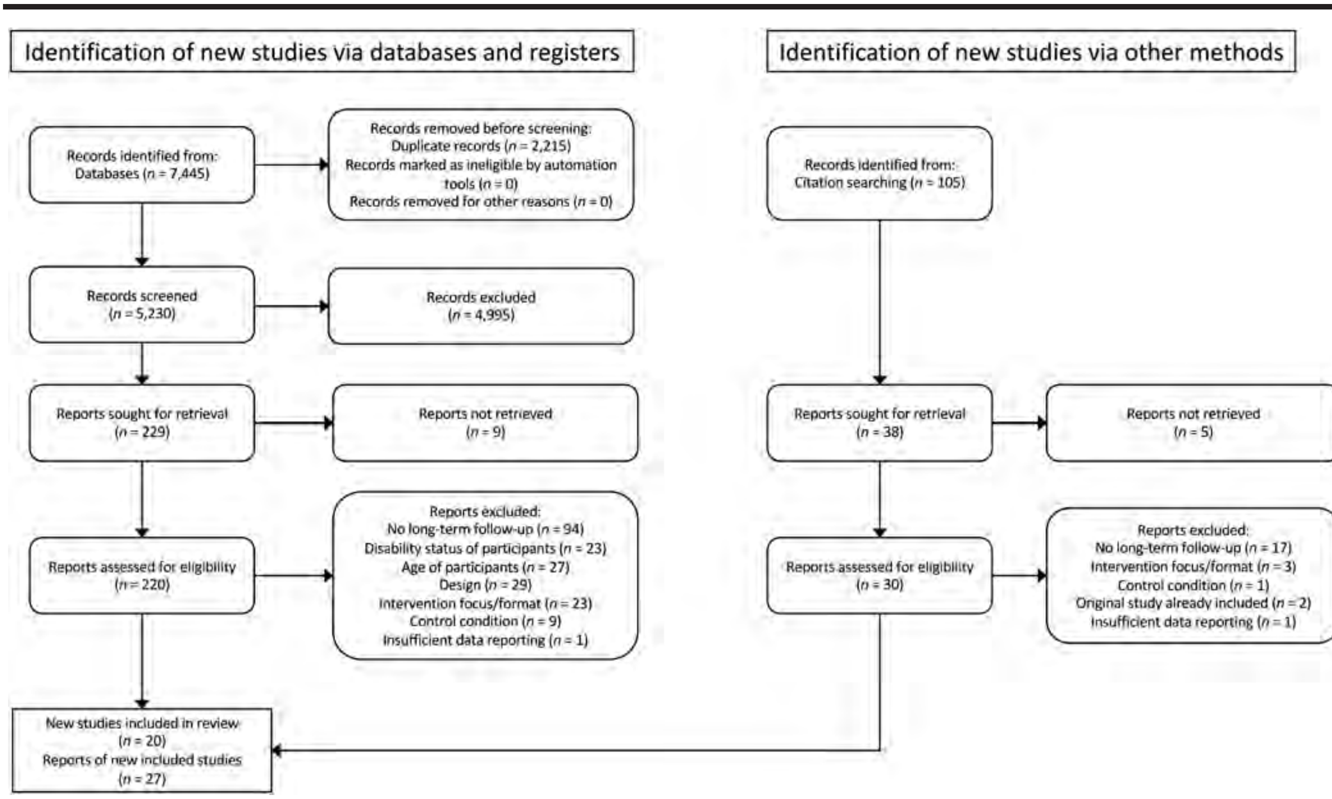
Figure 1 shows the number of records excluded at each stage of the search process (Page et al., 2021). From the online database search, we screened 5,230 records based on titles and abstracts and reviewed the full text of 220 articles. From backward searching, we screened 105 records based on titles and abstracts, and we reviewed the full text of 30 articles. After exclusions of studies that did not meet eligibility criteria, we included 20 studies with unique participant samples in the final review. Supplemental Material S2 lists the references for studies that nearly met criteria for inclusion and the reasons for exclusion. Supplemental Material S3 lists the references for articles associated with the included studies.

Descriptive Synthesis of Results

Participant Characteristics

The mean age of children across studies ranged from 10.6 to 58.3 months (M = 33.9 months). Thirteen studies included children with diagnoses or characteristics of autism (999 participants). The remaining seven studies included children with language disorders that were not specified as secondary to other diagnoses such as autism, genetic syndromes, or hearing loss (640 participants). Inclusion and exclusion criteria varied across these seven studies, but most included children with receptive and expressive language delays or disorders, children with specific expressive language delays or disorders, or both. For the purposes of this review, we refer to this population as children with DLD (Bishop et al., 2016; McGregor et al., 2020), despite the relatively young age of the children and the variability in receptive language scores. Language impairment seemed to be the primary diagnosis, and children with cognitive deficits were sometimes included in the study samples. Gender or sex was reported in all but one study. On average, 78% (range: 51%–91%) of children were males and 22% (range: 9%–49%) were females. Information about race and/or ethnicity of the child participants was included in 13 studies. Across studies, 20% (Siller et al., 2013) to 100% (Baxendale & Hesketh, 2003) of participants were White. Few studies included reports on proportions of participants of other races/ethnicities; however, two studies (Carter et al., 2011; Siller et al., 2013) included around 40% participants that were Hispanic or Latine. When reported or able to be inferred, children's home languages were English (k = 7), German (k = 2), Dutch (k = 1), or Norwegian (k = 1).

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 flow diagram. This figure was based on the PRISMA 2020 template (Page et al., 2021).



Across studies with family caregiver involvement (k = 17), there was minimal information about the caregivers. Most caregivers were female (88%–100%) in the seven studies that reported gender.

Design and Experimental Conditions

Study characteristics are summarized in Table 2. Each study's findings were reported in one to three publications between 1999 and 2020. The studies took place in seven countries. There were 14 RCTs and six nonrandomized trials. Sample sizes in the intervention and control groups combined ranged from 43 to 200 participants or dyads (M = 81.95) for a total of 1,639 participants or dyads across studies. Intervention periods were 1–13 months (M = 3.79). A caregiving family member was the primary implementer of the intervention in the majority of the studies (k = 14). Other implementers were interventionists (k = 2), teachers (k = 1), or the family caregiver and an interventionist (k = 3).

Outcome Characteristics

There were 129 long-term effect sizes for language or social communication outcomes across studies, with up to 14 effect sizes per study (M = 6.5, Mdn = 4.5). The duration of the period between the end of the intervention and follow-up was 3–71 months (M = 10.3, Mdn = 7.0). Pretest effect sizes (i.e., the differences in scores between groups at baseline) were available for 78% of long-term outcomes (n = 100), and posttest effect sizes were available for 45% of long-term outcomes (n = 58). Missing pretest or posttest data were often the result of assessments at follow-up that were not administered in the original study (e.g., Kasari et al., 2012, Expressive Vocabulary Test). In some cases, researchers did not expect effects immediately after intervention, so they measured intervention effects only at long-term timepoints and not at posttest timepoints (e.g., McConachie et al., 2005).

Statistical Synthesis of Results

Table 3 shows effect size estimates at long-term timepoints. The overall mean effect size was small and significant in favor of the intervention groups (g = .22, SE = .06, and p = .002).

Outcome Types

Outcome type was the first variable added to the model. Given the distribution, we categorized outcomes as prelinguistic (k = 13, n = 54) or linguistic (k = 18, n = 75). Prelinguistic outcomes included targets such as joint attention, joint engagement, turn-taking, or rate of vocalization. Linguistic outcomes included targets such as the number of different words used during a language sample, mean length of utterance, or standardized language assessment scores. Estimates were significant for prelinguistic outcomes (g = .36, SE = .08, and $p < .001$) and nonsignificant for linguistic outcomes (g = .14, SE = .08, p = .101). Given these differences in effect size estimates, as well as theoretical differences between prelinguistic and linguistic behaviors, the remaining meta-regression analyses were conducted separately for prelinguistic and linguistic outcomes. Table 3 shows the results of the meta-regression analyses.

Etiology

Based on the distribution of etiologies of language impairments for children in the selected studies, etiology could be categorized as DLD or autism. For prelinguistic outcomes, only three effect sizes from two studies were available from children with DLD. The remaining prelinguistic outcomes (k = 11, n = 51) were from children with autism. Given this imbalance in the numbers of studies and effect sizes across etiologies for prelinguistic outcomes, we did not continue with the analysis of the association between etiology and outcomes for prelinguistic outcomes. For linguistic outcomes, etiology was more balanced across studies with samples characterized by DLD (k = 7, n = 51) and autism (k = 11, n = 24). The effect size estimates were small and nonsignificant for children with DLD (g = .36, SE = .14, p = .052) and autism (g = .01, SE = .08, p = .881).

Implementer

Interventions were implemented primarily by a caregiving family member in most studies (k = 14). There were few studies with clinician-implemented or clinician plus caregiver implemented interventions in each data set (see Table 3). Without a clear theoretical rationale for aggregating the remaining implementer types into one category, we did not perform meta-regressions with implementer for either prelinguistic or linguistic effects.

Posttest Effect Sizes

Only 45% of long-term outcomes had corresponding posttest effect sizes available (k = 12, n = 58). For prelinguistic outcomes for which posttest data were available (k = 11, n = 32), the mean effect size for posttest outcomes was g = .36 (SE = .12, p = .013). For linguistic outcomes (k = 8, n = 26), the mean effect size for posttest outcomes was g = .12 (SE = .08, p = .200). As expected, posttest effect sizes were significantly related to long-term effect sizes for both prelinguistic (ß = .49, SE = .04, $p < .001$) and linguistic outcomes (ß = .43, SE = .10, p = .044). Findings should be interpreted cautiously, however, as these analyses included a relatively small subset of the data. Of note, the meta-regression analysis for the linguistic outcomes was sensitive to varying levels of correlated dependent effects (see Supplemental Material S4), possibly due to the especially small sample of studies and effects.

Table 2. Summary of study characteristics.

Study ID[a]	Country	n[b]	Age[c]	Etiology	Intervention condition	Control condition	Implementer(s)	Intervention duration (months)[d]	Long-term timepoints (months post-intervention)
Baxendale & Hesketh, 2003	UK	49	32(2)	DLD	Hanen Parent Programme (HPP)	BAU	Caregiver	2.75	3, 9
Brian et al., 2017	Canada	63	25(3)	Autism	Social ABCs	WL	Caregiver	3	3
Buschmann et al., 2015	Germany	58	25(1)	DLD	Heidelberg Parent-Based Language Intervention (HPLI)	WL	Caregiver	3	3, 9, 21
Carter et al., 2011	USA	62	20(3)	Autism	Hanen's "More than Words" (HMTW)	BAU	Caregiver	2	3, 7
Hampton et al., 2017	USA	97	30(5)	DLD	Enhanced Milieu Teaching (EMT)	BAU	Caregiver	3	6, 12
Hampton et al., 2020	USA	68	43(5)	Autism	Multicomponent Adaptive Intervention (JASP + EMT + SGD + DTT)	BAU	Caregiver + interventionist	4	4
Kaale et al., 2014	Norway	61	48(8)	Autism	Joint Attention Intervention	BAU	Teacher	2	6, 12
Kasari et al., 2012	NR	43	43(7)	Autism	Joint Attention Intervention	BAU	Interventionist	1.25–1.50	6, 12, 60
Kasari et al., 2014	USA	66	22(4)	Autism	Focused Playtime Intervention (FPI)	PE	Caregiver	3	9
Kasari et al., 2015	USA	86	31(3)	Autism	Joint Attention, Symbolic Play, Engagement, and Regulation (JASPER)	PE	Caregiver	2.5	6
Kruythoff-Broekman et al., 2019	Netherlands	63	25(1)	DLD	The Target Word programme	BAU	Caregiver	3	6, 18
Landa et al., 2011	USA	50	29(3)	Autism	Interpersonal Synchrony (IS) supplemental social curriculum	BAU	Interventionist	6.5	6
McConachie et al., 2005	UK	56	38(7)	Autism	Hanen's "More than Words" (HMTW)	WL	Caregiver	3	4
Motsch & Ulrich, 2012	Germany	54	NR	DLD	Lexicon Pirate	BAU	Caregiver + interventionist	1.25	6, 12
Pickles et al., 2016	UK	152	45(NR)	Autism	Preschool Autism Communication Trial (PACT)	BAU	Caregiver	12	71
Poslawsky et al., 2015		78	43(10)	Autism	Video-feedback Intervention to promote Positive Parenting adapted to Autism (VIPP-AUTI)	BAU	Caregiver	3	3
Schertz et al., 2018	USA	144	25(4)	Autism	Joint Attention Mediated Learning (JAML)	BAU	Caregiver	8	6
Siller et al., 2013	USA	70	58(13)	Autism	Focused Playtime Intervention (FPI)	PE	Caregiver	3	12
Wake et al., 2015	Australia	200	50(1)	DLD	Language for Learning intervention	BAU	Caregiver + interventionist	4.5	12
Ward, 1999	UK	119	10(2)	DLD	Intervention Program	BAU	Caregiver	2–8	8, 20

Note. UK = United Kingdom; DLD = developmental language disorder; JASP = Joint Attention, Symbolic Play, Engagement, and Regulation intervention; EMT = Enhanced Milieu Teaching; SGD = speech-generating device; DTT = discrete trial teaching; BAU = business-as-usual; WL = waitlist; USA = United States of America. PE = parent education; Caregiver = a caregiving family member; NR = not reported.

[a]Authors and publication year of most recently published article. [b]Excludes groups other than the intervention group and control group. [c]Age of the entire sample when reported; otherwise, the age of the children in the intervention group. Rounded to the nearest month. *M*(*SD*). [d]Intervention duration, when reported in weeks, was converted to months by dividing by 4.

Table 3. Meta-regression model results.

Data set & predictor	Subgroups	Studies	Effect sizes	Estimate	*SE*	*df*	*p*	95% CI
All outcomes		20	129	**.22**	.06	16.3	.002	[.10, .34]
Outcome type	Prelinguistic	13	54	**.36**	.08	15.5	< .001	[.19, .52]
	Linguistic	18	75	.14	.08	17.0	.101	[–.03, .31]
Prelinguistic outcomes		13	54					
Etiology	DLD	2	3	—	—	—	—	—
	Autism	11	51	—	—	—	—	—
Implementer	Caregiver	8	30	—	—	—	—	—
	Interventionist	3	20	—	—	—	—	—
	Both	2	4	—	—	—	—	—
Posttest effects	n/a	11	32	**.49**	.04	3.72	< .001	[.38, .60]
Time to follow-up	n/a	12	52[a]	.001	.02	5.75	.957	[–.04, .04]
Linguistic outcomes		18	75					
Etiology	DLD	7	51	.36	. 14	5.20	.052	[–.004, .73]
	Autism	11	24	.01	.08	9.59	.881	[–.17, .19]
Implementer	Caregiver	12	57	—	—	—	—	—
	Interventionist	3	8	—	—	—	—	—
	Both	3	10	—	—	—	—	—
Posttest effects	n/a	8	26	**.43**	.10	2.24	.044	[.03, .84]
Time to follow-up	n/a	17	73[a]	–.02	.01	2.56	.071	[–.04, .003]

Note. Estimates in bold are significant at $p < .05$. Dashes indicate that the analysis was not conducted. *SE* = standard error; *df* = degrees of freedom; CI = confidence interval; DLD = developmental language disorder.

[a]Outcomes measured more than 3 *SD*s above the mean were removed from analysis.

Time to Follow-Up

Nearly all long-term measurements (97%) occurred within 21 months post-intervention. Outcomes measured more than 3 *SD*s above the mean period post-intervention were omitted from analyses to prevent these outliers from skewing findings ($k = 2$, $n = 4$). The cutoffs were approximately 47 months post-intervention for the prelinguistic outcomes data set and 42 months post-intervention for the linguistic outcomes data set. Two effect sizes were omitted from each analysis. Time to follow-up did not significantly predict long-term effects for either prelinguistic or linguistic outcomes (see Table 3). It is important to note that long-term effects in this sample were from many different participant samples. Many measures were only gathered from a sample at a single long-term timepoint ($n = 67$). Longitudinal data sets with multiple long-term timepoints per outcome would be ideal for fully understanding how effects persist or fade-out over time.

Assessment of Risk of Bias and Quality

Across 14 RCTs, we analyzed risk of bias for 91 outcomes. Risk ratings in each domain are displayed in Figure 2. The overall risk of bias for outcomes measured in randomized controlled trials was *High Risk* for 45.1%, *Some Concerns* for 52.7%, and *Low Risk* for 2.2% of long-term outcomes. *High risk* of bias in the "Measurement of the Outcome" domain was often the result of covariation between condition and assessment contexts, or context-bound outcome measurement (Sandbank et al., 2021).

Missing outcome data was another domain with relatively high risk of bias across studies. Of the 91 outcome measures across the 14 RCTs, the average proportion of data available for all randomized participants was 85.8% (range: 60.5%–100%). For the "Selection of the Reported Result" domain, most outcomes received a rating of *some concerns,* because there was rarely enough information provided about analysis plans in clinical trial registries or protocols published before the analysis to ascertain whether selective reporting was present. In all domains except "Selection of the Reported Result" and "Measurement of the Outcome," the majority of outcomes received a rating of *low risk*. Generally, randomized studies had adequate procedures for randomization of participants to interventions, and any deviations from assigned conditions (e.g., a participant dropping out of the study) were not likely to be related to the research context. Most research teams used appropriate intention-to-treat analyses.

In the six studies with nonrandomized designs, we analyzed 39 outcomes. The overall risk of bias was high, with ratings of *serious* for 71.8% ($n = 28$) and *moderate* for 28.2% ($n = 11$) of outcomes. Most of the *serious* risk ratings were from bias due to confounding variables

Figure 2. Risk of bias summary for randomized (top) and nonrandomized (bottom) trials. This figure displays results of ratings from the Revised Cochrane Risk-of-Bias Tool for Randomized Trials (Sterne et al., 2019) and the Risk of Bias in Non-Randomized Studies of Interventions (Sterne et al., 2016).

(Baxendale & Hesketh, 2003; Kruythoff-Broekman et al., 2019) or in selection of the reported result (Ward, 1999). In two studies, the groups differed in age (Baxendale & Hesketh, 2003) or receptive language level (Kruythoff-Broekman et al., 2019) at baseline and the authors were not able to adequately control for these variables in the analyses. In one study, authors reported that two standardized language assessments were administered but scores from only one were reported (Ward, 1999). All outcomes across the six nonrandomized trials received *low-risk* ratings in the domains related to selection of participants and classification of interventions.

Risk of Bias Moderator Analysis

We conducted a random-effects meta-analysis with RVE with the overall risk of bias rating as a predictor. We conducted separate analyses for randomized and nonrandomized designs. Nearly all of the randomized study outcomes could be categorized as having *some concerns* or *high risk*, so we included the *low risk* outcomes and the *some concerns* outcomes in the same category (k = 12, n = 50). The outcomes with *high risk* of bias (k = 10, n = 41) had small and significant effect sizes (g = .29, SE = .07, p = .002), while the outcomes with *low risk* or *some concerns* had much smaller, nonsignificant effect sizes (g = .07, SE = .06, p = .268). There were not enough nonrandomized studies (k = 6) to conduct an analysis with risk of bias as a predictor (Tanner-Smith & Tipton, 2014).

Publication Bias and Selective Reporting

Only published studies were identified for inclusion in this synthesis. Unpublished studies were included in the search process, but none were identified that met inclusion criteria. Funnel plots for the whole data set and for prelinguistic and linguistic outcomes data sets are in

Supplemental Material S5. Data were not significantly asymmetrical in the funnel plots for the overall data set (p = .592), prelinguistic outcomes only (p = .424), or linguistic outcomes only (p = .873). Visual inspection of funnel plots revealed several outliers, which warranted a sensitivity analysis. Without the outliers (all linguistic outcomes from Ward, 1999), there were 19 studies and 121 effect sizes. The overall effect size estimate was somewhat smaller but remained significant (g = .18, SE = .05, p = .002) when the outliers were removed. When outcome type was added to the model without the outliers, the estimate for prelinguistic outcomes decreased but remained significant (g = .30, SE = .07, p < .001). The estimate for linguistic outcomes also decreased and remained nonsignificant (g = .09, SE = .06, p = .163).

Discussion

The purpose of this systematic review and meta-analysis was to investigate the long-term effects of early communication interventions. Twenty studies meeting inclusion criteria were identified via a systematic search of the literature. We conducted descriptive and meta-analytic syntheses of the studies.

The results of the descriptive synthesis illuminated intervention research priorities for infants, toddlers, and preschoolers with language impairments and the measurement of long-term outcomes beyond the posttest. The children in all studies in the sample were selected based on characteristics of autism or DLD. Most studies involved interventions with at least partial family caregiver implementation (k = 17), although specific caregiver characteristics were insufficiently reported to allow systematic analysis of the impact of individual differences. Long-term outcome measurement across studies typically occurred within 2 years post-intervention.

The descriptive synthesis also revealed gaps in the literature involving long-term outcomes of young children who received language and communication interventions. There was a notable lack of long-term measurement of intervention outcomes for children with primary etiologies other than autism or DLD, such as Down syndrome. This omission is especially notable given the high prevalence of Down syndrome and the early age at which it is identified (Presson et al., 2013). Data on the race/ethnicity of participants were reported in many studies, but the reporting formats across studies made it difficult to synthesize these findings. The home languages and language profiles of children in the sample remain relatively unknown. In some studies, children were all monolingual (e.g., Baxendale & Hesketh, 2003); in others, families were included if they were proficient in the language of the intervention and may have spoken additional languages (e.g., Kruythoff-Broekman et al., 2019). Specific descriptions of the cultural and linguistic characteristics of the sample are important for understanding the generalizability of the effect across children and families and how effects in a given language might be expected to persist or fade out over time.

Study results were statistically synthesized via multilevel meta-analysis using effect sizes and RVE. For outcome assessments at least 3 months postintervention (M = 10.26 months), there was a small and significant overall difference in outcomes in favor of the intervention group (g = .22, p = .002). Previous meta-analyses involving early communication interventions for young children with language impairments have reported similar small to moderate effects post-intervention (Fuller & Kaiser, 2020; Hampton & Kaiser, 2016; Heidlage et al., 2020; M. Y. Roberts et al., 2019).

Meta-regression analyses were employed to test the influences of variables with conceptual and evidence-based links to long-term intervention effects. These predictors included (a) type of outcome measure, (b) etiology of the children's language impairments, (c) intervention implementer, (d) posttest effect sizes, (e) time between end of intervention and follow-up, and (f) degree of risk of bias in the methods of the studies. Based on the characteristics of these studies, we categorized outcomes as prelinguistic or linguistic and etiologies as DLD or autism.

Intervention effects were larger for prelinguistic outcomes (g = .36) than linguistic outcomes (g = .14) overall at long-term assessment timepoints. However, the participant samples contributing to the average effect sizes for prelinguistic and linguistic outcomes differed. Almost all prelinguistic outcomes involved interventions tested with autistic children, whereas linguistic outcomes were measured in more heterogeneous participant samples. These differences must be considered when interpreting average effect sizes for each category of outcomes. The magnitude of the effects for linguistic outcomes for the subset of children with DLD (g = .36, p = .052) was similar to the magnitude of the effects for prelinguistic outcomes, which primarily applied to children with autism (g = .36, p < .001). Therefore, the findings from this meta-analysis indicate that autistic children may receive significant long-term benefits of intervention for prelinguistic outcomes, and children with DLD may receive similar benefits of intervention for linguistic outcomes.

The differences in outcomes measured for the subsamples of children with DLD and autism likely reflect intervention priorities based on child characteristics such as stage of communication development and areas of need. As discussed above, studies with samples of autistic children often included both prelinguistic and linguistic measures, whereas studies with children with DLD tended to only include linguistic measures at long-term timepoints. Autistic children demonstrate varying levels of difficulty with early social communication development, including delayed development of prelinguistic behaviors such as joint attention and engagement, whereas children with DLD often demonstrate less difficulty in these areas (Pence Turnbull & Justice, 2017). Accordingly, interventions in the current sample designed for autistic children may have focused more on development of prelinguistic skills, whereas interventions designed for children with DLD focused on linguistic skills such as vocabulary and grammar. The focus of intervention in combination with child foundational skills at entry may explain the larger *linguistic* effect sizes for children with DLD (g = .36, p = .052) than autistic children (g = .01, p = .881).

We posited that long-term effects would be positively and significantly associated with effect sizes at posttest. Posttest effects were available for a subset of the sample, and the analysis for linguistic outcomes was a particularly small sample (k = 8, n = 26). Posttest effect sizes significantly predicted long-term effect sizes for both prelinguistic and linguistic outcomes. These findings should be interpreted cautiously, however, as they were somewhat sensitive to different assumptions about the dependence among effect sizes. With this caveat, larger immediate effects of intervention may be more likely to persist than smaller immediate effects. A large posttest effect might still be significant at long-term timepoints, even if a small degree of fade-out occurs. The same may not be true for small posttest effect sizes. Surprisingly, we found no significant relation between the number of months between posttest and follow-up and the effect sizes at the long-term timepoints. It is important to note that this analysis was limited to 20 studies and 129 effect sizes. Many outcome measures (n = 67) had a single long-term data point. To fully understand the persistence of effects over time would require a sample of studies with multiple postintervention assessments occurring over a substantial period of time (e.g., up to 18 months after intervention). A larger sample of studies would also allow for more complex analyses with multiple variables in the same model. It is possible that the time from posttest to follow-up could be a significant predictor for a subset of the outcomes. Such analyses and findings would have implications for intervention planning. For example, researchers and clinicians might decide to maximize postintervention outcomes specifically for outcome categories that are more vulnerable to fade-out over time. Additional longitudinal research is needed to draw conclusions about the durability of outcomes over time and the extent of the resilience to fade-out.

We examined the risk of bias of this group of studies using the RoB 2 and the ROBINS-I tools (Sterne et al., 2016, 2019). In the current meta-analysis, randomized controlled trials with higher risk of bias were also more likely to have significant effect sizes at long-term timepoints. Prior meta-analyses have reported null associations between overall risk of bias ratings and effect sizes (Fuller & Kaiser, 2020; Hampton & Kaiser, 2016; M. Y. Roberts et al., 2019). One reason for the difference between current and prior findings could be the measurement tools used to estimate bias in the studies. The RoB 2 was published relatively recently and provides more thorough guidance for decision making regarding risk of bias than previous versions of the tool (Sterne et al., 2019). The use of this tool may be sensitive to risks that had not been detected in previous reviews using earlier tools (e.g., Gersten et al., 2005; Higgins et al., 2011). At the same time, the standards of RoB 2 and the ROBINS-I may be unduly restrictive in evaluating some domains for this group of studies. Nearly all outcomes had at least a moderate overall risk of bias. One common concern was missing outcome data. According to the RoB 2 and ROBINS-I, researchers should report on data from at least 95% of participants assigned to groups by the time of outcome measurements to be considered to have low risk of bias in this domain. At long-term follow-up timepoints, retaining participants for assessments is more difficult than at immediate postintervention timepoints. Thus, the 95% criterion may have been too strict for assessing long term outcomes; studies on average maintained 85.8% (60.5%–100%) of participants at follow-up.

Another common risk of bias measured by the RoB 2 and ROBINS-I was related to measurement. In the current sample of studies, context-bound measures were common (Sandbank et al., 2021; Yoder et al., 2018). For example, child communication behaviors were often measured during interactions with a caregiver, when presumably caregivers in the intervention group used more strategies to promote child communication. Similarly, trained caregivers could have been more sensitive to child communication skills when reporting their child's communication skills post-intervention on measures such as the MacArthur-Bates Communicative Development Inventories (Fenson et al., 2007). Studies with context-bound posttest measures tend to yield larger effect sizes than studies with generalized measures (Fuller & Kaiser, 2020; Sandbank et al., 2020).

We did not code for boundedness of long-term outcome measurement directly; rather, the outcome measurement domains of the RoB 2 and the ROBINS-I address the presence of context-bound outcomes indirectly through a series of questions including whether the assessors were blind to group assignment and whether measurement was likely to have differed between groups (Sterne et al., 2019). We could not isolate the association between context-bound measurement and average long-term effect sizes based on the current data. Nevertheless, it is likely that caregiver involvement in both intervention and assessment contributed to the larger effects for RCT outcomes with *high-risk* ratings. Future meta-analyses should directly address the association between boundedness of outcomes and long-term effect sizes.

Context-bound outcomes need not be dismissed but should not be interpreted as generalized developmental gains (Sandbank et al., 2021). Schertz et al. (2018) included context-bound outcome measures and acknowledged that "post-intervention assessment was meant to capture the interactional dynamic rather than child behaviors in isolation" (p. 864). The presence of context-bound measurement does not diminish the value of caregiver involvement in early childhood interventions. Because family members are children's most important communication partners in early childhood, these types of context-bound communication gains are undoubtedly valuable. However, the generalization of skills to interactions with other people (e.g., teachers and peers) will also be important for children's development, especially in the long term.

Strengths and Limitations

The current systematic review uses meta-analysis to review and synthesize long-term early communication intervention effects. This approach goes beyond the study of average intervention effects on child outcomes to specifically study the magnitude of long-term effects and presence or absence of fade out. We included all relevant effect sizes in this analysis rather than selecting one per study and used the CHE working model for analysis to account for between-study and within-study heterogeneity and dependence among effect sizes from the same study (Pustejovsky & Tipton, 2022) as opposed to selecting between models that account for one type of effect size dependency.

There are also limitations to the current findings. The first limitation is the relatively small number of studies and effect sizes. RVE analyses can be performed with as few as 10 studies, but a sample size of at least 40 studies with an average of five effect sizes per study would be ideal for conducting moderator analyses with RVE (Tanner-Smith & Tipton, 2014). We were unable to explore the relation between intervention implementer and follow-up effects in these studies, because few studies included therapists or teachers as interventionists ($k = 6$). We dichotomized some variables to increase sample sizes for analysis of the subgroups. For example, prelinguistic outcomes included a range of behavioral outcomes ranging from joint attention to affect sharing and turn-taking. Similarly, the linguistic outcomes included skills ranging from sentence comprehension to expressive vocabulary. With a larger sample of studies, we might have been able to explore differences in long-term effects with more precision.

Generalizability

Although the relative homogeneity of implementers across these interventions (85% of studies had family caregivers as implementers in some capacity) limited use of statistical analyses for detecting the relative influence of the specific implementer on long-term effects, it strengthened our confidence in the external validity of the findings. When interventions are primarily implemented by members of the research team, we cannot assume that the intervention procedures will be acceptable, feasible, or effective when implemented at scale by community members such as parents, teachers, or clinicians (Odom, 2009).

The generalizability of the current findings is limited in several ways. The overall number of studies was small. All studies were conducted in countries dominated by western cultural values and cannot be generalized to other cultural groups. Second, we could not summarize the wide variability in current studies' reports of socioeconomic status, race/ethnicity, and home languages adequately for the group of studies. Third, effect sizes for outcomes in this sample of studies spanned a relatively short time after the experimental intervention (less than 24 months later). There were two exceptions with outcomes 5–6 years post-intervention (Kasari et al., 2012; Pickles et al., 2016). For many children with DLD or autism, language and communication challenges persist long beyond early childhood (Bishop et al., 2016; Tager-Flusberg et al., 2011). More extended long-term follow-up studies are needed before we can confidently make inferences about persistence or fade-out of the effects of early communication interventions.

Implications

The current review indicated that early communication intervention effects generally persist for at least several months post-intervention. D. Bailey et al. (2017) defined "trifecta skills" as ideal intervention targets that are less likely to fade out over time than other skills. Trifecta skills are named for three characteristics: They are (a) malleable, (b) fundamental to continued development, and (c) potentially unlikely to develop without targeted intervention. Prelinguistic outcomes did not entirely fade out for autistic children in the current sample of studies; thus, prelinguistic intervention targets could be trifecta skills for this population. In contrast, long-term intervention effects for linguistic outcomes were nonsignificant for autistic children ($g = .01$, $p = .881$), possibly because they were less emphasized in intervention. If prelinguistic skills are indeed trifecta skills that are fundamental to continued development of language skills, concurrent high-quality intervention with specific linguistic targets appears to be essential for many autistic children demonstrating language impairments.

More research on the long-term effects of early communication interventions is needed. To support this type of research, funding agencies should prioritize long-term follow-up studies and proposals for initial studies that include long-term measurements. We did not code details related to funding sources for the current sample of studies, but access to substantial grant funding is likely to be a major factor in the ability to measure long-term outcomes. Given the substantial cost of intervention research, additional costs for conducting follow-up assessments are likely to be prohibitive for many researchers.

Conclusions

In this systematic review and meta-analysis, we synthesized long-term outcomes from early communication intervention studies. This review evaluated early language intervention research for children with language impairments from the perspective of fade-out of effects; this perspective has been informative when applied in early childhood educational research (Abenavoli, 2019; Burchinal et al., 2022). Based on the modest sample of studies meeting inclusion criteria, the meta-analysis indicated overall significant small effects were still present several months after interventions concluded. The magnitude and significance of the maintained effects varied by type of outcome, etiology of language impairment, and posttest effect sizes. Overall, at least some effects persisted beyond the end of intervention and were resilient against fade-out for several months. More research is needed to understand the short-term outcomes, participant characteristics, and intervention characteristics that contribute to lasting improvements in language skills over time for young children with language impairments.

Data Availability Statement

The data sets are available in the Open Science Framework repository, https://osf.io/9cd2a/, or from the first author.

Acknowledgments

This research was funded by a professional development grant (H325D180095, A. Kaiser, PI) from the U.S. Office of Special Education Programs and a graduate training fellowship from Peabody College at Vanderbilt University. We thank Caroline Wilkerson for her contributions to the data collection for this project.

References

Abenavoli, R. M. (2019). The mechanisms and moderators of "fade-out": Towards understanding why the skills of early childhood program participants converge over time with the skills of other children. *Psychological Bulletin, 145*(12), 1103–1127. https://doi.org/10.1037/bul0000212

Adamson, L. B., Bakeman, R., Deckner, D. F., & Romski, M. (2009). Joint engagement and the emergence of language in children with autism and Down syndrome. *Journal of Autism and Developmental Disorders, 39*(1), 84–96. https://doi.org/10.1007/s10803-008-0601-7

Adamson, L. B., Kaiser, A. P., Tamis-LaMonda, C. S., Owen, M. T., & Dimitrova, N. (2020). The developmental landscape of early parent-focused language intervention. *Early Childhood Research Quarterly, 50*(Pt. 1), 59–67. https://doi.org/10.1016/j.ecresq.2018.11.005

Baxendale, J., & Hesketh, A. (2003). Comparison of the effectiveness of the Hanen Parent Programme and traditional clinic therapy. *International Journal of Language & Communication Disorders, 38*(4), 397–415. https://doi.org/10.1080/1368282031000121651

Bailey, D., Duncan, G. J., Odgers, C. L., & Yu, W. (2017). Persistence and fadeout in the impacts of child and adolescent interventions. *Journal of Research on Educational Effectiveness, 10*(1), 7–39. https://doi.org/10.1080/19345747.2016.1232459

Bailey, D. H., Duncan, G. J., Cunha, F., Foorman, B. R., & Yeager, D. S. (2020). Persistence and fade-out of educational-intervention effects: Mechanisms and potential solutions. *Psychological Science in the Public Interest, 21*(2), 55–97. https://doi.org/10.1177/1529100620915848

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE consortium. (2016). CATALISE: A multinational and multidisciplinary Delphi consensus study. Identifying language impairments in children. *PLOS ONE, 11*(7), Article e0158753. https://doi.org/10.1371/journal.pone.0158753

Brian, J. A., Smith, I. M., Zwaigenbaum, L., & Bryson, S. E. (2017). Cross-site randomized control trial of the Social ABCs caregiver-mediated intervention for toddlers with autism spectrum disorder: RCT of caregiver-mediated early ASD intervention. *Autism Research, 10,* 1700–1711. https://doi.org/10.1002/aur.1818

Burchinal, M., Foster, T., Garber, K., Cohen-Vogel, L., Bratsch-Hines, M., & Peisner-Feinberg, E. (2022). Examining three hypotheses for pre-kindergarten fade-out. *Developmental Psychology, 58*(3), 453–469. https://doi.org/10.1037/dev0001302

Buschmann, A., Multhauf, B., Hasselhorn, M., & Pietz, J. (2015). Long-term effects of a parent-based language intervention on language outcomes and working memory for late-talking toddlers. *Journal of Early Intervention, 37*(3), 175–189. https://doi.org/10.1177/1053815115609384

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

Chow, J. C., Ekholm, E., & Coleman, H. (2018). Does oral language underpin the development of later behavior problems? A longitudinal meta-analysis. *School Psychology Quarterly, 33*(3), 337–349. https://doi.org/10.1037/spq0000255

Chow, J. C., Zhao, H., Sandbank, M., Bottema-Beutel, K., & Woynaroski, T. (2023). Empirically-derived effect size distributions of interventions for young children on the autism spectrum. *Journal of Clinical Child & Adolescent Psychology, 52*(2), 271–283. https://doi.org/10.1080/15374416.2021.2007485

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories, Second Edition*. Brookes.

Fuller, E. A., & Kaiser, A. P. (2020). The effects of early intervention on social communication outcomes for children with autism spectrum disorder: A meta-analysis. *Journal of Autism and Developmental Disorders, 50*(5), 1683–1700. https://doi.org/10.1007/s10803-019-03927-z

Gersten, R., Fuchs, L. S., Compton, D., Coyne, M., Greenwood, C., & Innocenti, M. S. (2005). Quality indicators for group experimental and quasi-experimental research in special education. *Exceptional Children, 71*(2), 149–164. https://doi.org/10.1177/001440290507100202

Hampton, L. H., & Kaiser, A. P. (2016). Intervention effects on spoken-language outcomes for children with autism: A systematic review and meta-analysis. *Journal of Intellectual Disability Research, 60*(5), 444–463. https://doi.org/10.1111/jir.12283

Hampton, L. H., Kaiser, A. P., & Fuller, E. A. (2020). Multi-component communication intervention for children with autism: A randomized controlled trial. *Autism, 24*(8), 2104–2116. https://doi.org/10.1177/1362361320934558

Hampton, L. H., Kaiser, A. P., & Roberts, M. Y. (2017). One-year language outcomes in toddlers with language delays: An RCT follow-up. *Pediatrics, 140*(5), e20163646. https://doi.org/10.1542/peds.2016-3646

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., Duda, S. N., & REDCap Consortium. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics, 95,* 103208. https://doi.org/10.1016/j.jbi.2019.103208

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*(2), 107–128. https://doi.org/10.3102/10769986006002107

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*(1), 39–65. https://doi.org/10.1002/jrsm.5

Heidlage, J. K., Cunningham, J. E., Kaiser, A. P., Trivette, C. M., Barton, E. E., Frey, J. R., & Roberts, M. Y. (2020). The effects of parent-implemented language interventions on child linguistic outcomes: A meta-analysis. *Early Childhood Research Quarterly, 50*(Pt. 1), 6–23. https://doi.org/10.1016/j.ecresq.2018.12.006

Higgins, J. P. T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., Savovic, J., Schulz, K. F., Weeks, L., Sterne, J. A., Cochrane Bias Methods Group, & Cochrane Statistical Methods Group. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ, 343,* d5928. https://doi.org/10.1136/bmj.d5928

Kaale, A., Fagerland, M. W., Martinsen, E. W., & Smith, L. (2014). Preschool-based social communication treatment for children with autism: 12-Month follow-up of a randomized trial. *Journal of the American Academy of Child & Adolescent Psychiatry, 53*(2), 188–198. https://doi.org/10.1016/j.jaac.2013.09.019

Kasari, C., Gulsrud, A., Freeman, S., Paparella, T., & Hellemann, G. (2012). Longitudinal follow-up of children with autism receiving targeted interventions on joint attention and play. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(5), 487–495. https://doi.org/10.1016/j.jaac.2012.02.019

Kasari, C., Gulsrud, A., Paparella, T., Hellemann, G., & Berry, K. (2015). Randomized comparative efficacy study of parent-mediated interventions for toddlers with autism. *Journal of Consulting and Clinical Psychology, 83*(3), 554–563. https://doi.org/10.1037/a0039080

Kasari, C., Siller, M., Huynh, L. N., Shih, W., Swanson, M., Hellemann, G. S., & Sugar, C. A. (2014). Randomized controlled trial of parental responsiveness intervention for toddlers at high risk for autism. *Infant Behavior and Development, 37*(4), 711–721. https://doi.org/10.1016/j.infbeh.2014.08.007

Kruythoff-Broekman, A., Wiefferink, C., Rieffe, C., & Uilenburg, N. (2019). Parent-implemented early language intervention programme for late talkers: Parental communicative behaviour change and child language outcomes at 3 and 4 years of age. *International Journal of Language & Communication Disorders, 54*(3), 451–464. https://doi.org/10.1111/1460-6984.12451

Landa, R. J., Holman, K. C., O'Neill, A. H., & Stuart, E. A. (2011). Intervention targeting development of socially synchronous engagement in toddlers with autism spectrum disorder: A randomized controlled trial: RCT of social intervention for toddlers with ASD. *Journal of Child Psychology and Psychiatry, 52*(1), 13–21. https://doi.org/10.1111/j.1469-7610.2010.02288.x

Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly, 45,* 155–176. https://doi.org/10.1016/j.ecresq.2018.03.005

McConachie, H., Randle, V., Hammal, D., & Le Couteur, A. (2005). A controlled trial of a training course for parents of children with suspected autism spectrum disorder. *The Journal of Pediatrics, 147*(3), 335–340. https://doi.org/10.1016/j.jpeds.2005.03.056

McGregor, K. K., Goffman, L., Van Horne, A. O., Hogan, T. P., & Finestack, L. H. (2020). Developmental language disorder: Applications for advocacy, research, and clinical service. *Perspectives of the ASHA Special Interest Groups, 5*(1), 38–46. https://doi.org/10.1044/2019_PERSP-19-00083

Motsch, H.-J., & Ulrich, T. (2012). Effects of the strategy therapy 'lexicon pirate' on lexical deficits in preschool age: A randomized controlled trial. *Child Language Teaching and Therapy, 28*(2), 159–175. https://doi.org/10.1177/0265659011432943

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry, 57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Odom, S. L. (2009). The tie that binds: Evidence-based practice, implementation science, and outcomes for children. *Topics in Early Childhood Special Education, 29*(1), 53–61. https://doi.org/10.1177/0271121408329171

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews, 5*(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., . . . Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372,* n71. https://doi.org/10.1136/bmj.n71

Pence Turnbull, K. L., & Justice, L. M. (2017). *Language development from theory to practice* (3rd ed.). Pearson.

Pickles, A., Le Couteur, A., Leadbitter, K., Salomone, E., Cole-Fletcher, R., Tobin, H., Gammer, I., Lowry, J., Vamvakas, G., Byford, S., Aldred, C., Slonims, V., McConachie, H., Howlin, P., Parr, J. R., Charman, T., & Green, J. (2016). Parent-mediated social communication therapy for young children with autism (PACT): Long-term follow-up of a randomised controlled trial. *The Lancet, 388*(10059), 2501–2509. https://doi.org/10.1016/S0140-6736(16)31229-6

Poslawsky, I. E., Naber, F. B., Bakermans-Kranenburg, M. J., van Daalen, E., van Engeland, H., & van IJzendoorn, M. H. (2015). Video-feedback intervention to promote positive parenting adapted to autism (VIPP-AUTI): A randomized controlled trial. *Autism, 19*(5), 588–603. https://doi.org/10.1177/1362361314537124

Presson, A. P., Partyka, G., Jensen, K. M., Devine, O. J., Rasmussen, S. A., McCabe, L. L., & McCabe, E. R. B. (2013). Current estimate of Down syndrome population prevalence in the United States. *The Journal of Pediatrics, 163*(4), 1163–1168. https://doi.org/10.1016/j.jpeds.2013.06.013

Protzko, J. (2015). The environment in raising early intelligence: A meta-analysis of the fadeout effect. *Intelligence, 53*(6045), 202–210. https://doi.org/10.1016/j.intell.2015.10.006

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with robust variance estimation: Expanding the range of working models. *Prevention Science, 23*(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Roberts, J. E., Price, J., & Malkin, C. (2007). Language and communication development in Down syndrome. *Mental Retardation and Developmental Disabilities Research Reviews, 13*(1), 26–35. https://doi.org/10.1002/mrdd.20136

Roberts, M. Y., Curtis, P. R., Sone, B. J., & Hampton, L. H. (2019). Association of parent training with child language development: A systematic review and meta-analysis. *JAMA Pediatrics, 173*(7), 671–680. https://doi.org/10.1001/jamapediatrics.2019.1197

Roberts, M. Y., & Kaiser, A. P. (2015). Early intervention for toddlers with language delays: A randomized controlled trial. *Pediatrics, 135*(4), 686–693. https://doi.org/10.1542/peds.2014-2134

Rodgers, M. A., & Pustejovsky, J. E. (2021). Evaluating meta-analytic methods to detect selective reporting in the presence of dependent effect sizes. *Psychological Methods, 26*(2), 141–160. https://doi.org/10.1037/met0000300

Romano, M., Kaiser, A., Lounds-Taylor, J., & Woods, J. (2019). Rates of prelinguistic communication and early symbol use in young children with Down syndrome: Using a progress-monitoring tool to model growth. *American Journal of Speech-Language Pathology, 29*(1), 49–62. https://doi.org/10.1044/2019_AJSLP-19-0016

Sandbank, M., Bottema-Beutel, K., Crowley, S., Cassidy, M., Dunham, K., Feldman, J. I., Crank, J., Albarran, S. A., Raj, S., Mahbub, P., & Woynaroski, T. G. (2020). Project AIM: Autism intervention meta-analysis for studies of young children. *Psychological Bulletin, 146*(1), 1–29. https://doi.org/10.1037/bul0000215

Sandbank, M., Chow, J., Bottema-Beutel, K., & Woynaroski, T. (2021). Evaluating evidence-based practice in light of the boundedness and proximity of outcomes: Capturing the scope of change. *Autism Research, 14*(8), 1536–1542. https://doi.org/10.1002/aur.2527

Schertz, H. H., Odom, S. L., Baggett, K. M., & Sideris, J. H. (2018). Mediating parent learning to promote social communication for toddlers with autism: Effects from a randomized controlled trial. *Journal of Autism and Developmental Disorders, 48*(3), 853–867. https://doi.org/10.1007/s10803-017-3386-8

Siller, M., Hutman, T., & Sigman, M. (2013). A parent-mediated intervention to increase responsive parental behaviors and child communication in children with ASD: A randomized clinical trial. *Journal of Autism and Developmental Disorders, 43*(3), 540–555. https://doi.org/10.1007/s10803-012-1584-y

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

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., . . . Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ, 366,* l4898. https://doi.org/10.1136/bmj.l4898

Tager-Flusberg, H., Edelson, L., & Luyster, R. (2011). Language and communication in autism spectrum disorders. In D. Amaral, D. Geschwind, & G. Dawson (Eds.), *Autism spectrum disorders* (pp. 172–185). Oxford University Press.

Takacs, Z. K., & Kassai, R. (2019). The efficacy of different interventions to foster children's executive function skills: A series of meta-analyses. *Psychological Bulletin, 145*(7), 653–697. https://doi.org/10.1037/bul0000195

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and spss. *Research Synthesis Methods, 5*(1), 13–30. https://doi.org/10.1002/jrsm.1091

Tipton, E. (2015). Small sample adjustments for robust variance estimation with meta-regression. *Psychological Methods, 20*(3), 375–393. https://doi.org/10.1037/met0000011

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., Smith, E., & O'Brien, M. (1997). Prevalence of specific language impairment in kindergarten children. *Journal of Speech, Language, and Hearing Research, 40*(6), 1245–1260. https://doi.org/10.1044/jslhr.4006.1245

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03

Wake, M., Levickis, P., Tobin, S., Gold, L., Ukoumunne, O. C., Goldfeld, S., Zens, N., Le, H. N. D., Law, J., & Reilly, S. (2015). Two-year outcomes of a population-based intervention for preschool language delay: An RCT. *Pediatrics, 136*(4), e838–e847. https://doi.org/10.1542/peds.2015-1337

Ward, S. (1999). An investigation into the effectiveness of an early intervention method for delayed language development in young children. *International Journal of Language & Communication Disorders, 34*(3), 243–264. https://doi.org/10.1080/136828299247405

Yoder, P. J., Lloyd, B. P., & Symons, F. J. (2018). *Observational measurement of behavior* (2nd ed.). Brookes.