RUNNING HEAD: Systematic review of speech interventions

A systematic review and classification of interventions for speech sound disorder in preschool children

Yvonne Wren[1,2], Sam Harding[1], Juliet Goldbart[3], Sue Roulstone[1,4]

1. Bristol Speech and Language Therapy Research Unit, North Bristol NHS Trust, Bristol, UK
2. University of Bristol, UK
3. Manchester Metropolitan University, UK
4. University of the West of England, UK

Corresponding author: Yvonne Wren, Bristol Speech and Language Therapy Research Unit, Pines and Steps, Southmead Hospital, Westbury-on-Trym, Bristol BS10 5NB, UK.

Yvonne.wren@speech-therapy.org.uk

Tel: +44 117 4143951

Abstract

Purpose: To systematically review the evidence for interventions for speech sound disorder (SSD) in preschool children within a classification of intervention procedures.

Method: Relevant search terms were used to identify evaluations of intervention with the following inclusion criteria: participants were aged between 2 years and 5 years, 11 months; participants exhibited speech, language and communication needs; and a primary outcome measure of speech was used. Studies that met inclusion criteria were quality appraised. Those which were judged as high quality were classified based on the procedures used in the intervention to effect change in a child's speech.

Results: The final review included 26 studies. Case series was the most common research design. Cognitive-linguistic and production approaches to intervention were the most frequently reported but the highest graded evidence was for three studies within the auditory-perceptual and integrated categories.

Conclusions: The evidence for intervention for preschool children with SSD is focused on seven out of 11 subcategories of interventions. Although all of the studies included in the review were good quality studies, they mostly represented lower graded evidence. Higher graded studies are needed to understand clearly the strength of evidence for different interventions.

Introduction

Speech sound disorder (SSD) is a high prevalence condition in preschool children (Broomfield & Dodd, 2004; Eadie, Morgan, Ukoumunne, Ttofari Eecen, Wake, & Reilly, 2015; McLeod & Harrison, 2009; Shriberg, Tomblin, & McSweeny, 1999). In response to this, a number of interventions have been developed which vary in the method used to achieve change in a child's speech or the way in which targets for intervention are selected (Baker & McLeod, 2011).

To date, there have been a number of systematic literature reviews that have examined the effectiveness of these interventions for children with SSD across the age range. Some of the reviews were part of a larger and more comprehensive review of speech and language therapy interventions for children with speech and language delay or disorder (Law, Garret, & Nye, 2003; Law, Lee, Roulstone, Wren, Zeng, & Lindsay, 2012; Law, Roulstone, & Lindsay, 2015) while others have focused specifically on speech (Baker & McLeod, 2011; Murray, McCabe, & Ballard, 2014) or on a specific type of intervention (Lee, Law, & Gibbon, 2009; Lee & Gibbon, 2015; McCauley, Strand, Lof, Schooling, & Frymark, 2009; Morgan & Vogel, 2008). While those focusing on specific interventions revealed a paucity of studies with sufficient strength to provide categorical support for the approaches (specifically, electro-palatography, Non Speech Oro Motor Exercises, and interventions for Childhood Apraxia of Speech), the results of the more extensive reviews were encouraging. Law, Garret, and Nye (2003) included only randomised controlled trials in their review and found convincing support for interventions where the outcome was the child's 'expressive phonology'. Similarly, the review by Law et al., (2012) found that out of 57 interventions included in the review, approximately one third (38%) targeted speech. Evidence for most

of these interventions was at a moderate level (68%), i.e. tested in either a randomised controlled trial or several quasi-experimental studies, whilst for others the evidence was at an indicative level, i.e. they have good face validity and are widely used by clinicians but have limited research evidence which can be generalised to the population concerned.

Baker and McLeod (2011) included a wider range of study designs in their narrative review of evidence based practice for children with SSD. Samples in these studies included participants with concomitant difficulties such as hearing loss, cleft lip and/or palate, or stuttering and spanned an age range of 1;11 to 10;5. They identified a total of 154 studies which described seven different methods for target selection and 46 different approaches to intervention. While a small number of these interventions had been subject to meta-analysis or included in a randomised controlled trial, the majority had been subject to less rigorous investigations such as quasi experimental or non-experimental case studies. Baker and McLeod concluded that more rigorous experimental design is required to enable the relative benefits of any intervention or approach to be determined.

The interpretation of the review findings in a clinical context is challenging and there is little to guide the clinician regarding which intervention to use for children with differing presentations of SSD. The 2006 special edition of Advances in Speech-Language Pathology on 'Jarrod', the 7-year-old boy with SSD, published a range of papers describing different interventions for this child showing the different interventions that could be applied for one individual presentation. However, there was no conclusion regarding which approach might be the most effective or efficient. In practice, it would seem that clinicians tend to favour a small number of interventions for children with a range of presentations (Joffe & Pring, 2008; Roulstone, Wren, Bakapoulou, Goodlad, & Lindsay, 2012).

To summarise, the existing reviews and current ways of grouping interventions are not aligned with clinical practice and therefore not very accessible for practice contexts. However, if the studies could be considered within a practice-based classification of interventions, it might be easier to determine the strength of the evidence in a way which has greater application for clinicians.

Existing classifications of interventions

Current classifications of SSD have linked interventions to the presentation of the child. Shriberg, Fourakis, Hall, Karlson, Lormeier, McSweeny et al., (2010) categorised children according to aetiology, leading to the option of selecting interventions which fit most closely to the child's underlying causal features. Dodd (2005) classified a child's SSD according to the surface level speech presentation, leading to clearly defined intervention approaches which are intended to address the underlying nature of the child's SSD. A third method has considered the needs of parents and produced a family friendly classification which enables clinicians to explain to parents the nature of a child's presenting SSD in terms which can be easily understood (Bowen, 2011).

Many of the interventions reported in the literature, however, do not fit easily into these classifications, and instead cross classification boundaries as they are utilised for SSD with a variety of aetiologies or surface level presentations. An alternative way to classify interventions for SSD is to focus on the nature of the task that the child is required to carry out. This approach has been adopted in descriptions of intervention approaches for SSD such as in Bernthal, Bankson, and Flipsen (2012), Rvachew and Brosseau-Lapré (2012) and Stackhouse and Wells (1997). Typically such approaches have grouped interventions into whether they are primarily targeting: 'input', where the child is required to respond to some auditory stimuli to effect change in their speech; 'storage', where the child is asked to

reflect on their stored representations of words as a means to challenge existing inaccurate representations; or 'output', which require the child to produce speech in response to imitation or some other stimuli.

This type of classification is a useful way of representing the range of specific tasks carried out in intervention for SSD. It does not, however, reflect the multidimensional nature of intervention which includes decisions about targets, dosage, and delivery amongst other factors. Figure 1 from McCauley, Fey and Gillam (in press) presents these dimensions effectively and shows the inter-relationships between the levels. Although the figure was developed to represent dimensions in intervention for grammar, the same basic principles can be applied to intervention for SSD. Moreover, each layer within the figure could be further divided into subcategories describing different approaches to target selection, dosage and delivery for example. It is at the level of 'procedures' however, where tasks which involve 'input', 'storage', and 'output' would most usefully fit.

[Figure one about here]

A model for classification of interventions for SSD

The broad categories of 'input', 'storage', and 'output' go some way to helping categorise interventions for SSD but this does not fully explain the range of procedures included. For example, the procedure 'modelling', one of the examples in Figure 1, could be considered 'input' as the child *hears* the target sounds and words used by adults. But this procedure generally does not require a response from the child and is therefore very different from an input task which requires the child to act on the auditory stimuli in some way.

The basic model of input, storage, output was expanded in work carried out by Wren (2005), using a bottom-up approach from the intervention procedures which are available

and identified as in use by clinicians (Roulstone & Wren, 2001). This was developed further as a result of background work to the development of software for use with children with SSD and the need to identify the most suitable intervention procedure to use as a theoretical basis to the software (Wren & Roulstone, 2008). The model itself is hypothetical and proposes one way of organising types of intervention procedures. It has changed since the original version described in Wren (2005) and is in ongoing development and as such, may continue to change and evolve as new intervention procedures and new evidence become available. Nonetheless, it provides an initial framework that is inclusive of the diverse range of intervention procedures that are available to clinicians. Specific approaches are not named in this model but the mechanism which is used to promote change has been identified and categorised accordingly (Figure 2).

[Figure 2 about here]

The model labels five categories of intervention procedure: environmental, auditory-perceptual, cognitive-linguistic, production and combined. The environmental approach is distinct from the others in that it encompasses intervention approaches which make use of everyday interactions, rather than specific directed activities, to promote change in a child's speech sound system. This would include procedures sometimes described as 'naturalistic intervention' as well as modelling and recasting of a child's spontaneous productions (Camarata, 2010). Auditory perceptual procedures target the child's perceptual skills as a means to induce change in speech output and include activities that aim to increase exposure to the sounds being targeted, as in focused auditory stimulation, and discrimination tasks designed to increase phoneme perception skills (Hodson & Paden, 1991; Rvachew & Brousseau-Lapré, 2010). Cognitive-linguistic procedures engage the child in higher level processing in which the child's awareness of their speech is consciously

addressed and used to promote change, either through confronting a child with their reduced set of contrasts or through increasing awareness of sounds in speech generally. Production procedures aim to effect change through performance of oromotor tasks, guidance on phonetic placement or manner, imitation and drills. Combined procedures are simply those that combine two or more of the other four through profiling of the child's specific needs as in the psycholinguistic approach (Stackhouse & Wells, 1997) or combining procedures into a programme of multiple interventions consistent with a Cycles approach to intervention for example (Hodson & Paden, 1991).

The model does not reflect decisions around target selection though undoubtedly, the decisions regarding procedure and target are related for many interventions. Nor does it attempt to link to aetiology. However, the model makes explicit the procedural aspect of intervention. It is anticipated that this would provide a summary of the current evidence which is more easily accessible to clinicians, and therefore addresses some of the concerns raised in Lancaster, Keusch, Levin, Pring, and Martin, (2010) regarding the incompatibility of research and clinical work.

Aim of the study

The aim of this study was to identify the evidence base for a range of interventions for preschool children with speech sound disorder and to map the procedures used in the interventions onto the model described above. It was part of a larger review of interventions for children with speech and language impairment in preschool children with no concomitant difficulties (Roulstone, Marshall, Powell, Goldbart, Wren, Coad et al., 2015) within the 'Child Talk' research programme, a series of research studies investigating the evidence base for speech and language therapy intervention for preschool children.

Method

The systematic review was guided by the principles outlined in the Cochrane Collaboration methodology (Higgins & Green, 2011), as far as they could be applied to the study methodologies, and built on the review undertaken by Pickstone, Goldbart, Marshall, Rees, and Roulstone (2009). The search strategy described below outlines the larger review carried out for the 'Child Talk' research program and describes how the studies relevant to SSD were identified within this. The systematic review was registered with PROSPERO (CRD42013006369), an international register of prospective systematic reviews.

Search strategy

The search strategy employed three key elements: development of a comprehensive and relevant list of search terms to ensure that all potentially valid studies in relation to interventions for speech and language impairment without concomitant difficulties were returned; exploration of a suitably broad range of databases to capture as many potentially valid studies as possible, including published, unpublished and conference proceedings; and identification of clear inclusion criteria against which to filter potentially valid studies and provide the dataset for analysis. The authors and co-applicants of the 'Child Talk' programme of research (Roulstone et al., 2015) identified a set of search terms based on their previous work in the field (Blackwell, Harding, Babayigit, Roulstone, 2014; Hambly, Wren, McLeod, & Roulstone, 2013; Marshall, Goldbart, Pickstone, & Roulstone, 2011; Pickstone, Goldbart, Marshall, Rees, & Roulstone, 2009; Wren, Hambly, & Roulstone, 2013) . Further potential search terms were identified from key papers. This expertise was augmented through consultation with information specialists. Through an iterative process of identification and discussion, a list of 92 search terms was determined to provide the most appropriate set to capture potentially valid studies. The same process was used to

select appropriate databases to ensure maximum inclusion of published data, unpublished data and conference proceedings.

In line with Booth and Fry-Smith (2003), the PICO model (Population, Intervention, Comparison, Outcome) guided the development of the inclusion criteria. All research design methodologies were considered and therefore the 'Comparison' element of the PICO model was not used to determine eligibility but recorded during data extraction. For inclusion in the larger 'Child Talk' review, studies had to meet the following requirements:

- Population: At least 80% of the sample were required to be within the age range 2 years to 5 years and 11 months at the start of the intervention or at recruitment; children would be diagnosed or considered 'at risk' of speech and language impairment without concomitant difficulties.
- Intervention: An empirical evaluation of an intervention, including randomised controlled trials, experimental and quasi-experimental studies and case studies which included multiple baseline or other systematic manipulation of the intervention.
- Outcomes: At least one of the primary outcome measures of included studies would address speech (articulation/phonology).

Studies were excluded if:

- They related to children whose speech or language appeared to be developing typically with no evidence to suggest that their language was 'at risk'.
- They related to children whose speech or language delays were associated with other developmental or pervasive conditions such as learning difficulties, autism, cleft palate and cerebral palsy.

- The only outcomes were social or behavioural, language or communication related and was not speech.

Search procedure

A combination of 'free text' terms with Boolean operators and truncations was used. Eighteen separate searches were conducted in electronic databases, to identify appropriate studies in articles published from the earliest entries of any of the databases until February 2012. Papers were initially reviewed by title and then by abstract.

Reliability

Two of the authors independently reviewed the titles of ten percent of the papers identified from the initial search of the databases to screen for relevance, removing any studies which did not fit the exclusion and inclusion criteria. There was 100% consensus and the remaining 33,000 references were shared between these two authors and papers were excluded at the title level. This process lead to the retention of 4,574 papers. The abstract review was undertaken by four members of the research team, with two people for each manuscript (one Speech Language Pathologist and one Psychologist). Where disagreements occurred, discussion took place within the team until consensus was reached. Those papers retained at this stage were then reviewed in their entirety in light of the inclusion and exclusion criteria.

The retained papers were further reduced to those that had interventions which related to SSD. Studies were included at this stage if the intervention described in the research was consistent with the definition: "Work that increases the accuracy of speech production or articulation, often focusing on specific sound(s)". Those studies which focused on phonological awareness skills only and did not relate to speech output were excluded. The remaining papers were then subjected to a quality appraisal.

Quality appraisal

The quality appraisal tools used in this review were selected to be relevant to the research designs used in the included studies Two tools were used for this purpose: the Physiotherapy Evidence Database quality assessment tool (PEDro-P, Perdices & Tate, 2009) had a score range of 0-9 and was used to appraise the methodological quality of randomised and non-randomised controlled trials and; and Single Case Experimental Design (SCED) had a score range of 0-10 and was used for single case studies (Tate, McDonald, Perdices, Togher, Schultz, & Savagem, 2008). All appraisers undertook and passed training on PEDro-P and SCED (http://speechbite.com/rating-research-quality/outline-rating-training-program/). Each article was reviewed by at two researchers and if disagreement had occurred it was planned to discuss and reach consensus. This process was not required as agreement on the quality assessment was 100 percent. For both tools, a higher score was associated with greater quality of the methodology applied and reported within the study. In line with previous reviews (Camarinos & Marinko, 2009; Maher, Sherrington, Herbert, Moseley, & Elkins, 2003), a score of six or over was used to identify studies of acceptable quality which would be retained in the review. These studies were then mapped onto the classification of intervention procedures model described above.

Data extraction and synthesis

The process of synthesis consisted of 2 stages. The first stage extracted the characteristics of the studies relating to country, culture, and language/s of the researchers and participants and to study designs categorised using the National Health and Medical Research Council levels of evidence guidelines (NHMRC, 2007). The second stage extracted information on location and agent of intervention, assessment and outcome measures used, number of treatment sessions and a description of the intervention provided.

Subsequently effectives for speech outcomes were calculated were possible. This was undertaken using the Campbell Collaboration effect size calculator (https://www.campbellcollaboration.org/escalc/html/EffectSizeCalculator-SMD-main.php) where data was appropriate and available. Studies using a within-subject pre-post methodology providing sufficient information were assessed using a second online calculation tool (http://www.cognitiveflexibility.org/effectsize/) and single-subject experimental designs were assessed using Improvement Rate Difference (IRD; Parker, Vannest & Davis, 2011)

Results

Figure 3 shows the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses, http://www.prisma-statement.org/ accessed 03/03/2016) flowchart and summary of papers retrieved at each stage of the review. Of the 147 studies matching the inclusion criteria for the Child Talk project as a whole, 55 could be mapped onto the speech theme. Twenty five of these papers, reporting on 26 studies, demonstrated a sufficient level of quality (i.e. obtained a score higher than 6) when assessed using the PEDro-P or SCED scale. Of the 30 that did not attain a score of six or more on these measures, 11 were reviewed using PEDro-P and 19 with SCED. The mean average score on these excluded studies were 4 and 3 respectively (median 4 and 3). The most frequent deficits in the randomised and non-randomised controlled studies were: lack of concealment during group allocation and lack of blinding of the assessor who measured at least one key outcome. In the single case experimental studies the top three deficits in reporting were: lack of raw data being reported; assessors not being independent of treatment/intervention; and lack of replication either across subjects, therapists or setting.

Formatted: Font color: Auto

[Figure 3 about here]

Categorisation of studies and reported outcomes

Of the 26 studies retained for inclusion, 18 were undertaken in the US, 4 in Canada, 3 in Australia and 1 in the UK. Fifteen of the studies used a case series design and 3 were case studies. A further 3 studies used a randomised controlled trial design and a further 4 used a between groups design. The 26 studies were categorised according to the procedure used in the intervention using the model in figure 2 (see figure 4). Table 1 details each of the studies in the review and provides summary information on each obtained from the data extraction.

[Figure 4 about here]

Environmental approaches are represented by one study. The study by Yoder, Camarata, & Gardener (2005) was categorised here due to the intervention using recasting and modelling within clinic contexts. This study found a significant positive impact of the intervention on the child's SSD in comparison with standard care.

Within the category of auditory perceptual approaches, the subcategory of phoneme perception approaches were used in two studies (Rvachew, 1994; Rvachew, Nowak, & Cloutier, 2004). Rvachew, Nowak, and Cloutier (2004) used speech sound discrimination tasks in their intervention and found a positive impact of the interventions. The children in the Rvachew (1994) study were randomly allocated to three groups and these children were given worksheet based tasks focused on treatment of misarticulated versions of target words. This study found a positive effect of the intervention. None of the studies in the review were classified under the focused auditory stimulation subcategory.

Cognitive-linguistic approaches were the most commonly reported interventions within the studies in the review. These studies focused on three subcategories of

intervention: 'meaningful minimal contrast' approaches, 'complexity' approaches and 'metalinguistic approaches'. Three studies focused on meaningful minimal contrast (Baker & McLeod, 2004; Dodd & Iacono, 1989; Robb, Bleile, & Yee, 1999) and a further six studies (from five papers) form the evidence base for complexity approaches (Gierut, 1989; 1990; Gierut & Champion, 1999; Gierut, Morrisette, Hughes, & Rowland, 1996; Rvachew & Nowak, 2001). These studies have small samples but suggest a positive impact of the interventions on the children, with one exception where change to the target of intervention was not observed (Gierut & Champion, 1999). No studies were included in the review under the category of metalinguistic approaches.

Studies within the review which came under the category of production were identified within the subcategories of 'oro-motor speech exercises', 'guidance on phonetic placement/manner' and 'imitations and drill'. No studies were categorized under 'oro-motor speech exercises' or 'guidance on phonetic placement/manner. The seven studies within the 'imitations and drill' subcategory all worked on increasing the complexity of articulation in graded steps such as breaking words into constituent sounds and subsequently recombining to form the word (Forrest & Elbert, 2001; Forrest, Elbert, & Dinnsen, 2000; Gierut, 1996; Gierut & Champion, 1999; 2001; Gierut, Morrisette, Hughes, & Rowland, 1996; Winner & Elbert, 1988). Five of these studies showed an improvement in the intervention group (Forrest & Elbert, 2001; Forrest et al., 2000; Gierut & Champion, 2000; 2001; Gierut & Morrisette, 1996), while in two studies there was no statistical impact of the intervention on the child's speech output (Gierut, 1996; Winner & Elbert, 1988).

'Integrated' approaches to intervention were represented by studies within the subcategories of 'combined' approaches and 'unspecified'. Combined approaches were adopted in five studies included in the review (Almost & Rosenbaum, 1998; Hart &

Gonzalez, 2010; McIntosh & Dodd, 2008; Saben & Ingham, 1991; Wolfe, Presley, & Mesaris, 2003). The studies used a combination of activities and strategies as interventions, described as being targeted at the individual child's needs or as routine one-to-one therapy. The studies provide mixed evidence for this approach, with only Almost and Rosenbaum (1998) showing a positive effect of active therapy. Unspecified approaches were used in the Glogowska, Roulstone, Enderby, & Peters (2000) study where no differences overall were found on the phonology score between control children and those receiving standard treatment. However, on a secondary outcome, a significantly greater proportion of children receiving standard treatment improved their phonology such that they no longer satisfied the original phonology eligibility criteria for the trial.

Delivery of intervention

All studies included in the review used interventions that were delivered by speech language pathologists. Several studies did not provide information on the number and length of intervention sessions, however where they did, the range was from three to 67 sessions lasting between 30 and 60 minutes.

Assessment measures used

Speech measurement was carried out for one or more of three purposes: to confirm eligibility for participation in the study; to identify targets for intervention; or to measure change in response to intervention (outcome measure). Three studies also measured change in speech perception (Wolfe, Presley, & Mesaris, 2003; Rvachew, Nowak, & Cloutier, 2004; Rvachew, 1994). Speech output was collected using published assessments (Hart & Gonzalez, 2010; McIntosh & Dodd, 2008; Rvachew & Nowak, 2001), confrontation picture naming tasks devised for the study (Saben & Ingham, 1991; Winner & Elbert, 1988), and spontaneous continuous speech samples (Dodd & Iacono, 1989; Hart & Gonzalez, 2010;

Saben & Ingham, 1991; Rvachew, 1994; Rvachew, Nowak & Cloutier, 2004; Winner & Elbert, 1988; Yoder et al., 2005). In all studies, reliability of the transcriptions was reported using point-to-point agreement for two transcribers, from between 20 to 100 percent of data collected. Some studies used a combination of two or three approaches to collecting speech samples. Several studies also used picture naming as part of a probe testing protocol (Baker & McLeod, 2004; Forrest, Elbert, & Dinnsen, 2000; Forrest & Elbert, 2001; Gierut, 1996, 1990, 1989; Gierut et al., 1996; Gierut & Champion, 2000, 1999; Robb, Bleile, & Yee, 1999; Saben & Ingham, 1991; Wolfe, Presley, & Mesaris, 2003).

In terms of analysis of the speech samples collected, those studies which included published assessments within their assessment protocol typically used the analysis procedures which accompanied those tools. These included process analysis (Assessment of Phonological Processes-Revised, Hodson, 1986, 2000), phonemic or phonetic inventories, (Productive Phonological Knowledge Profile, Gierut, Elbert, & Dinnsen, 1987), percentage phonemes/consonants/vowels correct (Diagnostic Evaluation of Articulation and Phonology, Dodd, Zhu, Crosbie, Holm, & Ozanne, 2002; Shriberg & Kwiatkowski, 1982), and accuracy of production (Goldman-Fristoe Test of Articulation, Goldman & Fristoe, 2000). Where spontaneous speech samples, confrontation picture naming or probe lists were used, a number of analyses were carried out, as detailed in table 1.

[table 1 about here]

Discussion

This systematic review of the literature has considered the evidence for a range of interventions for preschool children with SSD within a model in which interventions were classified based on the nature of the procedures used to effect change. A total of 55 papers were identified based on clearly defined search criteria. Following quality appraisal, 25

papers reporting 26 studies were considered robust enough to be included in the final review. These 26 studies were then mapped onto the model of interventions according to the description of the procedures within each paper.

Description of the review and limitations of the study

The systematic review had a specific remit to look at the evidence base related to intervention for SSD with preschool children (2;00 – 5;11). Studies with 20% or more of children outside the specified age range would not have been included. Given that interventions for children with SSD are often carried out when the child is school-aged, it is likely that some important studies relating to intervention for SSD in general will have been omitted.

While some previous reviews have limited their enquiry to children with phonological problems only (Baker & McLeod, 2011), this review included any study which targeted increased accuracy of speech production or articulation, encompassing both phonological and speech motor interventions. This was important given the aim of synthesizing the evidence for clinicians who will be faced with a broad spectrum of children with SSD in practice (Broomfield & Dodd, 2004; Shriberg, Lewis, Tomblin, McSweeny, Karlsson, & Scheer, 2005). It did not however include interventions which focused on prosodic skills or speech perception or other underlying speech processing skills unless these were included alongside a measurement of speech output.

The review included a range of research designs and did not limit itself to RCTs though most were at level III of the NHMRC Evidence Hierarchy (NHMRC, 2007) and therefore were either pseudorandomised controlled trials or comparative studies with or without concurrent controls. Previous reviews (Law, Garrett, & Nye, 2003; Lee & Gibbon, 2015; Morgan & Vogel, 2008) have followed more restrictive criteria with regards to study

design. However, in order to reflect the growing evidence base and the potential for lower graded studies to develop into larger studies with more robust research designs, the decision was made to include studies with a lower level of evidence, as defined by NHMRC (2007). This allowed an investigation of the current level of evidence for interventions and a clear picture regarding what is required to take the evidence forward. As a counter to the inclusion of studies with lower graded evidence, the quality appraisal tools were used to identify studies with the most robust operationalisations of these designs and reporting processes.

The data extraction process revealed that many studies did not report complete data regarding dosage but where these were reported, there was a wide range in the number of sessions provided (three to 67). However, there were no clear patterns to the dosage provided within the categories and subcategories of interventions. Rather, where it was reported, a wide range of number, frequency and duration of intervention sessions were offered. A lack of consistency in the provision of intervention makes it harder to compare across interventions and to determine the relative benefit of each.

With regards to measuring outcomes, a range of tools were used to assess speech output including published assessments, picture naming tasks and spontaneous continuous speech samples. As with dosage, there were no clear patterns within the categories and subcategories with regard to outcome data collection and analysis. Thus a narrative synthesis has been used rather than attempt a meta-analysis where the measures differed widely. The exception to this was the subcategories of imitation and drill and complexity approaches which both relied heavily on probe word lists to test outcomes. However, these studies were predominantly carried out by two groups of researchers which may explain the

tendency towards the same measurement tools rather than indicating consensus across research groups in favour of any particular measure.

The model for classifications of interventions for SSD

The classification model used as the basis for the review was developed using a bottom-up approach that considered the procedures used to effect change across a range of interventions for SSD, not limited to those in the review. The model assumed five main categories which distinguished the intervention procedures based on whether they used techniques within daily activities (environmental), or attempted to effect change through targeting input skills (auditory-perceptual), storage of representations (cognitive-linguistic) or output skills (production) or some combination of these (combined). The subcategories within attempts to capture more precisely what is being asked of the child in order to effect change. An exhaustive list of possibilities is not presented however and the model will undoubtedly evolve as new intervention procedures emerge and the evidence base grows.

Mapping the evidence to the model

The majority of studies in the review focused on just three of the eleven subcategories of the model: imitations and drill (7 studies), meaningful minimal contrasts (3 studies) and complexity (6 studies). The remaining studies covered a further 4 categories/subcategories. Thus no studies were identified for 4 of the subcategories of the model. Rather than suggesting that those with no studies in the review are ineffective, the more accurate conclusion is that currently, there is no strong evidence to support these intervention procedures with preschool aged children. This is consistent with the findings of reviews which have focused on other specific interventions for speech such as electropalatography and interventions for apraxia of speech (Lee, Law, & Gibbon, 2009; Lee & Gibbon, 2015; McCauley, Strand, Lof, Schooling, & Frymark, 2009; Morgan & Vogel, 2008).

Some degree of supporting evidence was identified for seven of the intervention categories and subcategories in the model. These covered all of the five main categories and a range of subcategories: environmental approaches; phoneme perception; guidance on phonetic placement/manner; imitations and drill; contrasts; complexity; and combined approaches. The number of quality studies varied across these subcategories, from just one each for 'environmental' and 'guidance on phonetic placement/manner' to seven for imitation and drill. Three subcategories in the model, imitations and drill, contrasts and complexity, were supported by a number of good quality studies but the level of evidence represented in each of these studies is low based on the NHMRC classification of levels of evidence (NHRMC, 2007). Across these three subcategories of intervention procedure, the highest graded study was at level III-2 – a comparative study with concurrent controls. This is comparable with a classification of indicative evidence based on the 'What Works' database of interventions (Law, Roulstone, & Lindsay, 2015). The fact that there are studies with higher grade evidence adds credence to the findings for the category or subcategory as a whole but there is still a need for more studies utilising a higher level of evidence methodologies to strengthen the evidence base for these types of intervention. This fits with the findings of Baker and McLeod (2011) who commented on the need for higher levels of scientific rigour and the importance of replication research to build on the findings of lower graded studies.

Higher grade evidence was identified in the review for three studies: one using phoneme perception (Rvachew, Nowak, & Cloutier, 2004), one which used a combined approach (Almost & Rosenbaum, 1998); and a third where the intervention procedure was unspecified (Glogowska et al., 2000). All three studies were randomised controlled trials with large sample sizes relative to most of the other studies (34, 26 and 26 respectively).

Given that a range of interventions were used within these three studies, this suggests that there is agreement that a variety of approaches to intervention can be effective for children with SSD (Lancaster et al., 2010).

Conclusion

To summarise, there is evidence to support certain types of intervention for preschool children with SSD. Whilst there are more studies to support those interventions working on imitation and drill procedures or using cognitive-linguistic approaches, the stronger evidence is linked to working on phoneme perception, combined and unspecified approaches to intervention. However, it is still not possible to provide clear guidance regarding which approach to use with individual preschool children with SSD. The work so far has been invaluable in establishing a preliminary evidence base in which different intervention types have been trialled and explored through small scale studies. As well as providing initial evidence, these studies have enabled researchers to explore the facets of a particular approach to intervention. It has allowed for the understanding of issues relating to delivery which can inform both clinical practice and further investigations. There is a need now for research activity to advance the knowledge base through the use of higher graded methodological studies which will provide more robust information on which approaches or combination of approaches are most suitable to use with this client group.

Acknowledgements

'Child Talk' - What Works presents independent research commissioned by the National Institute for Health Research (NIHR) under its Programme Grants for Applied Research funding scheme (RP-PG-0109-10073).

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.

Baker, E., & McLeod, S. (2004). Evidence-based management of phonological impairment in children. *Child Language Teaching and Therapy*, *20*, 261–285.

Baker, E., & McLeod, S.M. (2011). Evidence-based practice for children with speech sound disorder: part 1 narrative review. *Language, Speech and Hearing Services in Schools, 42,* 102-139.

Bernthal, J.E., Bankson,N.W., & Flipsen, P. (2012) *Articulation and Phonological Disorders: Speech Sound Disorders in Children* (6th ed.)*.* New York, NY: Pearson.

Blackwell, A.K.M., Harding, S., Babayiğit, S., & Roulstone, S. (2014). Characteristics of parent–child interactions: A systematic review of studies comparing children with primary language impairment and their typically developing peers. *Communication Disorders Quarterly, 36,* 67-78*.*

Booth, A., & Fry-Smith, A. (2003). Developing the research question. In Topfer, L., Auston, I., (eds). *Etext on Health Technology Assessment (HTA) Information Resources*. National Information Center on Health Services and Health Care Technology.

Bowen, C. (2011). *Classification of children's speech sound disorders.* Retrieved from www.speech-language-therapy.com [accessed 20/7/15].

Broomfield, J., & Dodd, B. (2004). Children with speech and language disability: caseload characteristics. *International Journal of Language and Communication Disorders, 39,* 303-324.

Camarata, S. (2010). Naturalistic Intervention for Speech Intelligibility and Speech Accuracy. In A. L. Williams, S. McLeod, & R. J. McCauley (Eds.), *Interventions for speech sound disorders in children*. Baltimore, MD: Paul H. Brookes.

Camarinos, J., & Marinko, L. (2009). Effectiveness of manual physical therapy for painful shoulder conditions: a systematic review. *The Journal of Manual & Manipulative Therapy*, *17*, 206–215.

Dodd, B. (2005). *Differential diagnosis and treatment of children with speech disorder (2nd Edition).* Chichester: Whurr.

Dodd, B., & Iacono, T. (1989). Phonological disorders in children: changes in phonological process use during treatment. *British Journal of Disorders of Communication*, *24*, 333–352.

Dodd, B., Zhu, H., Crosbie, S., Holm, A., & Ozanne, A. (2002). *Diagnostic Evaluation of Articulation and Phonology (DEAP).* Psychology Corporation, London.

Eadie, P., Morgan, A., Ukoumunne, O.C., Ttofari Eecen, K., Wake, M., & Reilly, S. (2015). Speech sound disorder at 4 years: prevalence, comorbidities, and predictors in a community cohort of children. *Developmental Medicine and Child Neurology, 57,* 578-584.

Forrest, K., & Elbert, M. (2001). Treatment for phonologically disordered children with variable substitution patterns. *Clinical Linguistics and Phonetics*, *15*, 41–45.

Forrest, K., Elbert, M., & Dinnsen, D. (2000). The effect of substitution patterns on phonological treatment outcomes. *Clinical Linguistics and Phonetics*, *14*, 519–531.

Gierut, J. A. (1989). Maximal Opposition Approach to Phonological Treatment. *Journal of Speech and Hearing Disorders*, *54*, 9–19.

Gierut, J. A. (1990). Differential learning of phonological oppositions. *Journal of Speech and Hearing Research*, *33*, 540–549.

Gierut, J. A. (1996). An experimental test of phonemic cyclicity. *Journal of Child Language, 23*, 81–102.

Gierut, J. A. (2001). Complexity in phonological treatment: clinical factors. *Language, Speech, and Hearing Services in Schools, 32* 229-241.

Gierut, J. A., & Champion, A. H. (1999). Interacting error patterns and their resistance to treatment. *Clinical Linguistics and Phonetics*, *13*, 421–431.

Gierut, J. A., & Champion, A. H. (2001). Syllable onsets II: three-element clusters in phonological treatment. *Journal of Speech, Language, and Hearing Research*, *44*, 886–904.

Gierut, J. A., Elbert, M., & Dinnsen, D. A. (1987). A functional analysis of phonological knowledge and generalization learning in misarticulating children. *Journal of Speech and Hearing Research*, *30,* 462–479.

Gierut, J. A., Morrisette, M. L., Hughes, M. T., & Rowland, S. (1996). Phonological treatment efficacy and developmental norms. *Language, Speech, and Hearing Services in Schools*, *27*, 215–230.

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* (7266), 923.

Goldman, R., & Fristoe, M. (1969). *Goldman-Fristoe Test of Articulation (GFTA).* Windsor: NFER-Nelson.

Goldman, R., & Fristoe, M. (2000). *Goldman-Fristoe Test of Articulation-Second Edition (G-FTA-2).* New York, NY: Pearson.

Hambly, H., Wren, Y., McLeod, S., & Roulstone, S. (2013). The influence of bilingualism on speech production: a systematic review. *International Journal of Language and Communication Disorders,* 48, 1–24.

Hart, S., & Gonzalez, L. (2010). The effectiveness of using communication-centered intervention to facilitate phonological learning in young children. *Communication Disorders Quarterly*, *32*, 13–25.

Higgins, J., & Green, S. (2011). *Cochrane handbook for systematic reviews of interventions.* (5.1.0 ed.). The Cochrane Collaboration. Retrieved January 28th 2016 from http://www.cochrane.org/training/cochrane-handbook

Hodson, B. W. (1986). *The Assessment of Phonological Processes—Revised.* Austin, TX: Pro-Ed.

Hodson, B. W. (2000). *The Hodson Assessment of Phonological Processes—Revised.* Austin, TX: Pro-Ed.

Hodson, B., & Paden, E. (1991). *Targeting intelligible speech: A phonological approach to remediation. (2nd ed.)*. Texas: Pro-Ed.

Joffe, V., & Pring, T. (2008). Children with phonological problems: a survey of clinical practice. *International Journal of Language and Communication Disorders*. 43, 154–164.

Lancaster, G., Keusch, S., Levin, A., Pring, T., & Martin, S. (2010). Treating children with phonological problems: Does an eclectic approach to therapy work? *International Journal of Language and Communication Disorders, 45*, 174-181.

Law, J., Garrett, Z., & Nye, C. (2003). Speech and language therapy interventions for children with primary speech and language delay or disorder. *Cochrane Database Systematic Review*. CD004110.

Law, J., Lee, W., Roulstone, S., Wren, Y., Zeng, B., & Lindsay, G. (2012). *What works: interventions for children and young people with speech, language and communication needs* (No. DFE-RR247-BCRP10), Better communication research programme. Department for Education, London: UK.

Law, J., Roulstone, S., & Lindsay, G. (2015). Integrating external evidence of intervention effectiveness with both practice and the parent perspective: development of "What Works" for speech, language, and communication needs. *Developmental Medicine and Child Neurology,* 57, 223–228.

Lee, A. S. Y., & Gibbon, F. E. (2015). Non-speech oral motor treatment for children with developmental speech sound disorders. *Cochrane Database of Systematic Reviews Issue 3.* Art. No.: CD009383.

Lee, A. S. Y., Law, J., & Gibbon, F. E. (2009). Electropalatography for articulation disorders associated with cleft palate. *Cochrane Database Systematic Reviews Issue 3.* Art. No.: CD006854.

Maher, C. G., Sherrington, C., Herbert, R. D., Moseley, A. M., & Elkins, M. (2003). Reliability of the PEDro scale for rating quality of randomized controlled trials. *Physical Therapy*, *83*, 713–721.

Marshall, J., Goldbart, J., Pickstone, C., & Roulstone, S. (2011). Application of systematic reviews in speech-and-language therapy. *International Journal of Language and Communication Disorders*. 46, 261–272.

McCauley, R.J., Fey, M., Gillam, R. (in press) *Treatment of Language Disorders in Children.* Baltimore, MD: Brookes Publishing.

McCauley, R.J., Strand, E., Lof, G.L., Schooling, T., & Frymark, T. (2009). Evidence-Based Systematic Review: Effects of Nonspeech Oral Motor Exercises on Speech. *American Journal of Speech and Language Pathology*, *18*, 343–360.

McIntosh, B., & Dodd, B. (2008). Evaluation of Core Vocabulary intervention for treatment of inconsistent phonological disorder: Three treatment case studies. *Child Language Teaching and Therapy*, *24*, 307–327.

McLeod, S., & Harrison, L. (2009). Epidemiology of speech and language impairment in a nationally representative sample of 4- to 5-year-old children. *Journal of Speech, Language and Hearing Research, 52,* 1213-1229.

Morgan, A. T., & Vogel, A.P. (2008). Intervention for childhood apraxia of speech. *Cochrane database of systematic reviews, Issue 3.* Art. No.: CD006278.

Murray, E., McCabe, P., & Ballard, K. (2014). A systematic review of treatment outcomes for children with childhood apraxia of speech. *American Journal of Speech-Language Pathology, 23,* 486-504.

NHMRC. (2007). *NHMRC standards and procedures for externally developed guidelines*. Canberra: National Health and Medical Research Council. Retrieved January 28th 2016, from http://www.nhmrc.gov.au/guidelines-publications/nh56

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect Size in Single-Case Research : A review of nine nonoverlap techniques. *Behav Modification.* 35(4) 303-322.

Perdices, M., & Tate, R. L. (2009). Single-subject designs as a tool for evidence-based clinical practice: Are they unrecognised and undervalued? *Neuropsychological Rehabilitation*, *19*, 904–927.

Pickstone, C., Goldbart, J., Marshall, M., Rees, A., & Roulstone, S. (2009). A systematic review of environmental interventions to improve child language outcomes for children with or at risk of primary language impairment. *Journal of Research in Special Educational Needs*, *9*, 66–79.

Robb, M. P., Bleile, K. M., & Yee, S. S. L. (1999). A phonetic analysis of vowel errors during the course of treatment. *Clinical Linguistics and Phonetics*, *13*, 309–321.

Roulstone, S. E., Marshall, J. E., Powell, G. G., Goldbart, J., Wren, Y. E., Coad, J., *et al.* (2015). Evidence-based intervention for preschool children with primary speech and

language impairments: Child Talk - an exploratory mixed-methods study. *Programme Grants Appl Res,* 3(5).

Roulstone, S.E., & Wren, Y.E. (2001) Investigation of theoretical models and therapy activities: Phonological difficulties. *International Journal of Language and Communication Disorders, 36 Supp, 441-446.*

Roulstone, S., Wren, Y., Bakopoulou, I., Goodlad, S., & Lindsay, G. (2012). *Exploring interventions for children and young people with speech, language and communication needs: A study of practice* (No. DFE-RR247-BCRP13). London: UK: Department for Education. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/219627/DFE-RR247-BCRP13.pdf

Rvachew, S. (1994). Speech perception training can facilitate sound production learning. *Journal of Speech and Hearing Research*, *37*, 347–357.

Rvachew, S., & Brousseau-Lapré, F. (2010). Speech perception intervention. In A. L. Williams, S. McLeod, & R. J. McCauley (Eds.), *Interventions for speech sound disorders in children* (pp. 295-314). Baltimore, MD: Paul H. Brookes.

Rvachew, S., & Brosseau-Lapré, F. (2012). *Developmental Phonological Disorders Foundation of Clinical Practice.* San Diego, CA: Plural Publishing.

Rvachew, S., & Nowak, M. (2001). The effect of target-selection strategy on phonological learning. *Journal of Speech, Language, and Hearing Research*, *44*, 610–623.

Rvachew, S., Nowak, M., & Cloutier, G. (2004). Effect of phonemic perception training on the speech production and phonological awareness skills of children with expressive phonological delay. *American Journal of Speech-Language Pathology*, *13*, 250–263.

Saben, C. B., & Ingham, J. C. (1991). The effects of minimal pairs treatment on the speech-sound production of two children with phonologic disorders. *Journal of Speech and Hearing Research*, *34*, 1023–1040.

Shriberg, L. D., Fourakis, M., Hall, S. D., Karlsson, H. B., Lormeier, H. L., McSweeny, J. L., et al. (2010). Extensions to the Speech Disorders Classification System (SDCS). *Clinical Linguistics & Phonetics, 24,* 795-824.

Shriberg, L. D., & Kwiatkowski, J. (1982). Phonological Disorders III: A procedure for assessing severity of Involvement. *Journal of Speech & Hearing Disorder,* 47, 256–270.

Shriberg, L. D., Lewis, B. L., Tomblin, J. B., McSweeny, J. L., Karlsson, H. B., & Scheer, A.R. (2005). Toward diagnostic and phenotype markers for genetically transmitted speech delay. *Journal of Speech, Language and Hearing Research, 48*, 834-852.

Shriberg, L. D., Tomblin, J. B., & McSweeny, J. L. (1999). Prevalence of speech delay in 6-year old children and comorbidity with language impairment. *Journal of Speech, Language & Hearing Research*, 42, 1461-1481.

Stackhouse, J., & Wells, B. (1997). *Children's Speech and Literacy Difficulties: A Psycholinguistic Framework.* London: Whurr.

Tate, R. L., McDonald, S., Perdices, M., Togher, L., Schultz, R., & Savage, S. (2008). Rating the methodological quality of single-subject designs and n-of-1 trials: introducing the Single-Case Experimental Design (SCED) Scale. *Neuropsychological Rehabilitation*, *18*, 385–401.

Winner, M., & Elbert, M. (1988). Evaluating the treatment effect of repeated probes. *Journal of Speech and Hearing Disorders*, *53*, 211–218.

Wolfe, V., Presley, C., & Mesaris, J. (2003). The importance of sound identification training in phonological intervention. *American Journal of Speech-Language Pathology*, *12*, 282–288.

Wren, Y. (2005). *An evaluation of the use of computers in phonology therapy.* Unpublished PhD thesis.

Wren, Y., Hambly, H., & Roulstone, S. (2013). A review of the impact of bilingualism on the development of phonemic awareness skills in children with typical speech development. *Child Language Teaching and Therapy, 29*, 11–25.

Wren, Y., & Roulstone, S. (2008). A comparison between computer and tabletop delivery of phonology therapy. *International Journal of Speech-Language Pathology, 10*, 346-363.

Yoder, P., Camarata, S., & Gardner, E. (2005). Treatment effects on speech intelligibility and length of utterance in children with specific language and intelligibility impairments. *Journal of Early Intervention, 28*, 34–49.

Table 1: Summary of studies from systematic review

Study Author(s) Country of origin	No of child participants (number of children in each group, if applicable)	Age range (months)	Study Deign (Type of Evidence†)	No. of therapy sessions/ Agent of Delivery	Length of each session	Frequency of sessions	Duration of intervention	Type of speech sampled	Analysis used to measure change	PEDro-P/SCED score	Effect Size Cohen d unless otherwise specified
Environmental											
Yoder, P., Camarata, S., & Gardner, E. (2005). USA	52 (26, 26)	Group 1 – average 44.3 Group 2 – average 43.2	Randomised (Type II)	Group 1 – Control 0; Group 2 (treatment group) 72/ SLP	30 minutes	Three times per week	6 months	Spontaneous speech	Percentage intelligible utterance PVC* PCC*	PEDro-P 7	49 (taken directly from article)
Auditory Perceptual: Phoneme Perception											
Rvachew, S. (1994). CA	27 (10, 9, 8)	Group 1 – average 53.4 Group 2 – average 53.6 Group 3 – average 51.5	Randomised (Type II)	6 / SLP	45 minutes	Weekly	6 – 11 weeks	Word identification Single word naming	Percentage correct word identification Number of single words produced correctly	PEDro-P 6	0.0092
Rvachew, S., Nowak, M., & Cloutier, G. (2004). CA	34 (17, 17)	Group 1 – average 52.88 Group 2 – average 50.29	Randomised (Type II)	16 (in addition to their regular therapy)/ SLP	15 minutes	Weekly	4.73 months	Conversation	PCC*	PEDro-P 6	0.8316
Cognitive-Linguistic: Meaningful Minimal Contrast											
Baker, E., &	2	Subject 1	Single	1 – 12	45	Twice	1 – 6 weeks	Probe	Percentage	SCED 7	0.001¥

McLeod, S. (2004). AUS		– 57 Subject 2 – 52	Subject studies – Case Report, A-B, Multiple Baseline Design (Type IV)	2 – 32 / SLP	minutes	weekly	2 – 16 weeks	conversation	correct production of trained cluster		
Dodd, B., & Iacono, T. (1989). AUS	7	36 – 57	Case Series Pre-Post Intervention Design (Type IV)	3 – 40 / SLP	Not available	Weekly	Average 23.6 weeks	Spontaneous speech (during play)	PCC* Phoneme Inventory Process analysis RIU*	SCED 6	-1.362¥
Robb, M. P., Bleile, K. M., & Yee, S. S. L. (1999). USA	1	48	Case study - Single Subject Pre-Post Intervention Design (Type IV)	20 / SLP	45 minutes	Twice weekly	10 weeks	Speech sample Probe list	Percentage accuracy Vowel inventory PVC* Acoustic analyses of vowels (duration, fundamental frequency)	SCED 6	Insufficient data
Cognitive-Linguistic: Complexity Approaches											
Gierut, J. A. (1989). USA	1	55	Case Study, Pre-Post Intervention Design (Type IV)	23 / SLP	30 minutes	Twice weekly	11.5 weeks	Probe lists	Percentage accurate production of target phonemes	SCED 8	Insufficient data
Gierut, J. A. (1990). USA	3	49 – 58	Alternating treatment design – Multiple Baseline	Not available / SLP	60 minutes	Three times a week	Not available	Probe list	Percentage accuracy correct on probe list	SCED 9	Figures are of insufficient resolution

			Design (Type III-3)								to extract data
Gierut, J. A., & Champion, A. H. (1999). USA	2	48 – 56	Single Subject studies – Multiple Baseline Design (Type III-3)	12 / SLP	60 minutes	Three times per week	Approximately 7 weeks	Probe	Percentage accuracy correct on probe list	SCED 6	Figures are of insufficient resolution to extract data
Gierut, J. A., Morrisette, M. L., Hughes, M. T., & Rowland, S. (1996). STUDY 1 USA	3	43 – 66	Single Subject studies – alternating treatment design (Type III-2)	Up to 19 / SLP	60 minutes	Three times per week	Not available	Probe list	Percentage accuracy correct on probe list	SCED 7	Figures are of insufficient resolution to extract data
Gierut, J. A., Morrisette, M. L., Hughes, M. T., & Rowland, S. (1996). STUDY 2 USA	6	41 – 66	Single Subject studies – Multiple Baseline Design (Type III-2)	Not available / SLP	Not available	Not available	Not available	Probe list	Percentage accuracy correct on probe list	SCED 7	Figures are of insufficient resolution to extract data
Rvachew, S.-, & Nowak, M. (2001). Ca	48 (24, 24)	Group 1 – average 51.46 Group 2 – average 49.63	Randomised (Type II)	12 / SLP	Not available	Weekly	12 weeks in two blocks of 6	PPKP* Conversation	PPKP* PCC*	PEDro-P 6	-0.1194

Production: Imitation & drills											
Forrest, K., & Elbert, M. (2001). USA	4	59 – 63	Single Subject studies – Multiple Baseline Design (Type III-2)	Not available / SLP	45 minutes	Twice weekly	Not available	Probe list	PCC* for target phonemes	SCED 6	Insufficient data
Forrest, K., Elbert, M., & Dinnsen, D. A. (2000). USA	10 (5, 5)	40 – 54	Comparative studies – Therapy approach (Type III-3)	Not available / SLP	Not available	Fortnightly	Not available	Probe	Percentage accuracy correct of probe	SCED 8	Insufficient data
Gierut, J. A. (1996). USA	7	40 – 68	Single Subject studies – Multiple Baseline Design (Type III-2)	Not available / SLP	60 minutes	Three time per week	Average 18 weeks	Probe	Change in phonemic inventory	SCED 6	Insufficient data
Gierut, J. A., & Champion, A. H. (2000). USA	1	53	Single Subject studies – Multiple Baseline Design (Type III-2)	19 / SLP	60 minutes	Three times per week	19	Probe list	Percentage accuracy correct on probe list	SCED 6	Insufficient data
Gierut, J. A., & Champion, A. H. (2001). USA	8	40 – 75	Single Subject studies – Multiple Baseline Design	Not available / SLP	60 minutes	Three times per week	Not available	Probe list	Percentage accuracy correct on probe list	SCED 9	IRD^{π} - between 84 & 100%

			(Type III-2)								
Gierut, J. A., & Morrisette, M. L. (1996). USA	2	47 – 62	Single Subject studies – Multiple Baseline Design (Type III-2)	Not available / SLP	60 minutes	Three time per week	Average of 16 weeks	Probes	Phoneme inventory	SCED 6	Insufficient data
Winner, M., & Elbert, M. (1988). USA	4	46 – 68	Single subject studies – Multiple Baseline Design (Type III-2)	25 / SLP	30 minutes	Three times per week	8 weeks	Speech sample Probe list Spontaneous speech (picture description)	Percentage correct scores of target sounds	SCED 7	IRD$^{\pi}$ - Between 50 & 100%
Integrated Approaches: Combined											
Almost, D., & Rosenbaum, P. (1998). CA	26 (13, 13)	33 – 61	Group studies - Randomised (Type II)	14 – 29 / SLP	30 minutes	Twice weekly	7 – 15 weeks	GFTA* APP-R* Standardised test of single words Conversational speech	Single words No of errors PCC*	PEDro-P 9	0.0004
Hart, S., & Gonzalez, L. (2010). USA	3	43 – 59	Single Subject studies – Multiple Baseline Design (Type III-2)	12 / SLP	30 minutes	Twice a week	6 weeks	HAPP-R 3* Spontaneous speech sample	Process analysis Percentage sample correct	SCED 8	IRD$^{\pi}$ - between 0 & 100%
McIntosh, B., & Dodd, B. (2008). AUS	3	36 - 45	Single Subject Pre-Post Intervention Design (Type IV)	Between 12 and 38 / SLP	30-40 minutes	Twice Weekly	Between 6 and 19 weeks (average 12.8 weeks)	Single word naming test (DEAP* phonology subtest) Connected	PVC* PCC* PPC* Percentage inconsistency	SCED 6	-42.187$^{¥}$

Formatted Table

								speech task (DEAP*) Repeated production of words (DEAP* – inconsistency subtest)			
Saben, C. B., F & Ingham, J. C. (1991). USA	2	Subject 1 -52 Subject 2 – 45	Single Subject studies – Multiple Baseline Design (Type III-2)	1 – 67 2 – 32 / SLP	Not available	Not available	1 – 9 months 2 – 4½ months	Probe list (spontaneous picture naming)	Percentage use of individual targeted phonemic processes	SCED 8	Insufficient data
Wolfe, V., Presley, C., & Mesaris, J. (2003). USA	9 (4, 5)	Group 1 – 47 – 55 Group 2 – 41 – 50	Comparative studies – Randomised Therapy approach (Type II)	Average 11 / SLP	30 minutes	Twice weekly	One academic quarter	Probe list	Accuracy of production Sound identification	PEDro-P 6	-0.3634
Integrated Approaches: Unspecified											
Glogowska, M., Roulstone, S., Enderby, P., & Peters, T. J. (2000). UK	159 (71, 84)	Group 1 – 18 – 42 Group 2 – 24 – 42	Comparative studies – Randomised Therapy approach (Type II)	Average 6.2 hours / SLP	Average of 47 minutes	Once a month	Average of 8.4 months	Unclear	Error rate	PEDro-P 8	0.0477

†NHMRC (2007) Evidence Hierarchy: Designations of 'levels of evidence' according to type of research question.

*APP-R: the Assessment of Phonological Processes – Revised (Hodson, 1986); DEAP: Diagnostic Evaluation of Articulation and Phonology, (Dodd, Zhu, Crosbie, Holm, & Ozanne, 2002); GFTA: Goldman Fristoe Test of Articulation (Goldman & Fristoe, 1969, 2000); HAPP-R: the Assessment of Phonological Processes – Revised (Hodson, 2004); PCC – Percent Consonants Correct; PPC – Percent Phonemes Correct; PVC: percentage vowels correct; (Shriberg & Kwiatkowski, 1982); PPKP – Productive Phonological Knowledge Profile (Gierut, Elbert & Dinnsen, 1987); Psycholinguistic Framework (Stackhouse & Wells, 1997); RIU – Relative Influence on Unintelligibility (Dodd & Iacono, 1989).

¥Effect size calculated using a within subject design and online calculator from http://www.cognitiveflexibility.org/effectsize/effectsizecalculator.php

Π- IRD =Improvement Rate Difference – a method of calculating effect size for single-subject experimental designs (Parker, Vannest & Davis, 2011)