The Efficacy of Treatment for Children With Developmental Speech and Language Delay/Disorder: A Meta-Analysis

James Law
Zoe Garrett
City University London

Chad Nye
University of Central Florida, Orlando

A meta-analysis was carried out of interventions for children with primary developmental speech and language delays/disorders. The data were categorized depending on the control group used in the study (no treatment, general stimulation, or routine speech and language therapy) and were considered in terms of the effects of intervention on expressive and receptive phonology, syntax, and vocabulary. The outcomes used in the analysis were dependent on the aims of the study; only the primary effects of intervention are considered in this review. These were investigated at the level of the target of therapy, measures of overall linguistic development, and broader measures of linguistic functioning taken from parent report or language samples. Thirty-six articles reporting 33 different trials were found. Of these articles, 25 provided sufficient information for use in the meta-analyses; however, only 13 of these, spanning 25 years, were considered to be sufficiently similar to be combined. The results indicated that speech and language therapy might be effective for children with phonological or expressive vocabulary difficulties. There was mixed evidence concerning the effectiveness of intervention for children with expressive syntax difficulties and little evidence available considering the effectiveness of intervention for children with receptive language difficulties. No significant differences were found between interventions administered by trained parents and those administered by clinicians. The review identified longer duration (>8 weeks) of therapy as being a potential factor in good clinical outcomes. A number of gaps in the evidence base are identified.

KEY WORDS: child language, child speech, intervention, efficacy, meta-analysis

Speech and language delay/disorder may present either as a *primary* condition, when it cannot be accounted for by any known etiology (Leonard 1998; Plante, 1998; Stark & Tallal, 1981), or as a *secondary* condition, where it can be accounted for by another primary condition such as autism, hearing impairment, general developmental difficulties, behavioral or emotional difficulties, or neurological impairment. The term *specific language impairment* is sometimes used, but exactly how specific such a condition is remains debatable given the reported comorbidity (Baker, & Cantwell, 1987; Cohen et al., 2000; Huntley, Holt, Butterfill, & Latham, 1988; Rice, Sell, & Hadley, 1991; Rutter, Mawhood, & Howlin, 1992; Stothard, Snowling, Bishop, Chipcase, & Kaplan, 1998).

Approximately 6%–8% of children may have speech and language difficulties in the preschool period (Boyle, Gillham, & Smith, 1996; Tomblin, Smith, & Zhang, 1997), of which a significant proportion will have primary speech and language delay/disorder. Primary speech and language delay/disorder is of significant concern to those involved with child development and has far-reaching implications for the child, caregiver, and school, both in terms of its immediate impact and in terms of its long-term effects. Natural history studies, in which children do not receive any specific intervention, suggest that difficulties persist (Law, Boyle, Harris, Harkness, & Nye, 2000), and the same pattern emerges for children who are provided with normal rehabilitative and educational services (Aram, Ekelman, & Nation, 1984; Bishop, & Adams, 1990; Catts, 1993; Haynes & Naidoo, 1991; Johnson et al., 1999).

Efficacy of Interventions for Speech and Language Disorder

Treatment efficacy for children with primary developmental language delay/disorder has been the subject of considerable debate in recent years and has been the focus of a number of narrative reviews of the literature, all of which conclude that intervention is effective (Enderby & Emerson, 1996; Gallagher, 1998, Goldstein & Hockenburger, 1991; Guralnick, 1988; Law, 1997; Leonard 1998; McLean & Woods Cripe, 1997; Olswang, 1998). In addition to these narrative reviews, one meta-analysis (Nye, Foster, & Seaman, 1987), one systematic review (Law, Boyle, Harris, Harkness, & Nye, 1998), and one *best evidence* review, including both group and single-subject experimental designs (Yoder & McDuffie, 2002), have been completed. These latter reviews have brought together a range of studies and methodologies and have used meta-analysis to summarize the data. These provide some support for the effectiveness of speech and language therapy, although the picture is by no means as clear as the narrative reviews tend to suggest.

There are a number of concerns with the above three analyses. Nye et al. (1987) included studies of both high and low methodological quality and therefore may have exaggerated effect sizes due to the possibility of lower quality studies producing higher effect sizes. Additionally, the effect sizes were aggregated in such a way that multiple dependent measures from a single study could be compared in tests of significance. For example, a single study could have measured both syntax and semantics as an outcome of language intervention. An effect size was calculated for both of these outcomes, averaged with similar outcomes from other studies, and analyzed for statistical differences. This type of aggregation would potentially result in a statistical bias due to the correlated nature of the outcomes taken from the same study. Law et al. (1998), while differentiating between study design and analyzing each language area separately, considered both primary and second-order effects. This may mean that effect sizes have been systematically altered because of participants' varying response to treatment depending on whether it was the area of primary difficulty. Finally, Yoder and McDuffie's (2002) review drew conclusions from a range of studies of varying methodological quality.

The Use of Meta-Analytic Techniques

Where there is controversy within the literature, it is useful to be able to synthesize the available data to provide an overall average effect of treatment. There are two methods of achieving this. First, it is possible to complete systematic reviews or research reviews that identify common themes and gaps in the literature without using statistical techniques. Second, a meta-analytic statistic can be applied to a group of studies to produce an average effect size across a range of studies that increases the power and precision of statistical procedures to provide an overall population effect. Each study included in a meta-analysis is summarized and then combined to produce an average effect size. Summary statistics are usually represented as an odds ratio, risk ratio, or hazard ratio for binary data, which provide a measure of effect based on a binary outcome of the intervention. For interventions in which the outcomes are measured on a continuous scale, the summary statistic used in the meta-analysis is either a standard mean difference (SMD) or a weighted mean difference (WMD). The SMD is used when the outcomes are measured in different ways (e.g., all studies measure language but use different tests). This allows for a standardization of values, so that results of individual studies can be combined. The WMD is used when the outcome measurements incorporate the same scale. The WMD is a computation of the weighted average of the differences in means across studies. Although meta-analysis is a technique commonly used in the intervention literature, it has been relatively slow to attract attention in the field of communication disorders (Law et al., 1998; Nye et al., 1987; Robey, 1998).

Interpretation of Effect Sizes and Confidence Intervals

Meta-analytic techniques involve the calculation of effect sizes based on either the difference in the posttest results of two groups after intervention or the difference in gains (e.g., pretest minus posttest results) between two groups of participants. A confidence interval can then

be constructed around the effect size, which shows how accurate the effect size is as an estimate of the effect size for the wider population (Gardner & Altman, 1989). If the confidence interval around the effect size is broad, it is considered a less precise estimate of what would be the effect size of the wider population.

As well as providing information regarding the precision of the estimate, the confidence interval also allows interpretation of the statistical significance of the effect size by noting whether the confidence interval (CI) crosses the zero line. If the mean effect size is 1.32 and the 95% CI is 0.45 to 2.47, the effect is said to be statistically significant at the $p > .05$ level, in favor of the treated group's performance. If the associated 95% CI includes a negative value (e.g., –0.22 to 2.34), the effect is said to be nonsignificant at the $p < .05$ level, and the conclusion would be that the intervention did not produce a statistically significant change in the treated group's performance. Should the 95% CI contain only negative values (e.g., –0.02 to –1.79), the outcome would be said to have significantly favored the control group's performance. Therefore, CIs are useful for two reasons: (a) to establish whether an estimate based on a sample would generalize to a wider population and (b) to establish whether effect sizes are statistically significant.

The Use of Randomized Studies in Meta-Analysis

Meta-analytic techniques have tended to draw heavily on the combination of effects from studies which randomize participants to experimental and control groups. Although there is by no means a consensus that interventions for speech and language difficulties are best evaluated by studies that use randomization techniques (Wertz, 2002), these studies do provide a conservative bottom-line judgment of the effectiveness of intervention both because of what they offer practitioners and policymakers (Egger, Davey Smith, & Altman, 2001) and because they give some indication of the extent to which speech and language skills are mutable in the context of environmental modification. The meta-analysis reported in this article extends the available evidence by including only data from randomized studies and combining only studies with similar aims through considering only the primary effects of intervention in each language area.

The present review was conducted to address the following questions: (a) What evidence is there that interventions can be shown to be effective when compared with untreated controls? (b) For which subgroups of children (characterized by their communication skills) has intervention been shown to be most effective? (c) To what extent does the administrator of intervention (e.g. clinicians or trained parents) or the length of intervention modify the effects of intervention?

Method

Identification of Studies

Before searches were carried out, inclusion criteria were developed based on aspects of study design, participants, intervention, and outcomes (see Table 1). It was considered necessary to maintain broad inclusion criteria relating to interventions and outcomes for two reasons: (a) The systematic review and meta-analysis needed to fully reflect the range of interventions offered to children with speech and language difficulties. (b) Studies using proxy measures (e.g., articulation measures to assess the outcomes of phonological interventions) would not be excluded, because this reflected the constraints within which many clinicians work.

Articles were then identified in two ways: through a systematic search of literature databases and through searches of existing systematic and narrative reviews. Eight databases were searched: Campbell Collaboration Social, Psychological, Education, and Criminological Trials Register; Cochrane Controlled Trials Register; Cumulative Index of Nursing and Allied Health; EMBASE; Educational Resources Information Center; MEDLINE; PsycINFO; and The National Research Register.

The database searches identified 630 papers, of which 82 were duplicates, leaving 548 papers to which the inclusion criteria were applied. A total of 277 papers were excluded because the participants were not considered to have primary speech and language difficulties, 166 papers were excluded because they did not report

Table 1. Inclusion criteria.

Design	Participants were randomly assigned either the experimental group or the control group (e.g., randomized between-subjects designs). All other study designs (i.e., case studies, case series, nonrandomized between-subjects designs, multiple baseline designs, and crossover designs) were excluded.
Participants	Children or adolescents with primary speech and language difficulties.
Intervention	To improve expressive or receptive phonology, syntax, or vocabulary.
	Did not focus on learned misarticulations, stammering, or voice.
Outcomes	Outcomes relate to speech or expressive or receptive phonology, syntax, or vocabulary.

an intervention study, and 79 were excluded on the basis that they were not randomized control trials. One study from the research register was found not to exist when the author was contacted, one author did not respond to requests to clarify information, and three studies were ongoing and could not provide data. In total, 21 papers from the database searches were judged to have met all the criteria and were included in the data set. This number was augmented by 8 papers identified from meta-analyses, 6 papers identified by the review authors, and 1 paper identified from the Campbell Collaboration trials database. This provided 36 papers, reporting 33 different studies. Of these, 13 included interventions that were considered sufficiently similar to warrant inclusion in meta-analyses.

Coding

Papers meeting the inclusion criteria were coded for information relating to the participants, the design, the intervention, and the outcomes. All coding was conducted by Zoe Garrett; Chad Nye recoded a random sample of five papers (16%). Agreement between the coders occurred for 94% of the occasions, and disagreements were resolved through discussions with James Law.

Participants

Participants were coded according to their age specified in the papers, and the severity of their difficulties was coded in terms of phonology or expressive or receptive language.

Interventions

The papers were coded according to whether the aims of intervention related to expressive or receptive phonology, morphology, or syntax. It was possible for a paper to have multiple intervention codes if the aims of the study and interventions targeted multiple language domains. Seven papers were given multiple codes, most frequently because they focused on both expressive and receptive language. If the aims of intervention were not explicit, then the paper was categorized based on the description of the intervention and on the outcomes measured. The nature of the intervention was also coded according to the administrator of the intervention, the intensity and duration of the intervention, and whether the intervention was interactive (e.g., child led) or directive (e.g., clinician led).

Outcomes

The papers were categorized according to whether their outcomes concerned expressive or receptive phonology, vocabulary, or syntax and were considered at three levels: the target of the therapy (e.g., production of the target sound); overall development, as assessed by a standardized measure of speech or language; and broader levels of functioning (e.g., percentage of consonants correct in conversation). To ensure greater homogeneity among the papers, only the primary effects of intervention were used in the analyses (e.g., only phonology interventions contributed effect sizes to the phonology meta-analyses).

Quality Assessments

While only papers reporting randomized allocation of participants are included in the present review, it is accepted that within this study type, not all studies are qualitatively comparable; thus, some assessment of the overall design quality was important to serve as a basis of consideration for inclusion for review and for data analysis (Juni, Altman, & Egger, 2001). Each of the papers was coded on a 3-point scale *(adequate, unclear,* or *inadequate)* for three aspects of methodological quality (explanation of randomization, blinding of assessors, and attrition; see Appendices A–F).

No quality assessment was made of the interventions reported in the papers, although it is accepted that the interventions may not all have been of equal quality. This decision was made for three reasons: (a) Interventions described in papers often reflect the quality of the reporting rather than the quality of the intervention. (b) Published research exists considering parameters and the impact of the quality of the study design (Juni et al., 2001), but the parameters for a *good-quality* intervention are less well defined. (c) Interventions are in part culturally defined; therefore, making judgments about the quality of interventions can be misleading.

Before combining studies in the meta-analysis, a final judgment was made considering the similarities of the techniques used in the interventions and the similarities of the responses to the interventions. Effect sizes for each study were plotted on Forest plots, and the results were visually interpreted by the authors. Where there was excessive heterogeneity (e.g., the responses to the interventions were markedly different) but the techniques used within those studies were similar, meta-analysis was completed. However, caution needs to be taken in interpreting the results. The studies cited were published over a considerable period of time and necessarily reflect clinical practice at the time of publication. Where there was excessive heterogeneity but the techniques used in the studies were different, no meta-analysis was completed. Therefore, no meta-analysis results are presented here for those studies that compared different speech and language therapy interventions, except for interventions comparing parent- and

clinician-administered interventions, or for studies comparing speech and language therapy interventions with general stimulation. A further study comparing speech and language therapy interventions with no intervention focused on language development in the 1st year of life and used the Receptive Expressive Emergent Language Scale (REEL; Bzoch & League, 1991) to provide a combined-outcome measure of receptive and expressive language. This study was judged not to be sufficiently comparable and was not combined with any other studies (Evans, in press; see list of included studies and their key characteristics in Appendices A–F).

Results

The summary table for the included studies is provided in Appendix A. Although all the included papers randomly allocated their participants to experimental and control groups, the studies were generally of low methodological quality, only 3 studies adequately described their methods of randomization, and only 14 studies stated that assessors were unaware of group allocation or that all transcripts were corrected by blinded transcribers. The studies reported low levels of attrition, although this may have been masked by nonreporting; 21 studies provided explanations of attrition with the maximum amount of attrition reported being 15%.

Of the 33 different studies identified through the searches, 8 studies could not be used in the meta-analyses because they did not report results using mean and standard deviation scores. This meant that treatment effect sizes could be calculated for only 25 studies, of which only 13 were judged to be sufficiently comparable to be combined in the meta-analyses. The meta-analysis process was completed in two stages. Initially, summary statistics were calculated for the outcomes measured in each study. This effect size was calculated as an SMD, that is, the difference in posttest means of the experimental and control group relative to the variability observed in the trial. This meant that the treatment effect was reported in units of standard deviation rather than in the units of the measurement scale used in the study. Although using the SMD made interpretation of the results more difficult, it allowed for comparison of effect sizes between meta-analyses and was necessary due to the wide range of assessments used to measure outcomes.

The second stage of the meta-analysis involved combining the individual summary statistics to create a pooled estimate of treatment effect. This was completed using a DerSimonian and Laird random effects model, which relaxes the assumption of a common treatment effect and provides a more conservative pooled estimate, which better takes into account study variations. By means of this model, studies were weighted on the basis of their inverse variance, which is related to the standard error of the study and therefore sample size (Clarke & Oxman, 2001; Deeks, Altman, & Bradburn, 2001).

A number of the studies included in the meta-analysis focused on multiple language domains, used multiple outcome measures within the same language domain, or had multiple conditions. In instances where the study focused on multiple language domains, more than one summary statistic could be calculated to represent each of the domains investigated in the study (e.g., a study investigating expressive and receptive language could provide two effect sizes, one based on the expressive language outcomes and one based on the receptive language outcomes). In instances where multiple outcome measures were used to measure the same domain, we chose one. In making this decision, standardized assessments were favored before criterion-referenced measures, which were favored over other types of measure. Where studies used multiple conditions (e.g., a parent group, a clinician group, and a control group), the clinician and parent groups were pooled in the overall meta-analyses. The treatment groups were then compared separately in additional analyses for effects of parent versus clinician intervention. In all instances, the above procedures were completed to maintain study independence; only one summary statistic from each study was used in calculating the combined treatment effect. All data were handled and analyzed in Review Manager software provided by the Cochrane Collaboration, using guidelines developed by the Cochrane Collaboration (Clarke & Oxman, 2001).

Analyses

Six main analyses were carried out, which combined all the studies that shared a focus on expressive or receptive phonology, vocabulary, or syntax in comparison with a no-treatment or delayed-treatment control group. Following this, secondary analyses were completed by removing from the main analyses studies with the following characteristics: (a) parent-administered intervention (see Table 2) and (b) duration of less than 8 weeks (see Table 3). During the coding process, it became apparent that the majority of the studies included in the analyses excluded children with severe receptive language difficulties from participating in the intervention programs. Therefore, a third secondary analysis was completed by removing studies with the following characteristic: (c) the explicit criterion that children have a severe receptive language difficulty to be included in the intervention (see Table 4).

The rationale for this secondary analysis was that severe receptive language difficulties may alter the effects of intervention. While it was not possible to remove all the children with receptive language difficulties from

Table 2. Average effect size of receptive and expressive language interventions compared with the no-treatment condition, according to overall measures of language development.

	Combined parent/clinician treatment			Clinician treatment only			Parent treatment only		
Area of intervention	*d*	95% CI	No. studies	*d*	95% CI	No. studies	*d*	95% CI	No. studies
Expressive phonology	0.44	0.01–0.86	6*	0.67	0.19–1.16	5*	–0.17	–0.72–0.39	2
Expressive vocabulary	0.98	–0.59–2.56	2	0.13	–0.65–0.91	1	1.06	–0.14–2.52	2
Expressive syntax	0.70	–0.14–1.55	5	0.28	–0.19–0.75	4	0.83	–0.96–2.63	3
Receptive phonology	0.53	–0.10–1.16	1	NC			0.53	–0.10–1.16	1
Receptive vocabulary	NC			NC			NC		
Receptive syntax	–0.04	–0.64–0.56	2	0.01	–0.53–0.56	2	–0.53	–1.41–0.34	1

Note. NC = not calculable.

*$p < .05$.

the analyses, it was considered that this might provide insight into the effects of expressive language intervention on this client group. Table 4 shows the numbers of studies that contributed to each meta-analysis.

Expressive Syntax Outcomes

The effect estimate, when measured using assessments of overall syntactic ability, did not show a significant difference between speech and language interventions and no treatment ($d = 0.70$, $n = 271$, CI = –0.14–1.55). The effect size decreased when only data from clinician studies were included ($d = 0.28$, $n = 214$, CI = –0.19–0.75), although this effect size increased when studies of less than 8 weeks were removed ($d = 0.43$, $n = 187$, CI = –0.06–0.93). When data were excluded from studies that explicitly involved only children with severe receptive difficulties, the effect estimate significantly favored the use of speech and language therapy ($d = 1.02$, $n = 233$, CI = 0.04–2.01). The results from studies comparing different approaches to parent involvement in intervention were varied. Direct comparisons of parent-delivered interventions with those in which a speech and language clinician had been responsible for the program delivery did not show any significant differences ($d = –0.04$, $n = 66$, CI = –0.56–0.48). The Forest plot of the expressive syntax outcomes is given in Figure 1.

Table 3. Clinician data for trials longer than 8 weeks' duration.

Area of intervention	Effect size	Confidence interval	No. studies
Expressive phonology	0.74	0.14–1.33	4*
Expressive vocabulary	NC		
Expressive syntax	0.43	–0.06–0.93	3
Receptive phonology	NC		
Receptive vocabulary	NC		
Receptive syntax	0.19	–0.12–0.51	1

*$p < .05$.

There were no significant differences between speech and language interventions and no treatment for the impact of speech and language therapy, when measured by total number of utterances ($d = 0.68$, $n = 99$, CI = –0.45–1.82); mean length of utterance ($d = 0.74$, $n = 95$, CI –0.33–1.81); and parental report of phrase complexity ($d = 1.02$, $n = 99$, CI = –0.17–2.22). When data were excluded from studies that explicitly included only children with severe receptive language difficulties, the effect estimates significantly favored expressive language intervention when compared with no treatment for total number of utterances ($d = 1.20$, $n = 61$, CI = 0.33–2.07); mean length of utterance ($d = 1.28$, $n = 57$, CI = 0.66–1.89); and parent report of phrase complexity ($d = 1.54$, $n = 61$, CI = 0.42–2.65). Direct comparisons of parent- and clinician-administered intervention did not show any significant differences for total utterances ($d = 0.15$, $n = 45$, CI = –0.45–0.74); mean length of utterance ($d = 0.28$, $n = 45$, CI = –1.41–1.96); and parental report of sentence complexity ($d = 0.01$, $n = 45$, CI = –0.63–0.66).

Expressive Phonology Outcomes

The overall effect estimate when measured using standardized assessment measures significantly favored the use of phonology interventions when compared with no treatment ($d = 0.44$, $n = 264$, CI = 0.01–0.86); the effect size increased when parent-administered treatments were removed ($d = 0.67$, $n = 214$, CI = 0.19–1.16) and increased again when interventions lasting less than 8 weeks were removed ($d = 0.74$, $n = 213$, CI = 0.14–1.33). Parent-administered interventions based on receptive auditory techniques (e.g., auditory bombardment or auditory discrimination therapy) did not significantly effect productive phonology ($d = –0.17$, n

Table 4. Number of studies contributing to each meta-analysis.

Area of intervention	No. studies in primary analysis	No. studies in secondary analysis: Clinician-administered interventions only	Studies of duration >8 weeks	Studies without receptive difficulties
Expressive phonology	6	5	4	6
Receptive phonology	1	0	0	1
Expressive vocabulary	2	1	0	1
Receptive vocabulary	0	0	0	0
Expressive syntax	5	4	4	4
Receptive syntax	2	2	1	1

Note. *Studies without receptive difficulties* refers to studies remaining after removal of those studies including only children with severe receptive language difficulties.

= 40, CI = –0.72–0.39). The Forest plot of the expressive phonology outcomes is given in Figure 2.

There was a large significant effect favoring speech interventions compared with no treatment, when measured by the percentage of consonants correct in conversation (d = 1.91, n = 26, CI = 0.96–2.86), but a nonsignificant effect for the retelling of a story with target consonants (d = 1.29, n = 11, CI = –0.11–2.69).

Expressive Vocabulary Outcomes

The use of standardized assessments to measure vocabulary growth following expressive language interventions showed a nonsignificant effect for expressive language intervention when compared with no therapy (d = 0.98, n = 74, CI = –0.59–2.56). When data were excluded from studies that included only children with severe receptive language difficulties, a significant effect was found (d = 1.79, n = 36, CI = 1.01–2.58). Direct comparisons of parent- and clinician-administered interventions did not show any significant differences (d = 0.20, n = 45, CI = –0.40–0.79). There was a significant effect in favor of the use of expressive language interventions when compared with no intervention, measured as number of words in a language sample (d = 1.08, n = 82, CI = 0.61–1.55) and parental report of vocabulary (d = 0.89, n = 136, CI = 0.21–1.56). When data were excluded from studies in which all children had expressive and receptive difficulties, the results remained the same, as measured by parent report of vocabulary (d = 1.00, n = 98, CI = 0.16–1.84). Direct comparisons of parent- and clinician-administered interventions did not show any significant differences for either parental report of vocabulary (d = –0.16, n = 45, CI = –0.76–0.44) or number of words in a language sample (d = –0.50, n = 17, CI = –1.48–0.47). The Forest plot for the expressive vocabulary outcomes is given in Figure 3.

Receptive Phonology Outcomes

The use of parent-administered receptive phonology interventions, either as auditory discrimination tasks or more general reading and talking, as a means of improving receptive phonology was not shown to be significantly effective when compared with no treatment (d = 0.53, n = 45, CI = –0.10–1.16). The Forest plot for the receptive phonology outcomes is provided in Figure 4.

Receptive Syntax Outcomes

The effect estimate did not show any significant differences between speech and language interventions versus no treatment (d = –0.04, n = 193, CI = –0.64–0.56), according to overall measures of receptive syntax. Removal of data from parent interventions (d = 0.01, n = 182, CI = –0.53–0.55) and from trials of less than 6 weeks' duration (d = 0.19, n=155, CI = –0.12–0.51) did not alter the results. Direct comparisons of clinician- and parent-administered interventions did not show any significant differences (d = –0.11, n = 28, CI = –0.87–0.65). The Forest plot for the receptive syntax outcomes is given in Figure 5.

Discussion

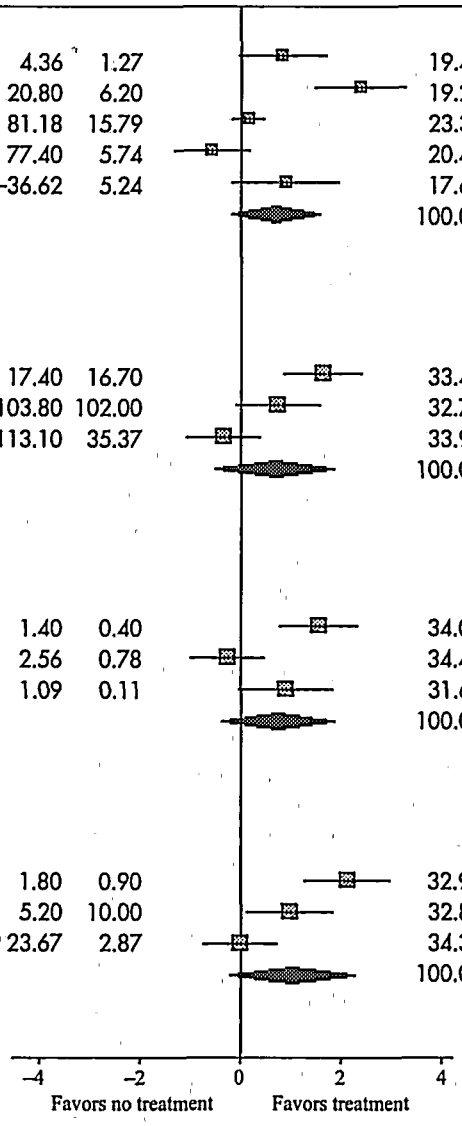
The results of the main analyses show a significant effect of intervention when children have phonological difficulties or expressive vocabulary difficulties. There is currently mixed evidence considering the effects of intervention for children with expressive syntax difficulties, and the results for children with receptive phonology or language difficulties are inconclusive due to the limited number of studies that have been carried out.

The results of secondary analyses show the following: (a) no significant differences between the use of

Figure 1. Expressive syntax outcomes (SMD = standard mean difference; CI = confidence interval).

Comparison: Speech and language intervention versus delayed or no treatment
Outcome: Expressive syntax outcomes

Study	Treatment n	Treatment M	Treatment SD	Control n	Control M	Control SD	SMD (95% CI random)	Weight (%)	SMD	95% CI random
Measures of overall expressive syntax development										
Fey et al., 1993	21	5.66	1.58	8	4.36	1.27		19.4	0.84	−0.01–1.69
Gibbard, 1994	18	38.70	8.60	18	20.80	6.20		19.2	2.33	1.47–3.20
Glogowska et al., 2000	70	83.87	15.13	82	81.18	15.79		23.3	0.17	−0.15–0.49
Law et al., 1999	28	74.74	4.71	10	77.40	5.74		20.4	−0.52	−1.25–0.21
Matheny & Panagos, 1978	8	−30.62	7.27	8	−36.62	5.24		17.6	0.90	−0.15–1.94
Subtotal (95% CI)	145			126				100.0	0.70	−0.14–1.55
Test for heterogeneity, $\chi^2(4, N = 140) = 28.99$, $p < .00001$										
Test for overall effect, $z = 1.64$, $p = .10$										
Total number of utterances in a language sample										
Gibbard, 1994	18	89.50	58.80	18	17.40	16.70		33.4	1.63	0.87–2.40
Girolametto et al., 1996b	12	182.90	103.00	13	103.80	102.00		32.7	0.75	−0.07–1.56
Law et al., 1999	28	102.94	31.42	10	113.10	35.37		33.9	−0.31	−1.03–0.42
Subtotal (95% CI)	58			41				100.0	0.68	−0.45–1.82
Test for heterogeneity, $\chi^2(2, N = 55) = 13.03$, $p = .0015$										
Test for overall effect, $z = 1.18$, $p = .2$										
Mean length of utterance from language sample										
Gibbard, 1994	18	2.30	0.70	18	1.40	0.40		34.0	1.54	0.79–2.30
Law et al., 1999	28	2.42	0.62	10	2.56	0.78		34.4	−0.21	−0.93–0.52
Robertson & Ellis Weismer, '99	11	1.32	0.32	10	1.09	0.11		31.6	0.90	0.00–1.81
Subtotal (95% CI)	57			38				100.0	0.74	−0.33–1.81
Test for homogeneity, $\chi^2(2, N = 54) = 11.02$, $p = .004$										
Test for overall effect, $z = 1.35$, $p = .18$										
Parent report of phrase complexity										
Gibbard, 1994	18	5.00	1.90	18	1.80	0.90		32.9	2.10	1.27–2.94
Girolametto et al., 1996b	12	16.70	13.00	13	5.20	10.00		32.8	0.96	0.13–1.80
Law et al., 1999	28	23.80	3.72	10	23.67	2.87		34.3	0.04	−0.69–0.76
Subtotal (95% CI)	58			41				100.0	1.02	−0.17–2.22
Test for heterogeneity, $\chi^2(2, N = 55) = 13.55$, $p = .0011$										
Test for overall effect, $z = 1.67$, $p = .09$										



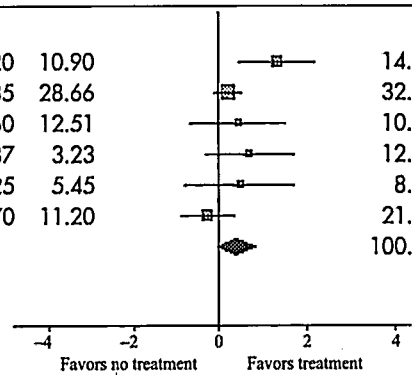
trained parent administrators and clinician-administered interventions, (b) possible indications that interventions of longer duration (>8 weeks) may be more effective than those over a shorter duration, (c) that there may be a differential effect of intervention for expressive syntax, with intervention being effective for those children who do not also have receptive language difficulties.

Parent interventions for expressive language difficulties have focused on child-led interactive approaches to intervention, such as the Hanen Early Language Parent Programme (Manolson, 1995). The effect sizes, when directly compared with clinician-implemented interventions, show very little difference, which may reflect the fact that the approach is well described in the literature, training programs are offered, and a relatively large number of studies have been carried out investigating the approach. The parent-intervention techniques in the phonology studies do not show significant differences when directly compared with clinician-administered intervention but are associated with much broader confidence intervals than those of expressive language interventions. This may reflect the fact that fewer studies have been carried out in this area and that intervention approaches tend to be more eclectic. The parent approaches used in the studies in this review have focused on parents implementing auditory discrimination

Figure 2. Expressive phonology outcomes.

Comparison: Speech and language intervention versus delayed or no treatment
Outcome: Expressive phonology outcomes

Study	Treatment n	Treatment M	Treatment SD	Control n	Control M	Control SD	SMD (95% CI random)	Weight (%)	SMD	95% CI random
Measures of overall phonological development										
Almost & Rosenbaum, 1998	13	−34.70	7.90	13	−48.20	10.90		14.9	1.37	0.50–2.24
Glogowska et al., 2000	70	−27.20	22.76	81	−34.35	28.66		32.0	0.27	−0.05–0.59
Lancaster, 1991	10	−36.59	19.17	5	−45.60	12.51		10.9	0.49	−0.61–1.58
Matheny & Panagos, 1978	8	−6.62	2.39	8	−8.87	3.23		12.0	0.75	−0.28–1.77
Munro, 1999	7	75.14	14.14	4	68.25	5.45		8.8	0.53	−0.73–1.79
Shelton et al., 1978	30	7.71	8.18	15	9.70	11.20		21.3	−0.21	−0.83–0.41
Subtotal (95% CI)	138			126				100.0	0.44	0.01–0.86
Test for heterogeneity, $\chi^2(5, N = 132) = 9.38$, $p = .095$										
Test for overall effect, $z = 2.01$, $p = .04$										



techniques and auditory bombardment techniques, while those implemented by the clinician have focused more on expressive production interventions. This may, in part, explain the nonsignificant findings of the receptive phonology studies, as both these studies used parent-implemented techniques (see Figure 6).

Interventions of duration longer than 8 weeks are suggested to be more effective than those of less than 8 weeks. Unfortunately, the role of intensity could not be calculated because too few trials offered intensities of therapy over 2 hr a week. However, those of longer duration did not necessarily also offer higher intensities or more hours of contact time than those of less than 8 weeks. This result can only be considered tentatively due to the range of studies included. Other reviews and trials have found different results; Nye et al. (1987) found the highest effect size to be among interventions lasting 4 to 12 weeks, and Fey, Cleave, and Long (1997) found that gains during a second 4.5-month treatment block were much smaller than those seen in the first 4.5-month treatment block.

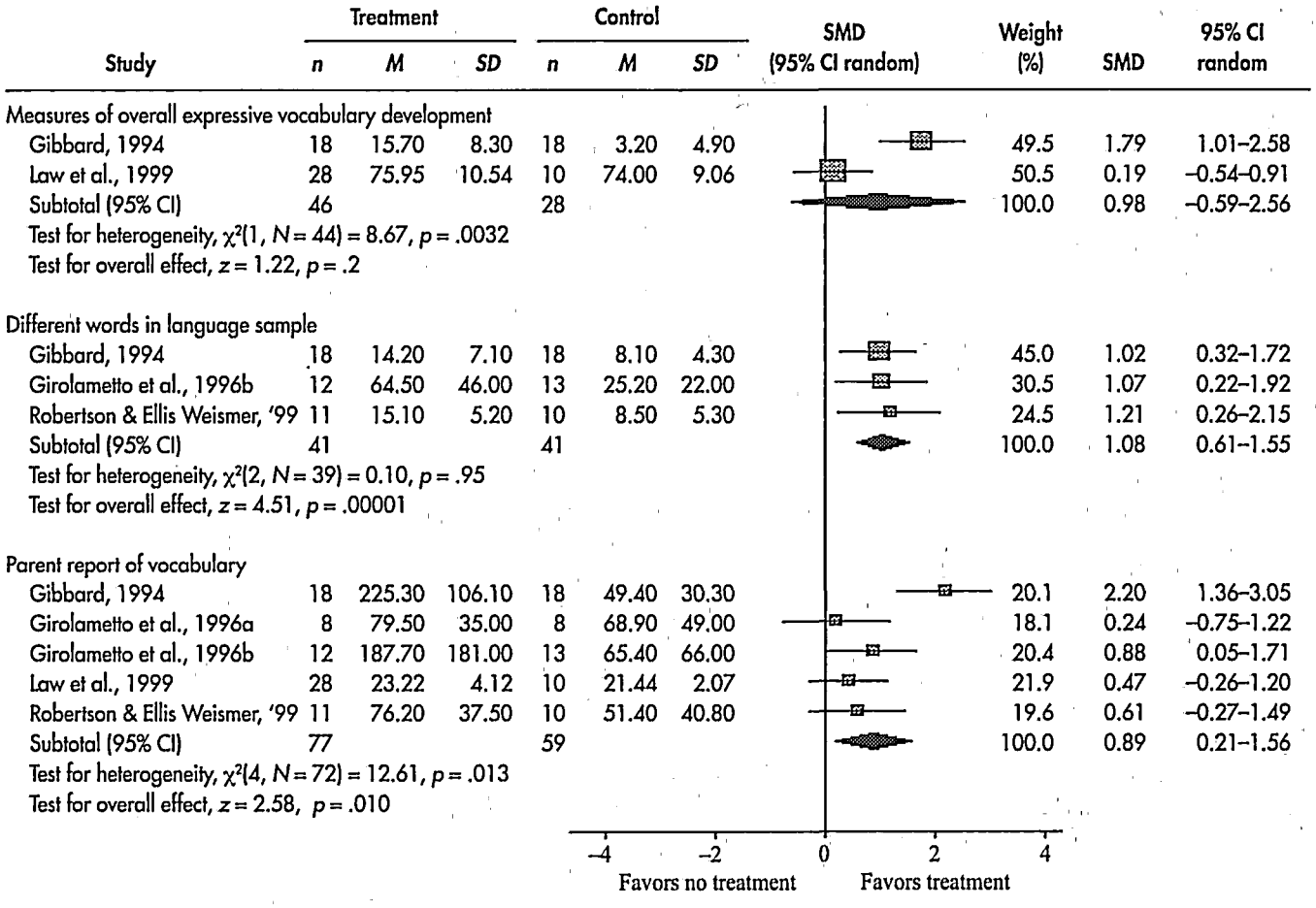
The results of this review suggest that there is a differential effect of intervention, with intervention being effective for those children who do not also have receptive language difficulties. This finding is suggestive only due to the nature of the correlational data on which it is based, but it reflects what is known about the long-term prognosis of this client group and would appear to have some biological plausibility. The finding that many children with receptive language difficulties are excluded from intervention studies is very important. The fact that these children are most likely to have long-term difficulties heightens the need for good-quality intervention studies, so that clinicians know how best to intervene in these cases.

The effect sizes reported in these meta-analyses are generally similar to the results reported in previous meta-analyses (Law et al., 1998; Nye et al., 1987), although the effect sizes for speech and receptive language outcomes are smaller. There are a number of potential reasons for this. The methodology of these meta-analyses differs from previous meta-analyses because the latter only consider the primary effects of intervention. A small number of the studies included in this review had large effect sizes for receptive language outcomes, but the children in these studies had not been identified on the basis of standardized assessment measures as having receptive language difficulties. These effects were not included in the meta-analysis, and therefore, the effect-size values are a more refined quantitative summary of the studies.

Limitations of the Review

In a number of areas, greater clarity would lead to a refinement of the review process in the domain of speech and language delay/disorder in the future. The first concerns the risk of detecting significance when it is not really there: a Type I error. The relatively low methodological quality of some of the studies given the prescribed criteria makes this a potential problem. Small sample sizes and lack of blinding inevitably increase the chances of a Type I error. There is little to be done about this at this stage, given that the studies met the other inclusion criteria, but it is a signpost for the development of future studies in this field. Similarly, there is a risk of an inflated Type I error rate from using the number of participants rather than the number of studies as the basis for computing the standard error rate.

Figure 3. Expressive vocabulary outcomes.



The second problem area concerns the nature of the interventions described. In most cases, too little information is provided about the interventions to truly replicate them; and this inevitably raises the question of whether it is appropriate to review interventions evaluated 30 years or more ago. If these interventions are outdated, is it appropriate to review them at all? The study selection process does not make it possible to eliminate studies based on the type of intervention alone. The lack of data also raises the question of whether it is really appropriate to compare some interventions directly. There are some interventions (e.g., focused stimulation)

Figure 4. Receptive phonology outcomes.

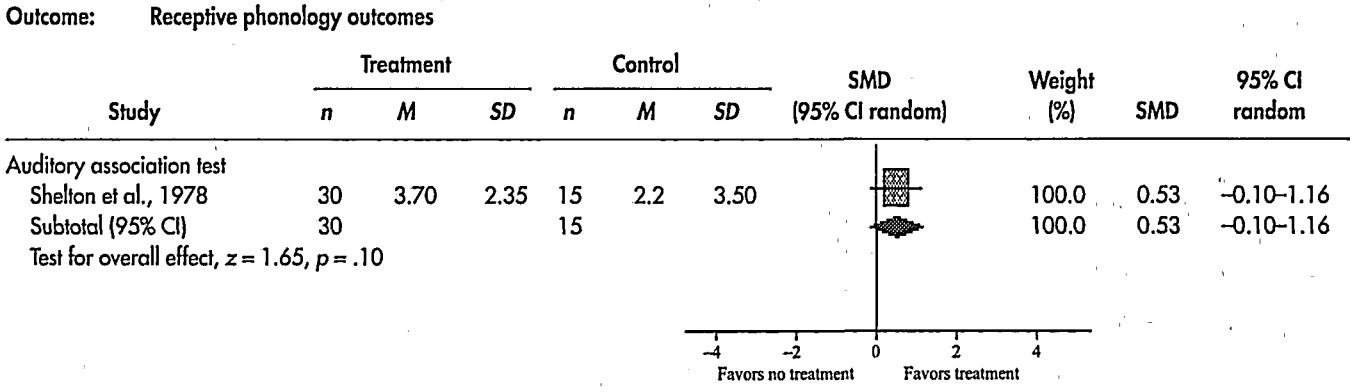
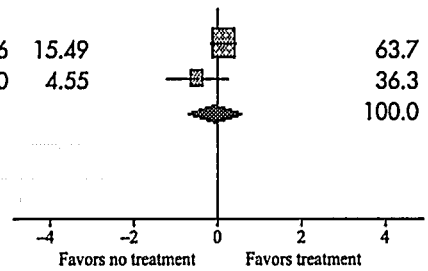


Figure 5. Receptive syntax outcomes.

Comparison: Speech and language intervention versus delayed or no treatment
Outcome: Receptive syntax outcomes

Study	Treatment n	Treatment M	Treatment SD	Control n	Control M	Control SD	SMD (95% CI random)	Weight (%)	SMD	95% CI random
Measures of overall receptive syntax development										
Glogowska et al., 2000	71	87.30	15.89	84	84.26	15.49		63.7	0.19	−0.12–0.51
Law et al., 1999	28	71.05	5.32	10	73.40	4.55		36.3	−0.45	−1.18–0.28
Subtotal (95% CI)	99			94				100.0	−0.04	−0.64–0.56
Test for heterogeneity, $\chi^2(1, N = 97) = 2.49$, $p = .11$										
Test for overall effect, $z = 0.13$, $p = .9$										

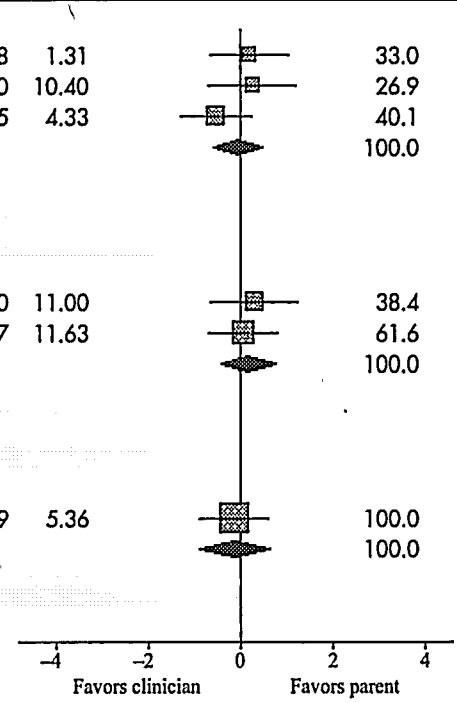


where this seems to be less of a problem, but in an ideal world, one would wish to combine only those evaluation studies with identical outcomes and identical interventions. The lack of specificity in some of the descriptions and the lack of benchmarks as to what are normal interventions in this area beg the question of whether the studies concerned are *efficacy* studies (i.e., the result of an optimum level of intervention) or *effectiveness* studies (i.e., reflecting current practice). For example the Glogowska, Roulstone, Enderby, and Peters (2000) study found that the average intervention time was six sessions. To what extent is it possible to generalize the findings of what would purport to be an effectiveness study to other contexts in which this level of input was not the norm? How relevant is an efficacy study that is optimally staffed but does not reflect current practice?

Figure 6. Language outcomes for parent and clinician interventions.

Comparison: Speech and language interventions versus traditional speech and language programs
Outcome: Subgroup analysis (clinician vs. parent)

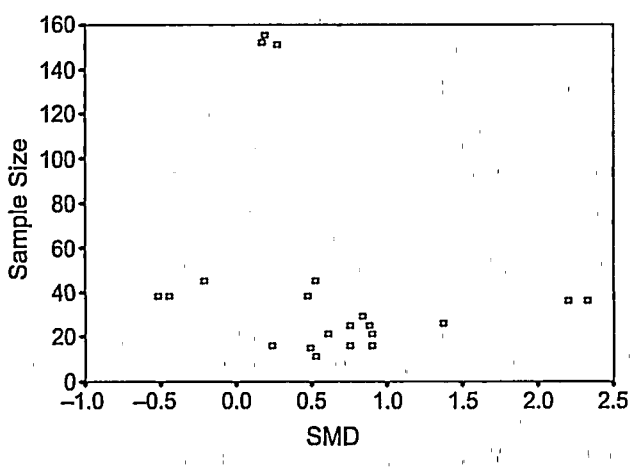
Study	Treatment n	Treatment M	Treatment SD	Control n	Control M	Control SD	SMD (95% CI random)	Weight (%)	SMD	95% CI random
Measures of overall expressive syntax development										
Fey et al., 1993	10	5.85	1.83	11	5.48	1.31		33.0	0.23	−0.63–1.08
Gibbard, 1994	9	34.60	8.20	8	31.50	10.40		26.9	0.32	−0.64–1.28
Law et al., 1999	11	73.27	5.27	17	75.65	4.33		40.1	−0.49	−1.26–0.28
Subtotal (95% CI)	30			36				100.0	−0.04	−0.56–0.48
Test for heterogeneity, $\chi^2(2, N = 27) = 2.21$, $p = .33$										
Test for overall effect, $z = 0.14$, $p = .9$										
Measures of overall expressive vocabulary development										
Gibbard, 1994	9	20.60	12.20	8	16.50	11.00		38.4	0.33	−0.63–1.29
Law et al., 1999	11	76.70	8.51	17	75.47	11.63		61.6	0.11	−0.65–0.87
Subtotal (95% CI)	20			25				100.0	0.20	−0.40–0.79
Test for heterogeneity, $\chi^2(1, N = 18) = 0.12$, $p = .72$										
Test for overall effect, $z = 0.65$, $p = .5$										
Measures of overall receptive syntax development										
Law et al., 1999	11	70.67	5.25	17	71.29	5.36		100.0	−0.11	−0.87–0.65
Subtotal (95% CI)	11			17				100.0	−0.11	−0.87–0.65
Test for overall effect, $z = 0.29$, $p = .8$										



Validity of the Meta-Analysis

There are two concerns with meta-analyses. The first is that a meta-analysis result is only as good as the studies that go into it; the second is that any meta-analysis may suffer from bias associated with a *bottom drawer* effect, which occurs with the use of analyses that systematically exclude studies with negative results or results with only small effect sizes that have not been submitted for publication. To analyze the possible impact of study quality, sensitivity analyses were carried out on the basis of the quality codes that had been assigned to the studies to assess methodological quality. All the studies reported that they were randomized, but only three studies explained their methods, meaning that a sensitivity analysis could not be carried out and it could not be determined that studies reporting that they were randomized were so both in terms of the generation and the concealment of allocation sequences. Sensitivity analyses could be completed that investigated the impact of nonreporting of attrition and blinding in effect estimates of interventions compared with no treatment, as measured by overall measures of development. Removing studies that did not report attrition did not affect the results (phonology outcomes $d = 0.40$, $n = 5$, CI = –0.08–0.89; expressive syntax $d = 0.67$, $n = 4$, CI = –0.33–0.66), although the confidence intervals became broader once the studies were removed. This suggests that attrition was not systematically altering the effect sizes. Removing studies that did not report blinding proved inconclusive. The effect for phonology interventions increased ($d = 0.66$, $n = 3$, –0.07–1.40), but the effect for expressive interventions decreased (syntax $d = 0.14$, $n = 3$, CI = –0.47–0.75; vocabulary $d = 0.19$, $n = 1$, CI = –0.54–0.91). This was likely to reflect the small number of studies that were included in the sensitivity analyses rather than any definite effect of study quality.

Figure 7. Funnel plot of SMDs against sample size.



The impact of bias in the averages in effect sizes was assessed by means of funnel plots (see Figure 7). A plot of the sample size over the effect size would show a symmetrical distribution with a characteristic funnel at the top if there was no associated bias. A funnel plot was completed for the 13 trials included in the meta-analysis based on their effect sizes from outcomes measured using standardized assessments, or broader outcome measures. In all, 21 outcome measures were used for the funnel plot across the 13 studies. The plot showed symmetry on either side, suggesting that there was no publication bias present. However, there did appear to be a small sample-size bias as shown by the large numbers of studies in the bottom half of the plot.

Further Research

The most substantial single gap in the literature revealed by this review is the lack of good-quality literature about intervention for children with severe receptive language difficulties. Given the relative risk of long-term difficulties associated with these conditions, this is a matter of priority for the field.

The sensitivity analyses and the methodological quality codes show that there are a number of issues with nonreporting even within trials that are considered to be of highest methodological quality. It is recognized that elements of study quality (e.g., blinding, attrition, randomization) can have important effects on the results through exaggerating effect sizes; therefore, reporting of study methods needs close consideration in articles. In addition, participants and treatments were not always described in sufficient detail to make a clear interpretation of the results and the potential effects of other important linguistic factors, like receptive language and nonlinguistic factors, such as behavior, attention, developmental history, and socioeconomic status. This is important, as current theory highlights the multifactorial nature of speech and language disorders. There is a need then for consistent reporting in treatment efficacy studies of a type recommended in the CONSORT statement (Moher, Schulz, & Altman, 2001).

In all, 16 studies in the full review compared either an intervention approach with no intervention or multiple intervention approaches with no intervention. Of these, 3 did not have usable outcome measures, and thus, 13 studies were included in the analysis. While comparisons involving no-treatment groups may not help clinicians consider what is the most effective means of delivering therapy, these trials are important because they help to develop a measure for the average effectiveness of speech and language therapy. Once this figure has been reached, trials comparing different therapy approaches can be used to improve on the average treatment effect.

Meta-analyses that have been completed in the last 15 years show a considerable range in their average effect size, and the confidence intervals are correspondingly broad. This necessarily affects the certainty that can be placed in the consistency of the results. Therefore, there is an ongoing need for trials that use no-treatment or delayed-treatment conditions. The most apparent area of need is receptive language difficulties, although there is still wide variation in the results of expressive language studies.

Clinical trials also need to be accompanied by epidemiological studies, such as cohort studies, or case-control studies, to investigate the factors that lead to better therapeutic outcomes in terms of participant characteristics like severity and age and intervention characteristics such as duration and intensity. The use of population data would facilitate the investigation of a broader range of factors that may be correlated with better outcomes, which could then be assessed using intervention studies that are more vigorously controlled. This would help direct resources to comparisons, most likely showing significant differences, and build on the general average developed from no-treatment comparisons. The information gained from these studies could help explain results from existing intervention studies in terms of potentially confounding variables and effect modifiers.

Acknowledgments

The review was funded by Nuffield Foundation Grant EDU/00326/9. Thanks go to Jane Dennis and Jo Abbott from the Developmental, Behavioural, and Psychosocial Group of the Cochrane Collaboration.

References

Almost, D., & Rosenbaum, P. (1998). Effectiveness of speech intervention for phonological disorders: A randomized controlled trial. *Developmental Medicine and Child Neurology, 40,* 319–325.

Aram, D., Ekelman, B., & Nation, J. (1984). Preschoolers with language disorders: 10 years later. *Journal of Speech and Hearing Research, 27,* 232–244.

Baker, L., & Cantwell, D. P. (1987). A prospective psychiatric follow-up of children with speech/language disorders. *Journal of the American Academy of Child and Adolescent Psychiatry, 26,* 546–553.

Bishop, D., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonology and reading retardation. *Journal of Child Psychology and Psychiatry, 31,* 1027–1050.

Boyle, J., Gillham, B., & Smith, N. (1996). Screening for early language delay in the 18-36 month age-range: The predictive validity of tests of production and implications for practice. *Child Language Teaching and Therapy, 12,* 113–127.

Bzoch, K. & League, R. (1991). *The Receptive Expressive Emergent Language Scale.* Gainesville, FL: Language Education Division, Computer Management Corp.

Catts, H.W. (1993). The relationship between speech-language impairments and reading disabilities. *Journal of Speech and Hearing Research, 36,* 948–958.

Clarke, M., & Oxman, A. D. (2001). *Cochrane Reviewers handbook* (Vol. 4.1.2, 2nd ed.). Oxford, England: Update Software.

Cohen, N. J., Vallance, D. D., Barwick, M., Im, N., Menna, R., Horodezjy, N. B., et al. (2000). The interface between ADHD and language impairment: An examination of language, achievement and cognitive processing. *Journal of Child Psychology and Psychiatry, 41,* 353–362.

Deeks, J. J., Altman, D. G., & Bradburn, M .J. (2001). Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In M. Egger, G. Davey Smith, & D. G. Altman (Eds.), *Systematic reviews in health care: Meta-analysis in context* (pp. 285–312). London: British Medical Journal.

Egger, M., Davey Smith, G., & Altman, D. (2001). *Systematic reviews in health care: Meta-analysis in context.* London: British Medical Journal.

Enderby, P., & Emerson, J. (1996). Speech and language therapy—Does it work? *British Journal of Medicine, 312,* 1655–1658.

Evans, C. (in press). *The Kenilworth Project: A Randomised Control Trial of WILSTAARs.* Unpublished report.

Fey, M. E., Cleave, P. L., & Long, S. H. (1997). Two models of grammar facilitation in children with language impairments: Phase 2. *Journal of Speech, Language, and Hearing Research, 40,* 5–19.

Fey, M. E., Cleave, P. L., Long, S. H., & Hughes, D. L. (1993). Two approaches to the facilitation of grammar in children with language impairment: An experimental evaluation. *Journal of Speech and Hearing Research, 36,* 141–157.

Gallagher, T. (1998) Treatment research in speech, language and swallowing: Lessons from child language disorders *Folia Phoniatrica Logopaedica, 50,* 165–182.

Gardner, M. J., & Altman, D. G. (1989). *Statistics with confidence–Confidence intervals and statistical guidelines.* London: British Medical Journal.

Gibbard, D. (1994). Parental-based intervention with preschool language-delayed children (Study 2). *European Journal of Disorders of Communication, 29,* 131–150.

Girolametto, L., Pearce, P. S., & Weitzman, E. (1996a). Interactive focused stimulation for promoting vocabulary in young children with delays: A pilot study. *Journal of Child Communication Development, 17,* 39–49.

Girolametto, L., Pearce, P. S., & Weitzman, E. (1996b). Interactive focused stimulation for toddlers with expressive vocabulary delays. *Journal of Speech, Language, and Hearing Research, 39,* 1274–1283.

Glogowska, M., Roulstone, S., Enderby, P., & Peters, T. J. (2000). Randomised controlled trial of community based speech and language therapy in preschool children. *British Medical Journal, 321,* 923–926.

Goldstein, H., & Hockenburger, E. (1991). Significant

progress in child language intervention: An 11 year retrospective. *Research Into Developmental Disabilities 12,* 401–424.

Guralnick, M. J. (1988). Efficacy in early childhood intervention programs. In S. J. Odom, & M. B. Karnes (Eds.), *Early intervention for infants and children with handicaps* (pp. 63–73). Baltimore: Brookes.

Haynes, C., & Naidoo, S. (1991). *Children with specific speech and language impairment.* Oxford, England: Blackwell.

Huntley, R. M. C., Holt, K., Butterfill, A., & Latham, C. (1988). A follow-up study of a language intervention programme. *British Journal of Disorders of Communication, 23,* 127–140.

Johnson, C. J., Beitchman, J. H., Young, A., Escobar, M., Atkinson, L., Wilson, B., et al. (1999). Fourteen-year follow-up of children with and without speech/language impairments: Speech/language stability and outcomes. *Journal of Speech, Language, and Hearing Research, 42,* 744–761.

Juni, P., Altman, D. G., & Egger, M. (2001). Assessing the quality of randomised control trials. In M. Egger, G. Davey Smith, & D. G. Altman (Eds.), *Systematic reviews in health care: Meta-analysis in context* (pp. 87–108). London: British Medical Journal.

Lancaster, G. (1991). *The effectiveness of parent administered input training for children with phonological disorders.* Unpublished master's thesis, City University London.

Law, J. (1997). Evaluating intervention for language impaired children: A review of the literature. *European Journal of Disorders of Communication, 32,* 1–14.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (1998). Screening for speech and language delay: A systematic review of the literature. *Health Technology Assessment, 2,* 1–184.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and natural history of primary speech and language delay: Findings from a recent systematic review of the literature. *International Journal of Language and Communication Disorders, 35,* 165–188.

Law, J., Kot, A., & Barnett, G. (1999). *A comparison of two methods for providing intervention to three year old children with expressive/receptive language impairment.* London: City University London.

Leonard, L. (1998). The nature and efficacy of treatment. In L. B. Leonard (Ed.), *Children with specific language impairment* (pp. 193–210). Cambridge, MA: MIT Press.

Manolson, A. (1995). *You make the difference: In helping your child to learn.* Toronto, Ontario, Canada: The Hanen Centre.

Matheny, N., & Panagos, J. M. (1978). Comparing the effects of articulation and syntax programmes on syntax and articulation improvement. *Language, Speech, and Hearing Services in Schools, 9,* 57–61.

Mclean, L. K., & Woods Cripe, J. W. (1997). The effectiveness of early intervention for children with communication disorders. In M. J. Guralnick (Ed.), *The effectiveness of early intervention* (pp. 349–429). Baltimore: Brookes.

Moher, D., Schulz, K. F., & Altman, D. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomised trials. *Journal of the American Medical Association, 285,* 1987–1991.

Munro, J. (1999). *A study of the efficacy of speech and language therapy for particular speech sounds in children.* Unpublished master's thesis, City University London.

Nye, C., Foster, S. H., & Seaman, D. (1987). Effectiveness of language intervention with language/learning disabled children *Journal of Speech and Hearing Research, 52,* 348–357.

Olswang, L. B. (1998). Treatment efficacy research. In C. Frattali (Ed.), *Measuring outcomes in speech-language pathology* (pp. 134–150). New York: Thieme.

Plante, E. (1998). Criteria for SLI: The Stark and Tallal legacy and beyond. *Journal of Speech, Language, and Hearing Research, 41,* 951–957.

Rice, M. L., Sell, M. A., & Hadley, P. A. (1991). Social interactions of speech and language impaired children. *Journal of Speech and Hearing Research, 34,* 1299–1307.

Robertson, S. B., & Ellis Weismer, S. (1999). Effects of treatment on linguistic and social skills in toddlers with delayed language development. *Journal of Speech, Language, and Hearing Research, 42,* 1234–1248.

Robey, R. (1998). A meta-analysis of clinical outcomes in the treatment of aphasia. *Journal of Speech, Language, and Hearing Research, 41,* 172–187.

Rutter, M., Mawhood, L., & Howlin, P. (1992). Language delay and social development. In P. Fletcher & D. Hall (Eds.), *Specific speech and language disorders in children* (pp. 63–80). London: Whurr.

Shelton, R. L., Johnson, A. F., Ruscello, D. M., & Arndt, W. B. (1978). Assessment of parent-administered listening training for pre-school children with articulation deficits. *Journal of Speech and Hearing Disorders, 43,* 242–254.

Stark, R. E., & Tallal, R. P. (1981). Selection of children with specific language deficits. *Journal of Speech and Hearing Disorders, 46,* 114–180.

Stothard, S. E., Snowling, M. J., Bishop, D. V. M., Chipchase, B. B., & Kaplan, C. A. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language, and Hearing Research, 41,* 407–418.

Tomblin, J. B., Smith, E., & Zhang, X. (1997). Epidemiology of specific language impairment: Prenatal and perinatal risk factors. *Journal of Communication Disorders, 30,* 325–344.

Wertz, R. T. (2002) Evidence-based practice guidelines: Not all evidence is created equal *Journal of Medical Speech-Language Pathology, 10,* xi–xv.

Yoder, P. J., & McDuffie, A. (2002). Treatment of primary language disorders in early childhood: Evidence of efficacy. In P. Accardo, B. Rogers, & A. Capute (Eds.), *Disorders of language development* (pp. 151–177). Baltimore: York Press.

Received February 4, 2003

Revision received June 29, 2003

Accepted December 31, 2003

DOI: 10.1044/1092-4388(2004/069)

Contact author: James Law, PhD, Department of Language and Communication Science, City University, Northampton Square, London EC1V OHB, United Kingdom. E-mail: jcl@juming.u-net.com

Appendix A (p. 1 of 2). Summary of expressive language intervention studies, classified by outcome.

Study	Rating			*n*	Mean age (year; months)	Intervention			Assessment	Effect size	CI	Meta-analysis
	Assignment	Blinding	Attrition			Type	Min per week	Length				
Measures of overall expressive syntax development												
Fey et al. (1993)	B	A	A	30	4;07	FS/cycles PT then	180 120 30	4.5 months 12 weeks 2 months	DSS	0.84	–0.01–1.69	Parent condition removed from analysis of clinician interventions only
Gibbard (1994)	B	C	A	36	2;09	PT	40	6 months	RDLS	2.33	1.47–3.20	Removed from analysis of clinician interventions only
Glogowska et al. (2000)	A	A	A	159	2;10	Routine	10 on average	8.4 months	PLS	0.17	–0.15–0.49	
Law et al. (1999)	A	A	C	43	3;02	IS PT	450 150	4 weeks 10 weeks	PLS	–0.52	–1.25–0.21	Removed from analysis of trials more than 8 weeks, trials including children with receptive difficulties, and parent condition from analysis of clinician interventions only
Matheny & Panagos (1978)	B	C	B	24	5;05–6;10	Monterey Language Programme	Not specified	5 months	PCLT	0.90	–0.15–1.94	
Number of utterances on the language sample												
Gibbard (1994)	B	C	A	36	2;09	PT	40	6 months		1.63	0.87–2.40	As previous
Girolametto et al. (1996b)	B	A	A	25	2;04	PT/FS	150	10 weeks		0.75	–0.07–1.56	
Law et al. (1999)	A	A	C	43	3;02	IS PT	450 150	4 weeks 10 weeks		–0.31	–1.03–0.42	As previous

Appendix A (p. 2 of 2). Summary of expressive language intervention studies, classified by outcome.

Study	Rating			*n*	Mean age (year; months)	Intervention			Assessment	Effect size	CI	Meta-analysis
	Assignment	Blinding	Attrition			Type	Min per week	Length				
Mean length of utterance in language sample												
Gibbard (1994)	B	C	A	36	2;09	PT	40	6 months		1.54	0.79–2.30	As previous
Law et al. (1999)	A	A	C	43	3;02	IS PT	450 150	4 weeks 10 weeks		–0.21	–0.93–0.52	As previous
Robertson & Ellis Weismer (1999)	B	C	C	24	2;01	IS/General stimulation	150	12 weeks		0.90	0.00–1.81	
Parent report of phrase complexity												
Gibbard (1994)	B	C	A	36	2;09	PT	40	6 months	CDI	2.10	1.27–2.94	As previous
Girolametto et al. (1996b)	B	A	A	25	2;04	PT/FS	150	10 weeks	CDI	0.96	0.13–1.80	Removed from analysis of clinician interventions only
Law et al. (1999)	A	A	C	43	3;02	IS PT	450 150	4 weeks 10 weeks	CDI	0.04	–0.69–0.76	As previous

Note. CI = confidence interval; A = adequate; B = unclear; C = inadequate; FS = focused stimulation; cycles = procedure of repeating target phonemes over the course of a treatment cycle; DSS = Developmental Sentence score; PT = parent training; RDLS = Reynell Developmental Language Scale; PLS = Preschool Language Scale; IS = interactive stimulation; PCLT = Programmed Conditioning Language Test; CDI = Communication Development Index.

Appendix B. Summary of phonological intervention studies, classified by outcome.

Study	Rating: Assignment	Rating: Blinding	Rating: Attrition	*n*	Mean age (year; months)	Intervention: Type	Intervention: Min per week	Intervention: Length	Assessment	Effect size	CI	Meta-analysis
Measures of overall phonological development												
Almost & Rosenbaum (1998)	A	A	C	30	3;06	Hierarchy/ cycles	80	4 months	Goldman Fristoe	1.37	0.50–2.24	
Glogowska et al. (2000)	A	A	A	159	2;10	Routine average	10	8.4 months	Overall phonological errors	0.27	−0.05–0.59	
Lancaster (1991)	B	C	A	15	3;07	Eclectic PT	17 9	6 months 6 months	Composite deviancy score	0.49	−0.61–1.58	Parent conditions removed from analysis of clinician interventions only
Matheny & Panagos (1978)	B	C	B	24	5;05–6;10	Monterey Language Programme	Not specified	5 months	PAT	0.75	−0.28–1.77	
Munro (1999)	B	A	B	13	4;06	Hierarchy	60	6 weeks	EAT	0.53	−0.73–1.79	Removed from analysis of only trials more than 8 weeks long
Shelton et al. (1978)	B	C	A	60	3;09	ADT	50	2 months	MDAS	−0.21	−0.83–0.41	Removed from analysis of clinician interventions only and only trials more than 8 weeks long
Percentage of consonants correct in conversation												
Almost & Rosenbaum (1998)	A	A	C	30	3;06	Hierarchy/ cycles	80	4 months		1.91	0.96–2.86	
Correct consonants when retelling a story using target consonants												
Munro (1999)	B	A	B	13	4;06	Hierarchy	60	6 weeks		1.29	−0.11–2.69	Removed from analysis of only trials more than 8 weeks long

Note. CI = confidence interval; Hierarchy = traditional articulation therapy hierarchy from minimal pairs teaching contrasts to production at word, sentence, and conversational levels; cycles = procedure of repeating target phonemes over the course of a treatment cycle; PT = parent training; PAT = Picture Articulation Test; EAT = Edinburgh Articulation test; ADT = Auditory Discrimination Training; MDAS = McDonald Articulatory Screening.

Appendix C. Summary of vocabulary intervention studies, classified by type of intervention and outcome.

Study	Rating			*n*	Mean age (year; months)	Intervention			Assessment	Effect size	CI	Meta-analysis
	Assignment	Blinding	Attrition			Type	Min per week	Length				
Overall measures of vocabulary development												
Gibbard (1994)	B	C	A	36	2;09	PT	40	6 months	RAPT	1.79	1.01–2.58	Removed from analysis of clinician interventions only
Law et al. (1999)	A	A	C	43	3;02	IS PT	450 150	4 weeks 10 weeks	BPVS	0.19	−0.54–0.91	Removed from analysis of only trials more than 8 weeks long, trials including children with receptive difficulties, and parent condition removed from analysis of clinician interventions only
Different words in language sample												
Gibbard (1994)	B	C	A	36	2;09	PT	40	6 months		1.02	0.32–1.72	As above
Girolametto et al. (1996b)	B	A	A	25	2;04	PT/FS	150	10 weeks		1.07	0.22–1.92	Removed from analysis of clinician interventions only
Robertson and Ellis Weismer (1999)	B	C	C	24	2;01	IS/General stimulation	150	12 weeks		1.21	0.26–2.15	
Parent report of vocabulary												
Gibbard (1994)	B	C	A	36	2;09	PT	40	6 months	CDI	2.20	1.36–3.05	As above
Girolametto et al. (1996a)	B	A	A	16	2;05	PT/FS	150	10 weeks	CDI	0.24	−0.75–1.22	Removed from analysis of clinician interventions only
Girolametto et al. (1996b)	B	A	A	25	2;04	PT/FS	150	10 weeks	CDI	0.66	0.05–1.71	Removed from analysis of clinician interventions only
Law et al. (1999)	A	A	C	43	3;02	IS PT	450 150	4 weeks 10 weeks	CDI	0.47	−0.26–1.20	As above
Robertson & Ellis-Weismer (1999)	B	C	C	24	2;01	IS/General stimulation	150	12 weeks	CDI	0.61	−0.27–1.49	

Note. CI = confidence interval; PT = parent training; RAPT = Renfrew Action Picture Test; IS = interactive stimulation; BPVS = British Picture Vocabulary Scale; FS = focused stimulation; CDI = Communication Development Index.

Appendix D. Summary of auditory association intervention studies, classified by type of intervention and outcome

Study	Rating: Assignment	Rating: Blinding	Rating: Attrition	n	Mean age (year; months)	Intervention: Type	Intervention: Min per week	Intervention: Length	Assessment	Effect size	CI
Measure of auditory association											
Shelton et al. (1978)	B	C	A	60	3;09	ADT	50	2 months	ITPA	0.53	–0.10–1.16

Note. CI = confidence interval; ADT = auditory discrimination training; ITPA = Illinois Test of Psycholinguistic Abilities.

Appendix E. Summary of receptive language intervention studies, classified by type of intervention and outcome.

Study	Rating: Assignment	Rating: Blinding	Rating: Attrition	n	Mean age (year; months)	Intervention: Type	Intervention: Min per week	Intervention: Length	Assessment	Effect size	CI	Meta-analysis
Measures of overall development of receptive syntax												
Glogowska et al. (2000)	A	A	A	159	2;10	Routine	10 on average	8.4 months	PLS	0.19	–0.12–0.51	
Law et al. (1999)	A	A	C	43	3;02	IS PT	450 150	4 weeks 10 weeks	PLS	–0.45	–0.18–0.28	Removed from analysis of only trials more than 8 weeks long, receptive difficulties, and parent condition from analysis of clinician interventions only

Note. CI = confidence interval; PLS = Preschool Language Scale; IS = interactive stimulation; PT = parent training.

Appendix F. Summary of parent and clinician intervention studies, classified by type of intervention and outcome.

Study	Rating			*n*	Mean age (year; months)	Intervention			Assessment	Effect size	CI
	Assignment	Blinding	Attrition			Type	Min per week	Length			
Comparison of measures of overall expressive syntax development											
Fey et al. (1993)	B	A	A	30	4;07	FS/cycles PT	180 120 then 30	4.5 months 12 weeks 2 months	DSS	0.23	–0.63–1.08
Gibbard (1994)	B	C	A	25	2;08	PT or clinician therapy	40 30	6 months 6 months	RDLSl	0.32	–0.64–1.28
Law et al. (1999)	A	A	C	43	3;02	IS PT	450 150	4 weeks 10 week	PLS	–0.49	–0.26–0.28
Comparison of measures of overall expressive vocabulary development											
Gibbard (1994)	B	C	A	25	2;08	PT Clinician therapy	40 30	6 months 6 months	RAPT	0.33	–0.63–1.29
Law et al. (1999)	A	A	C	43	3;02	IS PT	450 150	4 weeks 10 weeks	BPVS	0.11	–0.65–0.87
Comparison of measures of overall receptive syntax development											
Law et al. (1999)	A	A	C	43	3;02	IS PT	450 150	4 weeks 10 weeks	PLS	–0.11	–0.87–0.65

Note. FS = focused stimulation; cycles = the procedure of repeating target phonemes over the course of a treatment cycle; DSS = Developmental Sentence score; PT = parent training; RDLSl = Reynell Developmental Language Scale; IS = interactive stimulation; PLS = Preschool Language Scale; RAPT = Renfrew Action Picture Test; BPVS = British Picture Vocabulary Scale.