INT J LANG COMMUN DISORD, MARCH–APRIL 2011,
VOL. 46, NO. 2, 138–154

Review

Systematic review of the literature on the treatment of children with late language emergence

Amory Law Cable† and Celeste Domsch‡

†Bisceglie, Italy
‡Department of Communication Disorders, Texas State University, San Marcos, TX, USA

(*Received 25 August 2009; accepted 18 April 2010*)

Abstract

Background: Research has indicated that 24–36-month-olds with late language emergence ('late talkers') are at risk for later language-learning difficulties. Previous reviews have examined the efficacy of treatment for children with language delay/disorders; however, no systematic review has examined the effects of language treatment specifically for children with late language emergence.
Aims: This systematic review reports the effects of intervention studies conducted between 1985 and 2008 of 24–36-month-olds with late language emergence.
Main Contribution: Eleven studies that included a total of 275 participants were located, synthesized, and appraised for quality. Studies varied significantly with respect to methodological quality, with seven of the eleven receiving moderately strong quality scores.
Outcomes & Results: Treatment for children with late language emergence improved performance on formal measures of language, mean length of utterance, and target word use as indicated by medium to large effect sizes in these good-quality studies. Treatment descriptions and clinical implications are provided.
Conclusions & Implications: There is evidence to suggest that focused stimulation and modelling of single words can lead to improvements in the language of children with late language emergence. Subsequent treatment studies should examine the application of these effective treatments to larger, population-based samples. Future treatment studies would also benefit from specific examination of children's receptive language level and inclusion of children with late language emergence who vary in socio-economic status and/or ethnicity.

Keywords: late talkers, evidence-based practice (EBP), language delay, treatment.

What this paper adds
Systematic reviews have been completed regarding children with language delays/disorders but no review has specifically examined children in the 24–36-month range, known as children with late language emergence. Previous reviews have not examined whether treatment is effective for increasing the acquisition of specific target words, a question of clear clinical interest. There is good evidence that target word use improved as a result of treatment; however the actual number of target words learned was low, ranging from three to ten words for the treatment groups. Future research should investigate children's comprehension levels pre- and post-treatment, report individual responses to treatment, and examine whether these effective treatments for late language emergence remain so in larger, population-based samples.

Introduction

Children between the ages of 24–36 months with obvious expressive vocabulary delays have received a variety of diagnostic labels: 'late talkers' (Paul *et al.* 1991), 'early expressive language delay' (Whitehurst *et al.* 1991), 'early language delay' (Scarborough and Dobrich 1990, Dale *et al.* 2003), 'specific expressive language impairment' (SLI-E) (Rescorla and Goossens 1992), and 'slow expressive language development' (SELD) (Paul *et al.* 1996). Recently, they have been termed children with 'late language emergence' (LLE) (Zubrick *et al.* 2007), the person-first term that will be used in this review. Research has expanded significantly

Address correspondence to: Amory L. Cable, Via Siciliani, 44, Bisceglie I-70052, Italy; e-mail: amory.cable@gmail.com

International Journal of Language & Communication Disorders
ISSN 1368-2822 print/ISSN 1460-6984 online
http://www.informahealthcare.com
DOI: 10.3109/13682822.2010.487883

in the past two decades to describe this group, investigate the longitudinal effects of late language emergence, and examine intervention outcomes.

Characteristics of children with LLE

Two-year-old children with small vocabulary sizes in the absence of any underlying pathology (for example, neurological, sensory or cognitive deficit) are labelled LLE, with an estimated prevalence rate of 13.4% in a large epidemiological study (Zubrick *et al.* 2007). This same study, while including only 2-year-olds, notes that children labelled LLE in other studies have ranged in age from 18 to 35 months. A recent review (Desmarais *et al.* 2008) examined 25 publications describing the characteristics of this population and included studies in which children ranged in age from 12 to 39 months.

The expressive language criteria for diagnosing LLE have also varied, with some studies requiring fewer than 50 words produced on the Language Development Survey (Rescorla 1989) at the age of 2 years; some using a score below the 10th percentile on The MacArthur Communicative Development Inventories (CDI) (Fenson *et al.* 1993, Robertson and Ellis Weismer 1999); and one using a score of −1.0 SD (standard deviation) on the Communication scale of the Ages & Stages Questionnaire (Bricker *et al.* 1999, Zubrick *et al.* 2007).

In addition to the variability regarding ages and expressive language skills, Desmarais *et al.* (2008) reported two different definitions of LLE. In one, children have intact comprehension (Rescorla 1989), while in the other, children's comprehension levels freely vary (Thal *et al.* 1991). Desmarais *et al.* (2008) argue that this variability in language comprehension abilities means that children with LLE are not a homogenous group.

It is important to note that older studies of children with LLE have included primarily children from middle- to upper-middle class families (Ellis Weismer *et al.* 1993, Rescorla and Fechnay 1996, Paul *et al.* 1991) or have failed to report levels of socio-economic status (Thal *et al.* 1991). More recent work (for example, Horwitz *et al.* 2003, and Baxendale and Hesketh 2003) has included children from a wider socio-economic status range. The issue of ethnicity was examined by Rescorla and Achenbach (2002) who studied a population-based sample of 278 children between 18 and 35 months old, divided into groups of non-Latino White, African-American, Hispanic or 'other' (Asian, Native-American, South Asian). Although they reported that non-Latino White children had significantly higher vocabulary scores than children from any other ethnic group, questions remain regarding the issue of under-reporting. Roberts *et al.* (1999) studied vocabulary growth in 87 African-American children, primarily from low-income families, and found that vocabulary and utterance length grew, as expected, in a linear fashion from 18 to 30 months. They expressed concern, however, that African-American parents may be consistently under-reporting their children's vocabulary development, as an unexpectedly high proportion of children scored below the 10th percentile for vocabulary size.

A range of outcomes has been noted for children with LLE, perhaps due to the heterogeneity in ages, vocabulary sizes at intake, and comprehension abilities. Rescorla *et al.* (1997) reported that children diagnosed with LLE between 24 and 31 months scored in the average range for single-word vocabulary at 36 months of age, but remained 1.5 SDs below the mean for mean length of utterance (MLU). Paul *et al.* (1996) reported no significant differences in narrative skills between second-graders who had LLE between 20 and 34 months and those with typical language development. In this study, comprehension skills were not assessed when the children were between 20 and 34 months. In contrast, Manhardt and Rescorla (2002) found that oral narrative skills of 8- and 9-year-olds who were diagnosed with LLE between 24 and 31 months were weak in comparison with their typical peers. Interestingly, the children with LLE in this study had all achieved a receptive vocabulary score on the Reynell that was within 3 months of chronological age. Rescorla (2009) espoused a dimensional model of language development, such that children with LLE likely exhibit a weaker endowment for language acquisition. This weaker endowment results in reduced linguistic performance relative to peers.

Given these conflicting findings regarding outcomes, it is unsurprising that there is debate regarding appropriate clinical management for LLE. Paul (1996) argued that children with LLE 'typically move into the normal range of general language ability by school age' (p. 12) and her specific recommendation was to 'watch and see' (p. 15) children's growth in language skills. However, this approach was criticized for its lack of specificity by van Kleeck *et al.* (1997), on the grounds that 26% of children with LLE in Paul *et al.*'s sample did not have language skills in the average range at kindergarten entry. Van Kleeck *et al.* argued that Paul *et al.*'s data provided no reliable method to predict whether individual children with LLE would show recovery without intervention.

Interventions for children with LLE

When treatment is provided to children with LLE, techniques have included focused stimulation and modelling of single words. Focused stimulation employs a social-interactionist perspective on language development. Adults establish a joint focus with the child, and present a language target, which could include

an object label ('Dog!'), a comment ('Wow!'), or a two-word combination ('Big doggie'). Children are not required to provide a specific response or to imitate. The best-known focused stimulation approach is the Hanen Program for Parents (HPP) (Manolson 1992). Aspects of HPP include child-oriented techniques such as following the child's lead, promoting turn-taking to encourage the child's participation in conversation, and modelling language regarding the child's focus of attention. Outcome measures vary, but can include standardized language test scores or language use in conversation.

A second common treatment technique is modelling target words in which a clinician or parent repeatedly presents single words to the child with LLE in a play context. The child may or may not be required to imitate the word (Girolametto *et al.* 1995, 1996, Lederer 2001). Target words are chosen for their ability to be understood by the child, their ability to be represented by real objects, and their presence in young children's vocabularies. Words may also be chosen if they contain sounds that are already present in the child's phonetic inventory. The outcome measure of interest in studies examining this intervention technique is the number of target and/or control words learned.

Rationale and research questions

Previous meta-analyses have examined the efficacy of treatment for school-age children with language disorders (Cirrin and Gillam 2008), and for children with developmental speech and language delays/disorders (Law *et al.* 2004) and two reviews have explored the characteristics of children with LLE (Desmarais *et al.* 2008, Kelly 1998); however, no systematic reviews have examined the effects of language treatment specifically for children with LLE with an emphasis on techniques and outcomes that hold clinical significance for this group. The research questions are as follows:

- Does treatment improve performance on formal language measures for children with LLE?
- Does treatment improve MLU for children with LLE?
- Does treatment lead to the production of specific target words for children with LLE?

Method

To locate articles, a multifaceted search strategy used procedures similar to previous systematic reviews (Cirrin and Gillam 2008, Law *et al.* 2004) and published guidelines of systematic reviews (Schlosser and O'Neil-Pirozzi 2006, Chambers 2004). First, the following search terms were developed: late talk* or language delay* or developmental delay* or delayed language or late language emerg* AND preschool or infant or toddler or 2-year-old or 3-year-old or child,* AND intervention or treatment or therapy or instruction or parent training or program, NOT autism or syndrome or disease. The following databases were searched: CINAHL Plus with Full Text, ERIC, MasterFILE Premier, MEDline, PsycARTICLES, Psychology and Behavioral Sciences Collection and PsycINFO. The Cochrane and the Campbell Library of Systematic Reviews were searched separately. Searches included articles published between 1985 and 2008 and did not extend to articles that had not been peer-reviewed.

A total of 1936 articles were retrieved using these search terms and 236 were selected based upon title alone. Both authors read abstracts of the articles to eliminate those that did not fit the criteria. After this initial search, references of each selected article were also searched. A second database search was conducted after coding to identify any articles that had been recently published. A total of 48 full-text versions of articles were retrieved. Of these, six studies met the criteria for inclusion and five additional studies were located by scanning the bibliographies of the other studies.

Criteria for study inclusion

All studies selected for review in this systematic review met the following criteria:

- *Participants*. Studies that included participants with a mean age of 36 months or less, and who were identified as late-talking or as having a language delay, vocabulary delay, expressive language delay, or delays in language acquisition were selected for this review. Studies that included children with hearing loss, developmental delays, cognitive delays, autism, pervasive developmental disorder (PDD), or other neurological disorders were excluded.
- *Study design.* Research designs included treatment-comparison group designs, single group designs, and single-subject studies. Case studies that did not incorporate single subject research design, in which a subject serves as his/her control, were not included.
- *Dependent variable.* Dependent measures had to include measures of receptive or expressive language or both.
- *Independent variable.* Independent variables included any treatment designed to increase the expressive language abilities of the participants. Treatments were delivered by speech–language pathologists, graduate students, parents, or any combination.

- *Language of article.* Only articles published in English were selected.

Data analysis

Coding procedures

Each study was coded for age of the participants, research design, intervention variables, type of intervention, and effect sizes. All information except effect sizes was first coded by graduate students and then double-coded for all of the articles by one of the authors. Disagreements or ambiguities in coding were negotiated through discussion.

Effect size calculation

Effect size refers to the magnitude of effect that an intervention has on outcome (Klein 2004) and is considered a measure of practical significance of a treatment. When sufficient data were provided, effect sizes (Cohen's *d*) were calculated for studies with a treatment and a control group as the difference between the post-test means of the treatment group and control group divided by the pooled standard deviation (Cohen 1988). In a study with two treatment groups and a control group (Gibbard 1994), the effect size was calculated for both treatment groups by comparing each with the control group:

$$\text{Cohen's } d = \frac{\bar{X}_{\mathrm{T}} - \bar{X}_{\mathrm{C}}}{\sqrt{\frac{(n_{\mathrm{T}}-1)s_{\mathrm{T}}^2+(n_{c-1})s_C^2}{n_{\mathrm{T}}+n_c-2}}}$$

Due to the small sample sizes in the studies, Hedges' *g* was used as a correction factor. Hedges' *g* has been recommended for use in systematic reviews for both small and large samples for its ability to correct upward bias for small sample sizes ($n < 20$) and to converge to Cohen's *d* for large sample sizes (Turner and Bernard 2006, Lipsey and Wilson 2001, Hedges and Olkin 1985). The formula for Hedge's *g* is noted below with n_{T} as the number of participants in the treatment group and n_{C} as the numbers in the control group. The numbers '3', '4' and '9' are always constant in the formula:

$$\text{Hedges'g} = \text{Cohen's } d.\left(1 - \frac{3}{4(n_T - n_c) - 9}\right)$$

Effect sizes were not synthesized into an overall effect of treatment, or meta-analytic statistic, due to the range of study designs, intervention types and outcome measures. An effect size of 0.2 is considered to be small, 0.5 to be medium, and 0.8 or greater to be large (Cohen 1988). The confidence interval, the interval within which the population mean is expected to lie, was calculated for each effect size, so that the reader can better evaluate the significance of the effect. If the confidence interval around an effect size is large, it is considered a less accurate estimate of the effect size. In addition, if the confidence interval crosses 'zero', there is an increased chance that the results are not significant. For example, if the confidence interval is set at 95% and the effect size has a negative value, the effect is not significant at the $p < 0.05$ level. Effect sizes should be interpreted with some caution due to the different metrics used with regard to study design (Beretvas 2005).

Research quality

The American Speech–Language–Hearing Association's (ASHA) levels-of-evidence scheme (Mullen 2007, Cherney *et al.* 2008) was used to assess studies for methodological quality (table 1). Each study was assessed by both authors with regard to the following quality indicators: study design, assessor blinding, sampling/allocation, subject comparability/description, outcomes, significance, precision and intention to treat. Discrepancies in scoring were resolved through discussion. A study received 1 point for each indicator that met the highest quality level. For controlled trials, the maximum quality score was 9. The category 'intent to treat' was not relevant to single-subject and single-group studies, and the maximum score was 8.

Table 1. 1. ASHA's levels-of-evidence quality indicators

Indicator	Highest level of quality to receive 1 point
Study design	Controlled trial
Blinding	Assessors blinded
Sampling	Random sampling is adequately described
Group/participant similarity	Groups/participants are comparable at baseline and are adequately described
Treatment fidelity	Evidence of treatment fidelity
Outcomes	At least one primary outcome measure is valid and reliable
Significance	*p*-value is reported or calculable
Precision	Effect size is reported or calculable
Intent to treat (controlled trials only)	Participants are analysed by the group to which they were originally assigned; the study received 1 point if the number of participants at baseline was the same as at post-treatment

Source: Adapted from Mullen (2007).

Beyond quality assessment, Mullen (2007) advocates that studies be considered within their stage of research, and that future research should address gaps in the continuum of stages. In the first stage, 'exploratory' research, treatment approaches are assessed in a context in which they show promise of being efficacious (for example, pilot study). In the second stage, 'efficacy' research, the treatment shows promise within a highly controlled condition. In the third stage, 'effectiveness' research, the treatment is tested in a more 'real-world' situation. The last stage involves cost-benefit or public policy issues, in which research is considered within the political and economic environment. Treatment for children with LLE is considered within these stages of research.

Results

Eleven studies ($n = 275$) met the inclusion criteria. One publication contained two separate studies, each with different participants (Gibbard 1994). Table 2 provides a summary of the eleven studies reviewed in this article with regard to participants, intervention methods, and major findings. Table 3 reports effect sizes for studies that included sufficient information for effect size calculation and table 4 provides a summary of study quality.

Participants

Studies included children ranging in age from 20 to 47 months at intake, with an average age in each study no older than 36 months. Inclusion criteria included low expressive language ability as measured by standardized tests, by percentile rank on the MacCarthur CDI (Fenson *et al.* 1993), or vocabulary size. Exclusionary criteria included developmental delay, hearing impairment, oral motor problems, fine or gross motor skill impairments, or autism. Four studies specifically reported that children were in monolingual English homes, and no study indicated specifically that children were exposed to a language other than English.

Eight of the eleven studies reported measuring language comprehension to determine eligibility for participant inclusion. Of these eight, two studies included only children with typical comprehension skills, defined as scores in the average range on the Preschool Language Scale—3 (Zimmerman *et al.* 1997) or on the Peabody Picture Vocabulary Test (PPVT) (Whitehurst *et al.* 1991, Dunn and Dunn 1981). The remaining six studies included children with both receptive and expressive delays, as measured by the Battelle Developmental Inventory™ (Newborg 2005), the Sequenced Inventory of Communication Development (SICD) (Hedrick *et al.* 1984), the PLS-3 (PLS-3), and the Vineland Adaptive Behavior Scales (Sparrow *et al.* 2008).

Effect of treatment on performance on formal language measures

Nine of the eleven studies used formal measures to assess treatment outcomes for children with LLE. Formal measures included those testing overall expressive and receptive language ability and those measuring expressive vocabulary only. Five studies were randomized controlled trials (RCTs) with two studies reported in one publication (Girolametto *et al.* 1995, 1996, Gibbard 1994, Robertson and Ellis Weismer 1999). Three were non-randomized controlled trials (Baxendale and Hesketh 2003, McDade and McCartan 1998, Whitehurst *et al.* 1991), and one was a single-treatment group design (Lederer 2001). A total of 252 children participated in these studies.

Formal expressive and receptive language measures

Global language measures included the PLS-3, the Reynell Developmental Language Scales (Reynell and Gruber 1990), the Expressive One-Word Picture Vocabulary Test (EOWPVT) (Gardner 1990), and the Illinois Test of Psycholinguistic Abilities (ITPA) (Kirk *et al.* 1968). The use of these standardized tests varied among studies, making interpretation and synthesis of scores difficult. In two studies, raw scores rather than standard scores were used as outcome measures (Gibbard 1994). In other studies, gains in standardized test scores were noted by the authors but means and standard deviations were not provided (Whitehurst *et al.* 1991, McDade and McCartan 1998).

Gibbard (1994) completed two experiments, the first comparing parent-delivered focused stimulation to a delayed-treatment control. The treatment group demonstrated significant gains on both the receptive and expressive portions of the Reynell (ES = 1.42 and 2.33, respectively). The second experiment by Gibbard (1994) compared three conditions with each other: parent-delivered focused stimulation, individual speech–language therapy, and a parent-delivered control condition in which children received general cognitive stimulation. Both treatment groups showed significantly greater gains in raw scores on standardized language tests than the control group. Effect sizes were calculated for the parent-delivered treatment group versus the control group, and were moderate for receptive scores (ES = 0.68) and high for expressive scores (ES = 1.34). Effect sizes for the individual therapy group versus the control group were lower and differences were not significant on these measures (Receptive ES = −0.2; Expressive ES = 0.73).

McDade and McCartan (1998) also compared a parent-delivered treatment group (HPP) with a control condition (no treatment). Significant gains were reported on the PLS-3 for expressive language and total

Table 2. Intervention studies for children with LLE

Reference	Participants	Intervention	Major findings
Baxendale and Hesketh (2003)	*Number of participants*: 37 *Inclusion criteria*: Standard score of 80 or less on the expressive section of the PLS-3-UK; no profound sensorineural hearing loss; normal non-verbal development *Age*: 30–42 months *SES*: Inner city with 'high level of social deprivation' *Language status*: Monolingual English	*Design*: Pre-test/post-test with assignment by location to one of two experimental groups *Treatment goal*: To increase the child's language *Therapy methods*: Hanen Parent Program—HPP (T1): Hanen-trained SLPs followed the published programme designed to provide parents with naturalistic interaction strategies with the purpose of increasing the child's communication Individualized treatment (T2): SLPs who had not received Hanen training conducted individual, in-clinic treatment with parents present designed to improve children's receptive and expressive language abilities *Setting*: T1: Primarily clinic, but also included home visits; T2: clinic *Duration*: T1: 11 weeks (8 group sessions, lasting 2 hrs 15 min, plus three home visits, lasting 30–45 min); T2: 8–12 weeks (45-min sessions once a week)	Six months after the last treatment session, PLS-3-UK scores for both groups were either in or near the average range Differences between T1 and T2 group scores on the PLS-3-UK and on MLU were not significant at any assessment point
Ellis Weismer et al. (1993)	*Number of participants*: 3 *Inclusion criteria*: Restricted productive vocabulary and normal cognitive functioning; vocabularies fell below the 10th percentile on the MacArthur Communicative Developmental Inventories (CDI) *Age*: 27–28 months *SES*: Upper middle class *Language status*: Monolingual English	*Design*: Single-subject alternating treatments (words were taught under two types of treatment methods) *Treatment goal*: To increase the child's productive vocabulary *Therapy methods*: Modelling method (T1): Trained graduate student clinicians provided focused repetition of target words during functional group and individual activities. The child was not required to produce a verbal response but spontaneous productions were not discouraged Modelling plus evoked production (T2): Trained graduate student clinicians provided focused repetition of target words and intermittent opportunities for subjects to produce words and to receive feedback regarding their productions during functional group and individual activities *Setting*: Clinic *Duration*: 3 months (40–45 min of group sessions, 15–20 min individual), total 24 sessions	All children learned more target words than control words. Participant 1 gained most from modelling (M), participant 3 gained most from modelling plus evoked production (MEP) and participant 2 did not make significant gains

Table 2 – *continued*

Reference	Participants	Intervention	Major findings
Gibbard (1994), Experiment 1	*Number of participants*: 36 *Inclusion criteria*: Little or no expressive language (vocabulary of 30 or fewer single words); no diagnosis of a developmental delay or medical condition; no previous speech or language therapy *Age*: 27–39 months *SES*: Parents of participants distributed among all SES classes with most in middle class in skilled occupations *Language status*: Not reported	*Design*: Random assignment to treatment and control group (delayed treatment group) *Purpose*: To increase child's linguistic complexity from single word utterances to multi-word utterances *Therapy methods*: Parent-delivered intervention: Parents attended small group speech and language training sessions with other parents and were guided in devising activities to teach language objectives to children at home *Setting*: Parents were trained in a health centre; Parents provided treatment to children at home *Duration*: 6 months (11 group sessions; 1–1.25 h per session)	The children in the treatment group showed greater mean score improvement than the control group on standardized and informal measures
Gibbard (1994), Experiment 2	*Number of participants*: 25 *Inclusion criteria*: Little or no expressive language (vocabulary of 30 or fewer single words); no diagnosis of a developmental delay or medical condition; no previous speech or language therapy *Age*: 27–39 months *SES*: Participants distributed among classes with most in middle class in skilled occupations *Language status*: Not reported	*Design*: Random assignment to one of three groups *Purpose*: To increase child's linguistic complexity from single word utterances to multi-word utterances *Therapy methods*: Individual speech and language therapy (T1): Trained SLPs provided one-to-one therapy and gave parents general advice on the aims of the sessions and strategies used Parent-delivered language intervention (T2): Parents attended small group speech and language training sessions with other parents and were guided in devising activities to teach language objectives to children at home Parent-delivered cognitive training (T3): Parents attended small group for training in cognitive tasks that were unrelated to speech and language therapy *Setting*: Parents were trained in a health centre and provided treatment to children at home *Duration*: T1: 6 months (weekly 30-min sessions); T2: 6 months (11 group sessions; 1–1.25 h per session); T3: 6 months (11 group sessions; 1–1.25 h per session)	The parent delivered language intervention (T2) was found to be at least as effective as the individual language therapy (T1) based on standardized and informal measures. The parent language treatment and the individualized therapy groups made significantly greater gains than T3 which targeted only non-language activities

Table 2 – *continued*

Reference	Participants	Intervention	Major findings
Girolametto et al. (1995)	*Number of participants*: 16 (child/mother dyads) *Inclusion criteria*: Expressive vocabulary delay; no sensory impairment, oral motor problems, neurological problems, PDD, or autism; children at the single-word stage of language acquisition as determined by the MacArthur CDI *Age*: Range = 24–42 months *SES*: middle class *Language status*: Monolingual English	*Design*: Random assignment to treatment and control (delayed treatment) groups *Purpose*: To increase expressive vocabulary *Therapy methods*: Hanen Parent Program (HPP) with focused stimulation: Parents were taught to use naturalist interaction strategies HPP and were also trained to incorporate ten target words into daily routines *Setting*: Clinic and home *Duration*: 10 weeks (seven evening sessions to teach programme strategies/three home visits)	The children in the treatment group used a greater number of new words in semi-structured tasks than the children in the control group. There was no significant difference in the number of control words used or the total vocabulary size
Girolametto et al. (1996)	*Number of participants*: 25 (child/mother dyads) *Inclusion criteria*: Expressive vocabulary delay; no sensory impairment, oral motor problems, neurological problems, PDD, or autism; children scored in the lower 5th percentile for vocabulary on the MacArthur CDI *Age*: 23–35 months *SES*: middle class *Language status*: Monolingual English	*Design*: Random assignment to treatment and control (delayed treatment) groups *Purpose*: To increase expressive vocabulary *Therapy methods*: Hanen Parent Program (HPP) with focused stimulation: Parents were taught to use naturalist interaction strategies of HPP and were also trained to incorporate ten target words into daily routines *Setting*: Clinic and home *Duration*: 11 weeks (eight evening sessions to teach programme strategies/three home visits)	The children in the treatment group used a greater number (target words and non-target words) and greater diversity of words as compared with the children in the control group. No significant differences in measures of talkativeness were found between the two groups
Lederer (2001)	*Number of participants*: 10 (child/parent dyads) *Inclusion criteria*: Single-word stage of language development; expressive vocabulary ranging from three to 42 words *Age*: 23–29 months *SES*: Middle class *Language status*: Monolingual English	*Design*: Single-treatment group, no comparison group *Purpose*: To increase expressive vocabulary *Therapy methods*: Focused stimulation approach – TOTalk: Supervised graduate clinicians modelled a focused stimulation approach in child-directed play activities and adult-directed activities (e.g., circle time, rhythm and movement, sensory play, snack, story time, closing circle time). Parents were active participants during all activities and also modelled language-facilitation techniques. Vocabulary targets were presented multiple times during activities *Setting*: Clinic *Duration*: 11 weeks (90-min sessions, once a week)	All children increased overall and target vocabulary. Parents reported satisfaction with the programme

Table 2 – *continued*

Reference	Participants	Intervention	Major findings
McDade and McCartan (1998)	*Number of participants*: 20 *Inclusion criteria*: Fewer than 50 word expressive vocabulary; No use of word combinations; Normal receptive language *Age*: 24 months (No specific group or individual data reported) *SES*: Not reported *Language status*: Not reported	*Design*: Non-random assignment to treatment and control groups (control group families were unable to attend treatment) *Treatment goal*: Increase joint engagement *Therapy methods*: Treatment (Hanen Parent Program): Hanen-trained SLPs followed the published programme designed to provide parents with naturalistic interaction strategies Setting: Primarily clinic with several home visits Duration: 12 weeks (nine group sessions and three home visits)	Children in the treatment group exhibited significantly higher scores on the receptive and expressive portions of the PLS-3 at post-test, while the control group displayed no change
Robertson and Ellis Weismer (1999)	*Number of participants*: 21 *Inclusion criteria*: Vocabulary level below the 10th percentile on the McArthur Communicative Developmental Inventories *Age*: 21–30 months *SES*: Middle class *Language status*: Monolingual English	*Design*: Random assignment to treatment and control (delayed treatment) groups *Purpose*: To increase language and social skills *Therapy methods*: Interactive child-centred intervention (T): Trained clinicians provided general language stimulation techniques such as: parallel talk, expansion/expatiation, and recasts No treatment (C): Children in the control group received treatment after the children in the experimental group *Setting*: Early intervention centre *Duration*: 12 weeks (75-min group sessions, twice a week)	Significant group differences in favour of the treatment group were found on all variables including: MLU, total number of words, number of different words, and percentage of intelligible utterances

Table 2 – *continued*

Reference	Participants	Intervention	Major findings
Whitehurst et al. (1991)	*Number of participants*: = 62 (at first post-test) *Inclusion criteria*: A standard score below 65 on the EOWPVT; normal non-verbal IQ, receptive language, and hearing; no autism or PDD *Age* = mean age 27.95 (SD = 3.4) *SES*: Not reported *Language status*: Not reported	*Design*: Prospective longitudinal study with a control group *Purpose*: Increase labelling, decrease gestural communication, and initiate use of two-word combinations *Therapy methods*: Parent-delivered treatment (T): Parents were trained in use of seven assignments which encouraged and/or required child to label or imitate labels for objects, and later to combine words No treatment (C): Children did not receive the intervention and were referred to other community-based treatment services *Setting*: Home *Duration*: 6 months (bi-weekly office visits, lasting 30 min each); parents were encouraged to complete assignments with children several times per day	Treatment group had significantly higher post-test scores than the control group on the EOWPVT and ITPA immediately following treatment
Wilcox et al. (1991)	*Number of participants*: 20 *Inclusion criteria*: A score at least 1.5 SD below the mean on the receptive and expressive portions of the Sequenced Inventory of Communication Development or on the communication subscale of the Battelle Developmental Inventory; No diagnosed neurological impairment *Age*: 20–47 months *SES*: Not reported *Language status*: Not reported	*Design*: Random assignment to one of two treatment groups (classroom based or individual therapy) *Purpose*: Establishing productive use of a small vocabulary of core words *Therapy methods*: Classroom-based intervention (T1): Intervention was delivered in the classroom by an early childhood special educator and an SLP student clinician. Each target word modelled a minimum of ten times per day for each child. Procedures included interactive modelling, with no responses required or elicited Individualized intervention (T2): Intervention delivered in a clinic setting by an SLP student clinician. Clinicians followed engaged in parallel or cooperative play. Each target word modelled a minimum of ten times per session *Setting*: Preschool classroom and clinic *Duration*: T1: 12–16 weeks (classroom sessions were 3 h and occurred twice weekly; T2: Individual sessions were 45 min and also occurred twice weekly), total of 24 sessions for both groups	The classroom and individual treatment groups were similar for the number of target words used at post-test. Children in the classroom condition were more likely to demonstrate generalization of their target words in a home setting. Three children demonstrated no productive target word use

Table 3. Effect sizes

Study	Measures	Treatment *n*	Treatment Mean	Treatment SD	Control *n*	Control Mean	Control SD	Effect size (Hedges *g*)	95% Confidence interval	
Formal Measures of Language										
Gibbard (1994), Experiment 1	Reynell expressive	18	38.7	8.6	18	20.8	6.2	2.33	1.48	3.19
	Reynell receptive	18	40.5	9.4	18	29.3	5.6	1.42	0.68	2.15
Gibbard (1994), Experiment 2 (P)	Reynell expressive	9	34.6	8.2	8	25.5	3.5	1.34	0.28	2.40
	Reynell receptive	9	38	7.3	8	32.5	8.1	0.68	−0.3	1.66
Gibbard (1994), Experiment 2 (I)	Reynell expressive	8	31.5	10.4	8	25.5	3.5	0.73	−0.28	1.75
	Reynell receptive	8	30.4	11.6	8	32.5	8.1	−0.20	−1.18	0.78
Girolametto et al. (1995)	MacArthur (CDI)	8	79.5	35	8	68.9	49	0.24	−0.75	1.22
Girolametto et al. (1996)	MacArthur (CDI)	12	187.7	181	13	65.4	66	0.88	0.06	1.7
Robertson and Ellis Weismer (1999)	MacArthur (CDI)	11	76.2	37.5	10	51.4	40.8	0.61	−0.26	1.49
Informal Measures of Language										
MLU										
Gibbard (1994), Experiment 1		18	2.3	0.7	18	1.4	0.4	1.54	0.80	2.29
Gibbard (1994), Experiment 2 (P)		9	2.82	0.9	8	1.44	0.38	1.85	0.55	2.77
Gibbard (1994), Experiment 2 (I)		8	1.84	0.65	8	1.44	0.38	0.71	−0.30	1.72
Robertson and Ellis Weismer (1999)		11	1.32	0.32	10	1.09	0.11	0.90	0.0	1.80
Number of different target words used by child										
Girolametto et al. (1995)		8	3.9	2.4	8	1.5	2	1.03	−0.02	2.07
Girolametto et al. (1996)		12	3	2.1	13	1	1.2	1.14	.30	1.99

Notes: Effect sizes (Hedges *g*) are reported for studies with treatment and control groups and were calculated using the standardized mean difference (SMD) between post-test scores of the treatment and control groups.
(P), Parent-based language treatment; (I), individual speech–language treatment (one-on-one with a therapist).

Table 4. Study quality

Study	Design	Assessor blinding	Random sampling described	Participants comparable/ described	Treatment fidelity reported	Valid outcome measure	Significance	Intent to treat	Precision	Total
Baxendale and Hesketh (2003)	Controlled trial[a]	No	No	Yes	No	Yes	Yes	Yes	Yes	6/9
Ellis Weismer et al. (1993)	Single subject	No	No	Yes	Yes	Yes	No	n.a.	No	3/8
Gibbard (1994), (Experiment 1)	Controlled trial	No	Yes	Yes	No	Yes	Yes	Yes	Yes	7/9
Gibbard (1994), (Experiment 2)	Controlled trial[a]	No	Yes	Yes	No	Yes	Yes	Yes	Yes	7/9
Girolametto et al. (1995)	Controlled trial	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8/9
Girolametto et al. (1996)	Controlled trial	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8/9
Lederer (2001)	Single-treatment group	No	No	Yes	No	Yes	No	n.a.	No	2/8
McDade and McCartan (1998)	Controlled trial	No	No	No	No	Yes	Yes	Yes	No	4/9
Robertson and Ellis Weismer (1999)	Controlled trial	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	8/9
Whitehurst et al. (1991)	Controlled trial	No	No	No	No	Yes	No	No	No	2/9
Wilcox et al. (1991)	Controlled trial[a]	No	Yes	Yes	No	No	Yes	Yes	Yes	6/9

Notes: [a]Multiple treatment groups.
n.a., Not applicable.

language scores for the treatment group, and no gains for the control group. Pre- and post-test means and standard deviations were not included, so effect sizes could not be calculated.

Baxendale and Hesketh (2003) compared HPP with individualized speech–language treatment. Post-tests indicated gains in scores on the PLS-3-UK for both groups, but differences between groups were not significant. Due to the absence of a control group, it is not possible to determine if gains were due to intervention or to maturation.

Whitehurst *et al.* (1991) employed a strict imitation treatment technique where parents were told to feign misunderstanding of children's gestures and deny a requested object until the child verbally imitated the parent. The study reported significantly higher post-test scores at 34 months for the treatment group on measures of language age on the ITPA. However, an effect size could not be calculated because group mean scores and standard deviations were not provided.

Thus, of the five studies examining the effects of treatment on children's performance on formal language measures, two were high-quality RCTs and reported significant gains in receptive and expressive test scores (Gibbard 1994). Although standardized tests were used by Gibbard (1994), only raw scores were presented, making it difficult to determine how both groups performed relative to their peers. One controlled trial did not find significant differences in test scores between two types of treatment groups, HPP and individualized treatment (Baxendale and Hesketh 2003) but did find overall improvement in both groups. The remaining two studies reported significantly higher post-test scores for treatment groups but did not provide adequate information so that effect sizes could be calculated (McDade and McCartan 1998, Whitehurst *et al.* 1991). These two studies received lower study quality scores due to the absence of random assignment, a lack of information regarding the compatibility of treatment and control groups, a lack of treatment fidelity description, and the absence of assessor blinding.

Formal measures of expressive vocabulary

Three RCTs, one non-randomized study, and one single group study examined change in measures of expressive vocabulary due to treatment. The most common expressive-only measure was the MacArthur CDI (Fenson *et al.* 1993). The CDI is a parent report form used to quantify vocabulary growth. In all studies that used the CDI, the reported scores represented the participants' inventory of words rather than a standard score. A similar vocabulary checklist called the Language Development Survey (LDS) (Rescorla 1989) was used in one study (Lederer 2001). The LDS is a screening

measure that was designed to identify language delay in children around the age of 2 years old.

Although measures were standardized, reported scores in all studies represented the inventory of words as documented by parents rather than a standard score. Girolametto *et al.* (1995) compared HPP with a delayed-treatment control, and reported no significant difference for the number of different words reported by parents on the CDI post-treatment (ES = 0.24). However, in a later study, Girolametto *et al.* (1996) again compared HPP with a delayed-treatment control and found a significant difference for number of different words on the CDI following treatment, with a large effect size of 0.88. Robertson and Ellis Weismer (1999) compared individual treatment with a delayed-treatment control and reported significant gains for the number of different words reported by parents on the CDI post-treatment, with a moderate effect size (0.61).

Whitehurst *et al.* (1991) reported that the scores for the treatment group on the EOWPVT were significantly higher than those of the control group at a 34 months post-test. However, the scores between these groups were not statistically different at the 44- and 65-month post-tests. At these times, the majority of the participants in both control and treatment groups were scoring within the normal range on this test.

Lederer (2001) provided treatment to a single group by using focused stimulation techniques in a group setting. Effect sizes were not calculated because of the single group design of the study. Lederer reported a mean gain of 36.2 words (SD = 16.34, range = 16–75) on the LDS.

Four of the five studies examining vocabulary growth as measured by formal tests reported significant effects of treatment on expressive vocabulary. Three of the five studies received high-quality ratings (8/9) due to their design, inclusion of detailed participant description, report of treatment fidelity, and inclusion of sufficient pre- and post-test data to calculate effect sizes. Two of these three studies indicated moderate to large effect sizes, while one study did not report any significant difference between the treatment and control group.

Effect of Treatment on MLU

Four studies examined the effect of treatment on child MLU with a total of 119 participants. Baxendale and Hesketh (2003) reported gains in MLU in 47% of the participants in the study but no difference in change in MLU when comparing children receiving HPP with children receiving individual treatment. Gibbard (1994) reported significant gains in MLU of the treatment group over the control group for children receiving parent-delivered treatment (ES = 1.54). In a second experimental study, Gibbard found similar results when comparing a parent treatment with a control condition (ES = 1.85), and when comparing individual treatment with the same control condition (ES = 0.71). Results of a study by Robertson and Ellis Weismer (1999) also indicated a significant effect on MLU when comparing treatment and control groups (ES = 0.90). Thus, of the three studies that compared a treatment group with a control group, all demonstrated significant effects of treatment on MLU.

Effect of treatment on target word acquisition

Five studies ($n = 74$) measured the effect of treatment on the acquisition of specific target words in children with LLE. Girolametto *et al.* (1995) asked parents to model ten specific target words during naturalistic routines. Target and control words were chosen based upon these criteria: (1) comprehended by the child but not produced, (2) present in typical vocabularies of normally developing 24-month-olds, (3) contained an initial phoneme in the child's repertoire, and (4) represented objects or early developing function words (for example, more). Gains were noted for the ten treated words as compared with the control words, with an ES of 1.03. On average, children in the treatment group gained 3.9 words (SD = 2.4), while the control group they gained 1.5 words (SD = 2.0). In a second study, Girolametto *et al.* (1996) asked parents to target ten single words using the same techniques. In this study, gains in target words were also significant, with an ES of 1.14.

Wilcox *et al.* (1991) compared target word acquisition in classroom and individual treatment. Effect sizes were not calculated due to the absence of equivalent pre- and post-test measures. Ten words were identified as targets based upon criteria similar to Girolametto *et al.* (1995, 1996). Results indicated that spontaneous target word use in the treatment setting was similar for the classroom-based group and the individual treatment group. However, children in the classroom instruction group were more likely demonstrated generalization of word learning by producing target words at home than children in individual treatment.

Lederer (2001) targeted twelve specific vocabulary words that were typically early developing and that started with a phoneme in child's inventory. One to two target words were modelled naturally for a minimum of five times per session, and clinicians attempted to elicit words. Children were reported to have gained an average of seven target vocabulary words (range = 5–10 words).

Ellis Weismer *et al.* (1993) employed a single-subject alternating treatment design to compare the effects of modelling specific words (without required imitation) to modelling with required imitation in three children. Treatment was delivered in group and individual settings

by graduate student clinicians. Results indicated that both treatments were equally effective for one participant, but that another participant benefited more from modelling with required imitation. Two participants acquired ten out of the 14 target words. The third participant did not make gains under either treatment. The authors compared the use of target words versus control words and concluded that target words were learned more rapidly than the control words that were sampled.

Thus, targeting specific single words was examined by two high-quality RCTs, both of which reported significant effects. Unfortunately, it is impossible to determine whether gains reported by Wilcox *et al.* (1991) and Lederer (2001) occurred as a result of treatment or as a result of maturation. Both studies relied on frequency count for target words without reference to control words or a control group. The single-subject study which examined target word and control word growth (Ellis Weismer *et al.* 1993) yielded mixed results for different treatments.

Discussion

The purpose of this systematic review was to examine whether treatment for late language emergence (LLE) resulted in improved language skills, as measured by performance on formal language measures, MLU in conversation, and acquisition of target words. The evidence from this review indicates that treatment for children with LLE is effective in improving performance on formal language measures and in improving mean length of utterance (MLU). The acquisition of specific target words appeared to be influenced by treatment, but the total number of words learned was low. It is also clear from this review that there are relatively few studies examining the effects of treatment for 2- and 3-year-old children, despite mandates for early intervention. Of the eleven studies described here, sample sizes ranged from three to 62 children (mean = 25, standard deviation (SD) = 15.7). Quality ratings ranged from 2/9 points to 8/9 points, with no study reporting blinding of assessors for formal assessments and only four of eleven studies reporting checks of treatment fidelity. In addition, the gains resulting from treatment are typically modest, perhaps reflecting the duration of treatment, which ranged from 10 weeks to 6 months.

Types of treatment

Studies reported use of four different treatment techniques: (1) focused stimulation, (2) modelling of single words, (3) imitation of single words, and (4) traditional individual speech–language therapy.

Focused stimulation

Nine of the eleven studies in this review employed focused stimulation; however, treatment methods varied even within this framework. For example, clinicians provided direct treatment in four studies, and parents were the primary treatment providers in seven studies. Due to the different treatment methods and outcome measures used, no clear conclusion can be drawn regarding the relative effectiveness of parent versus clinician-delivered therapy. However, results from the focused stimulation studies indicate that parent training is associated with medium to large effect sizes. In addition, parent training addresses the child's language needs in a naturalistic environment, thereby maximizing communicative opportunities and participation as recommended by the Royal College of Speech & Language Therapists (RCSLT) (2005). This model also conforms to a more family-centred approach recommended by the American Speech–Language-Hearing Association (ASHA) (2008).

Modelling of single words

Five of the studies in this review had a clinician, graduate student, and/or parent repeatedly presenting a single word to the child with LLE in a play context. The child was not required to repeat the word, but was often encouraged to do so. Studies in this review targeted anywhere from ten to 14 single words per child over a 10–12-week time span. Results indicated that children learned from three to ten treated words, and far fewer of their control words. Because three of the five studies that used modelling of single words also used focused stimulation as a treatment technique, it is not possible to disentangle the effects of the two. It appears that the two techniques can be combined and can result in changes both to children's single-word vocabulary and to more global language measures.

Imitation of single words

Another technique used by studies in this review was direct, required imitation of single words. Of the three participants in Ellis Weismer *et al.* (1993), imitation was effective for one child and not for the others. Whitehurst *et al.* (1991) reported improvements in standardized test scores, but did not provide group means or standard deviations. In addition, a high rate of attrition was noted in this study, reducing confidence in the equivalence of the control and treatment groups. Support for required imitation of single words to treat children with LLE appears limited.

Individual speech–language therapy

Individual speech–language therapy was used as a comparison treatment group in three studies. Therapists were described as using traditional techniques to promote language growth. Many of the techniques used in the clinician-delivered treatment were similar to those used in parent-delivered treatment.

Study quality

While three of the eleven studies obtained a score of 8/9 in the study quality ratings, many of the studies had methodological weaknesses that negatively affected quality scores. Seven of the eleven studies included a control or comparison group, but only five of these used random assignment. Other weaknesses included a lack of assessor blinding, a lack of valid measures, and the absence of treatment fidelity procedures. Treatment fidelity, while challenging to measure in parent-centred interventions, is critical for providing a detailed description of treatment and a measure of confidence in the procedures used.

It has been previously noted that ASHA's levels-of-evidence scheme has limitations for evaluating single-subject research designs (Cherney *et al.* 2008). There was only one single-subject research study in this review (Ellis Weismer *et al.* 1993) and its low-quality score (3/8) reflects the incompatibility of this quality scheme with regard to this type of study design more than significant limitations of the study itself (Cherney *et al.* 2008). In addition, a study did not receive a '1' for intent to treat if there was insufficient information regarding the number of participants who began and finished the treatment study. It could be argued that some of the criteria in the scheme (for example, 'design') are more important for a study's methodology quality than others.

Limitations

Studies of varying methodology and methodological quality were included in this review. In addition, studies used a variety of intervention techniques and outcome measures. As a result, effect sizes cannot be reliably compared with each other or averaged together as in a meta-analysis.

This review included only studies published in English, and so may not represent a complete survey of findings. It included only published studies, and so may have excluded studies finding that treatment was not effective, as researchers may be less likely to seek or achieve publication of these studies.

Summary

In answer to the research questions, there is good evidence that treatment for children with LLE improves their performance on formal language measures. There is also evidence that treatment improves vocabulary use and MLU in conversation, as well as the acquisition of specific target words. Treatment techniques included focused stimulation, modelling of single words, and individual speech–language therapy. Effective treatment was provided by both therapists working under typical clinical conditions and by parents who received intensive training by therapists.

Future research

Three limitations are noted in the literature on treatment for LLE. First, while treatment gains are reported at the group level, clinicians and parents also have an obvious interest in individual responses. Of eleven studies in this review, four provided data on individual results. In each of these studies, there were individual children who either did not make significant gains or made no gains at all. The provision of individual outcome data would be of benefit to both researchers interested in the issue of non-response, as well as to practising therapists and families.

Second, as noted by Desmarais *et al.* (2008) and Law *et al.* (2004), receptive language skills of children with LLE are often not measured at intake, not specifically targeted in treatment, and may well be a confound in terms of response or non-response to treatment. Future studies would benefit from consistent measurement and reporting of receptive language skills.

Third, the studies in this review, with one notable exception (Baxendale and Hesketh 2003), included primarily Caucasian children from middle-class or upper- to middle-class homes. Work by Hart and Risley (1995) and Huttenlocher *et al.* (1991) has reported that vocabulary growth is strongly associated with family socio-economic status. Families where parents have high levels of education have been reported to use more words, and a greater variety of words, which contributes to language growth in children. Thus, studies of children with LLE that are restricted to high socio-economic status families may not be representative of all children with LLE. It should be noted, though, that the largest study of children with LLE to date (Zubrick *et al.* 2007) reported no link between maternal education and LLE, and argued that emergence of first words is likely under neurobiological control and less influenced by environmental factors. In addition to restricted socio-economic status, these studies rarely report inclusion of African-American children.

Finally, with regard to the stages of research described by Mullen (2007), the work on children with LLE shows promise in the third stage, with several studies investigating the treatment protocols within a 'real-world' situation (for example, Hanen Program for Parents (HPP) parent groups and classroom settings). However, the research stops short of the last stage, cost–benefit/public policy research. Future research should examine the setting and the political and economic environment in which treatment for LLE is most effective.

Acknowledgement

The authors would like to thank Lacey Beeson and Crystal Roberts for their assistance in coding articles.

References

American Speech–Language–Hearing Association (ASHA), 2008, *Roles and Responsibilities of Speech–Language Pathologists in Early Intervention. Technical Report* (available at: http://www.asha.org/policy) (accessed on 1 July 2009).

Baxendale, J. and Hesketh, A., 2003, Comparison of the effectiveness of the Hanen Parent Programme and traditional clinic therapy. *International Journal of Language and Communication Disorders*, **38**, 397–415.

Beretvas, S. N., 2005, Methodological challenges encountered in summarizing evidence-based practice. *School Psychology Quarterly*, **20**, 498–503.

Bricker, D., Squires, J., Mounts, L., Potter, L., Nickel, R., Twombly, E. and Farrell, J., 1999, *Ages & Stages Questionnaires: A Parent-Completed, Child-Monitoring System* (Baltimore, MD: Brookes).

Chambers, E. A., 2004, An introduction to meta-analysis with articles from *The Journal of Educational Research* (1992–2002). *Journal of Educational Research*, **98**, 35–44.

Cherney, L. R., Patterson, J. P., Raymer, A., Frymark, T. and Schooling, T., 2008, Evidence-based systematic review: effects of intensity of treatment and constraint-induced language therapy for individuals with stroke-induced aphasia. *Journal of Speech, Language and Hearing Research*, **51**, 1282–1299.

Cirrin, F. M. and Gillam, R. B., 2008, Language intervention practices for school-age children with spoken language disorders: a systematic review. *Language, Speech, and Hearing Services in Schools*, **39**, S110–S137.

Cohen, J., 1988, *Statistical Power Analysis for the Behavioral Sciences* (Hillsdale, NJ: Lawrence Erlbaum Associates).

Dale, P. S., Price, T. S., Bishop, D. V. M. and Plomin, R., 2003, Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language and Hearing Research*, **46**, 544–560.

Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I. and Rouleau, N., 2008, Systematic review of the literature on characteristics of late-talking toddlers. *International Journal of Language and Communication Disorders*, **43**, 361–389.

Dunn, L. M. and Dunn, D. M., 1981, *Peabody Picture Vocabulary Test (rev.)* (Circle Pines, MN: American Guidance Service).

Ellis Weismer, S. E., Murray-Branch, J. and Miller, J. F., 1993, Comparison of two methods for promoting productive vocabulary in late talkers. *Journal of Speech and Hearing Research*, **36**, 1037–1050.

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S. and Reilly, J. S., 1993, *The MacArthur Communicative Development Inventories: User's Guide and Technical Manual* (San Diego, CA: Singular).

Gardner, M. F., 1990, *Expressive One-Word Picture Vocabulary Test—Revised (EOWPVT-R)* (Novato, CA: Academic Therapy).

Gibbard, D., 1994, Parental-based intervention with pre-school language-delayed children. *European Journal of Disorders of Communication*, **29**, 131–150.

Girolametto, L., Pearce, K. and Weitzman, E., 1995, The effects of focused stimulation for promoting vocabulary in young children with delays: a pilot study. **17**, 39–49.

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

Hart, B. and Risley, T., 1995, *Meaningful Differences in the Everyday Experience of Young American Children* (Baltimore, MD: Brookes).

Hedges, L. V. and Olkin, I., 1985, *Statistical Methods for Meta-analysis* (Orlando, FL: Academic Press).

Hedrick, D. L., Prather, E. M., Tobin, A. R., Allen, D. V., Bliss, L. S. and Rosenberg, L. R., 1984, *Sequenced Inventory of Communication Development, Revised Edition* (Seattle, WA: University of Washington Press).

Horwitz, S. M., Irwin, J. R., Briggs-Gowan, M. J., Heenan, J. M. B., Mendoza, J. and Carter, A. S., 2003, Language delay in a community cohort of young children. *Journal of the American Academy of Child and Adolescent Psychiatry*, **42**, 932–940.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. and Lyons, T., 1991, Early vocabulary growth: relation to language input and gender. *Developmental Psychology*, **27**, 236–248.

Kelly, D. J., 1998, A clinical synthesis of the 'Late Talker' literature: implications for service delivery. *Language, Speech, and Hearing Services in Schools*, **29**, 76–84.

Kirk, S. A., McCarthy, J. and Kirk, W. A., 1968, *Illinois Test of Psycholinguistic Abilities (ITPA)* (Urbana, IL: University of Illinois Press).

Klein, R. B., 2004, *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research* (Washington, DC: American Psychological Association).

Law, J., Garrett, Z. and Nye, C., 2004, The efficacy of treatment for children with developmental speech and language delay/disorder: a meta-analysis. *Journal of Speech, Language, and Hearing Research*, **47**, 924–943.

Lederer, S. H., 2001, Efficacy of parent-child language group intervention for late-talking toddlers. *Infant–Toddler Intervention*, **11**, 223–235.

Lipsey, M. W. and Wilson, D. B., 2001, *Practical Meta-analysis* (Thousand Oaks, CA: Sage).

Manhardt, J. and Rescorla, L., 2002, Oral narrative skills of late talkers at ages 8 and 9. *Applied Psycholinguistics*, **23**, 1–21.

Manolson, A., 1992, *It Takes Two to Talk* (Toronto, ON: The Hanen Centre).

McDade, A. and McCartan, P., 1998, 'Partnership with Parents.' A pilot project. *International Journal of Language and Communication Disorders*, **33**, 556–561.

Mullen, R., 2007, The state of the evidence: ASHA develops levels of evidence for communication sciences and disorders. *ASHA Leader*, **12** 8–9: 24–25.

Newborg, J., 2005, *Battelle Developmental Inventory*™ (Rolling Meadows, IL: Riverside).

PAUL, R., HERNANDEZ, R., TAYLOR, L. and JOHNSON, K., 1996, Narrative development in late talkers: early school age. *Journal of Speech and Hearing Research*, **39**, 1295–1303.

PAUL, R., LOONEY, S. S. and DAHM, P. S., 1991, Communication and socialization skills at ages 2 and 3 in 'late-talking' young children. *Journal of Speech and Hearing Research*, **34**, 858–865.

PAUL, R., 1996, Clinical implications of the natural history of slow expressive language development. *American Journal of Speech–Language Pathology*, **5**, 5–21.

RESCORLA, L., 1989, The Language Development Survey: a screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders*, **54**, 587–599.

RESCORLA, L., 2009, Age 17 language and reading outcomes in late-talking toddlers: support for a dimensional perspective on language delay. *Journal of Speech, Language and Hearing Research*, **52**, 16–30.

RESCORLA, L. and ACHENBACH, T. M., 2002, Use of the Language Development Survey (LDS) in a national probability sample of children 18 to 35 months old. *Journal of Speech, Language and Hearing Research*, **45**, 733–743.

RESCORLA, L. and FECHNAY, T., 1996, Mother–child synchrony and communicative reciprocity in late-talking toddlers. *Journal of Speech and Hearing Research*, **39**, 200–208.

RESCORLA, L. and GOOSSENS, M., 1992, Symbolic play development in toddlers with expressive specific language impairment (SLI-E). *Journal of Speech and Hearing Research*, **35**, 1290–1302.

RESCORLA, L., ROBERTS, J. and DAHLSGAARD, K., 1997, Late talkers at 2: outcome at age 3. *Journal of Speech, Language, and Hearing Research*, **40**, 556–566.

REYNELL, J. K. and GRUBER, C. P., 1990, *Reynell Developmental Language Scales* (Los Angeles, CA: Western Psychological Services).

ROBERTS, J. E., BURCHINAL, M. and DURHAM, M., 1999, Parents' report of vocabulary and grammatical development of African American preschoolers: child and environmental associations. *Child Development*, **70**, 92–106.

ROBERTSON, S. and 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.

ROYAL COLLEGE OF SPEECH & LANGUAGE THERAPISTS (RCSLT), 2005, Clinical Guidelines (available at: http://www.rcslt.org/resources/clinicalguidelines) (accessed on 1 July 2009).

SCARBOROUGH, H. S. and DOBRICH, W., 1990, Development of children with early language delay. *Journal of Speech and Hearing Research*, **33**, 70–83.

SCHLOSSER, R. W. and O'NEIL-PIROZZI, T. M., 2006, Problem formulation in evidence-based practice and systematic reviews. *Contemporary Issues in Communication Science and Disorders*, **33**, 5–10.

SPARROW, S. S., CICCHETTI, D. V. and BALLA, D. A., 2008, *Vineland Adaptive Behavior Scales* (San Antonio, TX: Pearson).

THAL, D. J., TOBIAS, S. and MORRISON, D., 1991, Language and gesture in late talkers: a 1-year follow-up. *Journal of Speech and Hearing Research*, **34**, 604–612.

TURNER, H. M. and BERNARD, R. M., 2006, Calculating and synthesizing effect sizes. *Contemporary Issues in Communication Science and Disorders*, **33**, 42–55.

VAN KLEECK, A., GILLAM, R. B. and DAVIS, B., 1997, Second Opinion. When is 'watch and see' warranted? A response to Paul's 1996 article, 'Clinical implications of the natural history of slow expressive language development'. *American Journal of Speech–Language Pathology*, **6**, 34–39.

WHITEHURST, G. J., FISCHEL, J. E., LONIGAN, C. J. and VALDEZ-MENCHACA, M. C., 1991, Treatment of early expressive language delay: if, when, and how. *Topics in Language Disorders*, **11**, 55–68.

WILCOX, M. J., KOURI, T. and CASWELL, S. B., 1991, Early language intervention: a comparison of classroom and individual treatment. *American Journal of Speech–Language Pathology*, **1**, 49–62.

ZIMMERMAN, I., STEINER, V. and POND, R., 1997, *Preschool Language Scale—3* (San Antonio, TX: Harcourt Assessment).

ZUBRICK, S. R., TAYLOR, C. L., RICE, M. L. and SLEGERS, D. W., 2007, Late language emergence at 24 months: an epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language, and Hearing Research*, **50**, 1562–1592.