Comparing a controlled levelled vocabulary with a language rich vocabulary in a beginner reading scheme

Ruth Maria Price-Mohr

Doctor of Philosophy

University of York

Education

November, 2015

Abstract

The primary aim of this research was to compare the effect of using vocabulary that is within a child's current decoding ability in a reading text, with vocabulary that is beyond it. The original contribution to knowledge presented here is the discovery that children of all abilities and both genders can make greater gains in early reading when using reading books that go beyond their current phonic decoding ability (Intervention A), than from reading books that have a controlled, levelled vocabulary (Intervention P). The secondary aim of this research was to compare the use of a synthetic-phonics only approach with mixed teaching methods.

Three separate, but related studies were completed in schools across two counties. In total, there were 16 schools and 372 children who participated (4 schools acting as controls). A novel reading intervention, which had been purposely developed for the research (ωeebee Reading Programme), was used in 12 classes randomised to one of two possible Intervention strands (A or P). Measures of word reading and comprehension were used at both pre-test and post-test. All three studies were carried out over a 12-month period, although they began at staggered intervals.

The main findings were: first, children who used vocabulary that included many words which were beyond their current decoding ability, made greater gains in word decoding and comprehension than children using only vocabulary within their current decoding ability. Second, those children who had been taught using a mixed approach in addition to synthetic phonics made greater gains (particularly in comprehension) than those children taught using only synthetic phonics. In addition, gender analyses for all three studies, which included a small group of struggling readers, indicated greater gains for boys compared to girls, from using both the mixed approach to teaching as well as the non-decodable vocabulary.

Table of Contents Page No.

List of Tables

List of Figures

Acknowledgements

Deepest thanks to Louise Tracey for her guidance and wisdom in the supervision of this Thesis. Thanks also to Tracey Bywater and Claudine Bowyer-Crane for their support and advice.

Thanks to the Bowland Trust for their generous fee waiver.

Thanks to Mum and Dad for amazing spiritual and financial support.

Thanks to Ellie Simpkin and my daughter Louise for listening, and answering endless questions.

Special thanks to the schools and all the wonderful teachers, teaching assistants, head teachers and other members of staff, and of course the children, without whom none of this would have been possible.

Huge thanks to my wonderful husband Colin, for being there.

Author's Declaration

The results from Study 1 have previously been presented at the British Federation of Women Graduates (BFWG) student conference, in London, May 16, 2015.

I, Ruth Price-Mohr, declare that this Thesis and the work presented in it are my own and have been generated by me as the result of my own original research.

I confirm that: this work was done wholly while in candidature for a research degree at the University of York; I have acknowledged all main sources of help; where I have consulted the published work of others, this is always clearly attributed; where I have quoted from the work of others, the source is always given; with the exception of such quotations, this Thesis is entirely my own work.

Introduction

> "If Sara had been older or less punctilious about being quite polite to people, she could have explained herself in a very few words. But, as it was, she felt a flush rising on her cheeks. Miss Minchin was a very severe and imposing person, and she seemed so absolutely sure that Sara knew nothing whatever of French that she felt as if it would be almost rude to correct her." (From A Little Princess by Frances Hodgson Burnett, 1905)

When I was seven this was my favourite book and I would read it over and over. When I read the story I became Sara (the protagonist) and lived every moment with her. I knew the intentions of all the characters and empathised with all of Sara's feelings. Of course I could not have given a dictionary definition of all of the words at that age, or even now, such words as 'punctilious' have largely fallen out of use, but I knew what they *meant*.

Just suppose that you are Martha, aged five. Your home is full of books, pets and people talking. In your own room, you have a large selection of lovely picture books. Every night you are allowed to choose three of your favourites and you curl up in bed with your dad and read them together while mum gets your younger sister ready for bed. You already recognise nearly all the words in your books; you even recognise the word biscuit. You have just started school and are in nice Mrs Jones' class. In the morning you had to sit with the other children and learn that the letter 'a' makes the sound 'æ'. The next day you had to sit with the others and learn that the letter 'b' makes the sound 'b' and you tried not to fidget because you didn't want the nice Mrs Jones to get cross, but 'b' is so easy when you can already read 'biscuit'.

Now suppose that your name is George and you are five. There are no books or pets in your house, but there are ipads, tablets, laptops, wii devices and all sorts of great video games to play. Your dad thinks you are really clever because you can build a castle in 'Minecraft' faster than he can, and your mum thinks you are clever because you can upload her 'YouTube' clips to her website for her. You have just started school and are in nice Mrs Jones' class. In the morning you had to sit and learn that the letter 'a' makes the sound 'æ'. You don't really know why. You have seen letters before but they were never on their own. A few days later you learn that the letter with the stick and the round shape at the bottom makes the sound 'd'. You are convinced that the nice Mrs Jones has got it wrong because last time it was a 'b'. Sometime later Mrs Jones shows you a book. You have seen books

in the classroom; when Mrs Jones tells a story, she has one on her lap. There is a picture of an animal on one page. Mrs Jones points to a word and asks you to 'sound it out'. The first letter has a stick and a round shape at the bottom, you say 'b' and Mrs Jones says 'No, d'. You carry on and look at the next letter 'u'(^) then 'c' then you see the other letter which has the same sound but you can't be tricked. You pronounce 'c' (k) correctly. You look back and try to remember the sounds b…u…c…k… but you can't think of a word that it sounds like. 'Duck' says Mrs Jones and you wonder why you should.

Over the last few years I have met lots of children like Martha and George as well as lots of nice teachers like Mrs Jones. There are, of course, many children who fall between these extremes and this makes the work of the teacher both difficult and complex. As of September 2014, the statutory requirements in the National Curriculum for teachers are that they teach children to: 'apply phonic knowledge and skills as the route to decode words […] read accurately by blending sounds in unfamiliar words containing GPCs (grapheme phoneme correspondences) that have been taught […] read aloud accurately books that are consistent with their developing phonic knowledge and that do not require them to use other strategies to work out words' (Department for Education, 2013, p19). One could imagine a child such as Martha becoming frustrated by having to restrict her reading diet in the way described in the National Curriculum. Similarly, for a child like George who is more familiar with strings of information, to have to isolate and sound out single letters may become a cause for frustration. For Mrs Jones, who has a class of up to 30 children of differing reading experiences, to have to follow a single 'one-size-fits-all' approach with controlled-reading texts may limit her ability to support all the children in her care with equality.

I.1 Origins of the Thesis and the Research Questions

In 2005, the House of Commons Education and Skills Committee, Eighth Report of Session: Teaching Children to Read (House of Commons, 2005) recommended that the Department for Education and Skills commission a large-scale comparative study to establish: the relative effectiveness of different teaching approaches for reading; the most effective use of phonics; and "the effect of teaching texts which go beyond a child's existing knowledge of phonics compared to that of limiting instructional texts to those within a child's current decoding abilities" (p36). The Universities of

York and Sheffield were commissioned to conduct a review of experimental research into the use of phonics teaching (Torgerson, Brooks & Hall, 2006). However, no comparative study was commissioned into the use of instructional texts. The research studies presented in this Thesis attempt to establish the effects of teaching with texts that go beyond a child's phonic knowledge compared with those which are constrained by them.

The House of Commons Committee (2005) highlighted two fundamental questions for the teaching of early reading: what is the best form of instruction, and which are the best instructional texts to use? The first, method of instruction, relates to phonics (synthetic, analytic, onset-rime, analogy, blending and segmenting), whole-word sight recognition, whole language or any possible mix of these. The second, instructional texts, relates to the type of text to be used by beginner readers: reading schemes, 'real books' or a mix of texts. Much of the debate in the twentieth century focused on the phonics (skills-based) versus whole language (meaning-based) approach, but towards the end of the century, following a general acceptance of the importance of phonics, the debate has shifted to the type of phonics to be used and in what proportions.

There are two questions which this research was intended to address. Firstly, are there measurable differences when the vocabulary being used in reading books for beginner readers is either within their existing decoding ability, or is not so constrained? Secondly, are there measurable differences when comparing a synthetic phonics only approach with more eclectic methods; in other words, how does the method of instruction impact on children's learning to read?

I.2 Aims of the Thesis

The main aim of the research was to see if there is any evidence to suggest that the type of vocabulary used in an early reading scheme makes any significant difference to children's level of word recognition and passage reading comprehension. Specifically, does the vocabulary need to be restricted to words that are within the children's current decoding ability? A secondary question relates to the teaching methods associated with the reading scheme. Specifically, does teaching using an eclectic approach, while learning through play, differ significantly from a synthetic phonics only approach in respect of children's level of word recognition and passage reading comprehension?

There have been numerous studies comparing these different approaches, but very little exploration of the reading texts being used. The perceived value of a text to some extent depends on the intended learning outcomes. If the purpose of the process is for a child to learn the alphabetic principles used in written English, then a controlled text may well be more effective. If the purpose of the process is for a child to gain a higher score on a word recognition test then a whole word approach may be more effective for some children. If the purpose is to help the child to acquire the tools to gain meaning from written text, then it is likely that neither of these in isolation will satisfy most of the children most of the time. Further, it is likely that the learning styles of individual children will impact on the effectiveness of any one approach over another.

I.3 Research Strategy

Both types of text, real books and basal readers (instructional reading books) from schemes, do have a number of commonalities. These include: illustrations, a storyline, a limited number of words and size and style of font. The most significant difference is in the choice of vocabulary. By controlling for all possible confounding variables (illustration, storyline, number of words, font and teaching styles) it was hoped that vocabulary could be isolated as the independent variable in a comparison of the two kinds of text. The aim of this research was to discover if there are any differential outcomes for beginning readers where vocabulary is the independent variable in the taught text (phonically decodable compared to non-phonically decodable), and if so, what these differences are. This knowledge should inform teachers, publishers and policy makers.

Three different trials were devised. The first focused mainly on the teaching methods and the practicalities of using the materials in a typical working classroom; an effectiveness trial. The second focused on the main question of comparing the different types of vocabulary used in text, using an 'ideal conditions' scenario; an efficacy trial. The third trial investigated the comparative teaching methods and comparative written vocabulary as intervention strategies for struggling readers.

I.4 Research Technique

The first study was a three-armed randomised controlled trial with 12 schools at the outset and a total of 282 children. The control schools self-selected and the

remaining 8 schools were randomised to condition (phonics-based vocabulary or 'real books' style vocabulary). The second and third studies involved 4 schools with a total of 90 children at the outset. The children were randomly assigned to each of the phonics-based or 'real books' condition so that there were equal numbers overall, but roughly half in each school. Data for both trials was collected at pre and post-intervention. Assessment tools used were the British Picture Vocabulary Scale III (Dunn et al, 2009) and the York Assessment of Reading for Comprehension (Snowling et al, 2009). Additional data was collected using audio observations, lesson observations, questionnaires and record sheets.

I.5 Contribution to Knowledge

The research presented here makes an original contribution to the understanding of how children learn to read in the earliest stages. First, the research presented here challenges the existing literature concerned with instructional texts for beginner readers. There is no indication from the evidence in these trials that the use of carefully controlled vocabulary, which is kept within a child's current decoding ability, is a more effective approach than vocabulary which goes beyond it. Indeed, the evidence suggests that children benefit, particularly in reading comprehension, from reading more complex vocabulary situated in more complex sentences of the kind that occur in 'real books'.

Second, it challenges research concerned with teaching methods. The results provide evidence to support the view that children should be presented with multiple strategies in their approach to reading. In particular, the evidence from the trials presented in this Thesis indicates that using a single approach, such as synthetic phonics, results in lower gains in reading comprehension than using other approaches in addition to synthetic phonics.

Third, the results have indicated that some narrowing of the gender gap is associated to some extent with a mixed approach to teaching methods, but more so with the use of more complex written vocabulary. This was particularly noticeable for reading comprehension.

Fourth, these gains in comprehension, associated with a mixed teaching approach and more complex vocabulary, were also observed with struggling readers, whose teachers also reported gains in confidence and motivation.

I.6 Outline of Chapters

Chapter One discusses the phonics debate in the context of teaching methods for beginning readers. Chapter Two considers the arguments and evidence regarding the use of instructional texts for beginning readers, which are related to the issues discussed in Chapter One. In Chapter Three details of the design of the weebee Reading Programme, which is used as the Intervention for this research, are given. This includes: the rationale for the programme; a description of the reading books and teaching resources; details of supervision, fidelity to the programme, training and the manual; and the selection of specific written vocabulary for each condition.

Chapter Four is a description of the general methodology used in all three trials, and Chapters Five to Seven cover the specific methodology, results and analysis for each of the three trials respectively. In Chapter Eight key additional findings are considered that relate to the types of children who responded to different aspects of the Intervention, and some of the issues associated with educational research. This is followed by a general discussion in Chapter Nine, which considers the two main research questions, regarding teaching methods and instructional texts, in the light of the results from the three trials. There is also a discussion of the impact of the weebee Reading Programme on the gender gap and its potential use with struggling readers. Chapter Nine concludes with a discussion of the implications for policy makers and suggestions for future research.

Chapter One

The Phonics Debate

> "ONCE upon a time there were four little Rabbits, and their names were Flopsy, Mopsy, Cotton-tail and Peter. They lived with their Mother in a sand-bank, underneath the root of a very big fir-tree." (from The Tale of Peter Rabbit by Beatrix Potter, 1902)

The very first word in this well-known children's story book illustrates beautifully the dilemma for those wishing to help young children to learn to read. How does a teacher present, to a four-year old, a logical explanation for the pronunciation of the word 'once'? In addition, the first sentence contains two different spellings for two words that sound exactly the same (there and their). The above text is more likely to be read to children rather than given to them to read; however 'Once upon a time' is both a traditional and common start to many children's stories. Clearly, there are rules and conventions which have emerged over the centuries which determine how spoken words and meanings are conveyed in written form. What is less clear is the process children use to interpret the written form.

This chapter will firstly chart, in brief, the historical context for the phonics debate and present both sides of the argument that has become polarised over the decades. It will then explore the most prominent models of how children learn to read, and the literature that discusses the wider debate: the perceived benefits from a skills-based approach (phonics), compared to a meaning-based approach (whole language or whole word). This is followed by a discussion of the debate in the context of children with learning difficulties, and finishes with arguments supporting a mix of both phonics and whole-word approaches.

1.1 Historical Context

At one time, school teachers were left very much to their own devices. The first teachers of reading in English were seventh century priests (Hempenstall, 1997). Teachers taught children in their care the way they themselves had been taught. They would learn the letters of the alphabet by learning the initial letter sounds of words for example 'A is for apple, B is for ball and C is for cat', followed by syllables. They would have read from a primer or prayer book. By the sixteenth century, they would have access to the Bible as a reading text. Later teachers would follow a sequence of letter names (upper and lower case), two letter combinations, three

letter combinations, and then gradually increase the number of syllables. They would then introduce texts such as fairy stories or nursery rhymes. As long ago as 1828, a primer was produced that used the idea of whole-word recognition rather than sounding out words, with the analysis of letters to follow later (Hempenstall, 1997). So began a difference in approach to the teaching of early reading that has led to the polarisation of opinion on the best way to teach reading: firstly amongst teachers and educators; then amongst psychologists and other experts; and then, as now, amongst politicians (Ellis & Moss, 2014).

Historically, teachers of the youngest children followed the introduction of the alphabet and demonstrating initial letter sounds with other phonemes, which were usually treated as part of the writing curriculum and learned as part of spelling tests. Early reading was, for the most part, modelled by the class teacher, as it still is today, in guided reading. Whole word reading became popular during the 1920s, and before the 1960s, whole word sight reading was the dominant approach. But concerns over the lack of skills-based teaching led to exploratory research into other approaches (Elliott & Grigorenko, 2014). This included the use of the Initial Teaching Alphabet (ITA), an experimental approach using a phonetic alphabet (Downing & Latham, 1967). Using the ITA, children were encouraged not only to read using phonic sounds, but also to write using only phonetically regular sounds (those that are the initial letter position sounds) rather than using traditional spelling patterns and rules. By contrast, other teachers were using a 'look and say' method in which children were shown flash cards and taught no phonics at all. This approach was inspired by the work of Kenneth Goodman who emphasised the predictive nature of reading based on the reader's understanding of spoken language (Goodman, 1967).

In the early 1970s in England, at the request of the government of the day, a report was commissioned to review best practice in the teaching of reading, known as The Bullock Report (DES, 1975). Their conclusion then was much the same as a similar review commissioned thirty years later (Torgerson et al, 2006).

> "There is no one method, medium, approach, device or philosophy that holds the key to the process of learning to read […] some would put so much emphasis on the 'mechanics' of reading that certain children would be handicapped rather than helped. Others advocate so keenly the virtues of mature reading from the beginning

> that they are in danger of leaving it too much to trust that the skills will be acquired on the way." (p77)

The conclusion in the Bullock Report was that it was not possible to state that there was one approach to reading and that reading was too complex a process to try to reduce to a single method. Nevertheless, the debate continued. As recently as 2011, a House of Commons Parliamentary Group for Education published a report into Barriers to Literacy, in which they stated that there are different ways for children to learn to read and that synthetic phonics alone is insufficient (House of Commons, 2011).

During the latter part of the twentieth century, there was an increasing demand for a scientific basis for reading reform to improve standards of literacy. There was a demand for experimental-style research that was expected to be both reliable and replicable, following a medical model with large samples and randomisation of assignment to condition (Pearson, 2004). By the end of the century, there was a significant body of evidence supporting the use of systematic phonics and an emphasis on developing phonemic awareness; disagreement continued, however, over how phonics should be taught.

In 2000, the American National Reading Panel report (NICHHD, 2000), using meta-analysis of research considered to be reliable, recommended a balanced approach to teaching. Phonics was found to be useful in the early stages of learning to read, but neither synthetic phonics (blending sounds of letters from left to right through a word) in particular, or analytic phonics (looking at letter patterns or word families), was emphasised (Pearson, 2004). Phonics instruction was recommended as part of a balanced programme and not considered to constitute a total reading programme in itself (NICHHD, 2000). Support for the whole-language approach had been largely based on qualitative research such as case studies. However, in a climate where quantitative research was considered to be more scientific, the evidence in support of whole-language teaching was considered to be weak (Elliott & Grigorenko, 2014). Pressure to produce measurable results led to an increasing emphasis on a skills-based approach.

In 1997 in England, the Literacy Task Force emphasised the systematic teaching of phonics, stressing the importance of graphic knowledge and sound-symbol relationships (Literacy Task Force, 1997). This led directly to the National Literacy Strategy (DfEE, 1998) with its 'searchlight' model, which emphasised:

context, word recognition, graphic knowledge, and grammatical knowledge, as well as phonics knowledge. The aim was to have a balanced approach that included the direct teaching of systematic phonics (Solar & Openshaw, 2007). The Searchlights Model remained in place until 2006 when it was replaced by the Primary National Strategy (DfES, 2006). This was a direct result of recommendations in the Rose Report (2006) that there should be an emphasis on the use of synthetic phonics in the teaching of reading.

This focus on teaching methods was partly triggered by the debate surrounding the use of synthetic phonics as a teaching tool following the publication of a longitudinal study in Clackmannanshire, Scotland (Johnston and Watson, 2005). The Clackmannanshire study claimed to have demonstrated improved word level reading as a result of teaching through a systematic synthetic phonics approach compared to the traditional approach, described as analytic. According to this report, the programme had made a strong impact on pupils' ability to sound out, spell and recognise words.

The remit for the Rose Review (2006) was to discover what was considered to be best practice in the teaching of early reading and systematic phonics. The report drew on research reviews, written and oral accounts, papers submitted, HMI surveys, and OFSTED reports and data. Particular mention was made of the Clackmannanshire Trial. The Rose Review (2006) included the recommendation that high quality systematic phonics should be taught discretely, and as the prime approach in learning to decode, but also within a broad and rich language curriculum. Two controversial issues that followed the Review were: a recommendation that the existing 'searchlights' model should be replaced by the 'Simple View of Reading', a framework that places word recognition and comprehension on two distinct dimensions (Gough & Tunmer, 1986); and that the phonics teaching should follow the principles of synthetic phonics. These principles were embedded in subsequent National Curriculum documents.

As part of the Rose Review, a systematic review of research was commissioned. The authors of this review concluded from the available evidence that a systematic approach to phonics was to be recommended, but added that there was insufficient evidence to advocate the use of one type of phonics over another, or that phonics should be taught exclusive of other methods (Torgerson et

al, 2006). Their key relevant findings were that systematic phonics instruction within a broad literacy curriculum had a positive effect on word accuracy, but that there was no evidence of any positive effect on reading comprehension. Nevertheless, following Rose's final report (Rose, 2006), the use of synthetic phonics became the recommended method of instruction and has since become a statutory requirement in the national curriculum (Department for Education, 2013).

1.2 Models of Reading

As far back as 1908, Huey (1908) said: "…it is perfectly certain that words are not perceived by a successive recognition of letter after letter, or even by any simultaneous recognition of all the letters as such [...] it is certainly a recognition of whole words…" (p111), and "until the insidious thought of reading as word pronouncing is well worked out of our heads, it is well to place the emphasis strongly where it belongs, on reading as thought-getting…" (p350). There is, as yet, no real consensus as to a definition of reading, although there have been suggestions. For example, Morris (1963), like Huey considered reading to be thinking in response to print. There is, nevertheless, a general consensus that reading is a far more complex process than decoding an alphabet and involves the use of previous knowledge, sampling, hypothesis forming, prediction and comprehension (Levin and Williams, 1970; Gibson and Levin, 1975; Smith, Goodman & Meredith, 1976; Southgate, Arnold & Johnson, 1981).

Theories of learning that emerged during the last century reflected a general assumption that there is a developmental progression in learning. These ideas were based on the work of cognitive theorists, such as Piaget, Vygotsky and Bruner (Atherton, 2013). Two very influential developmental models of learning to read suggest a progression of stages (Frith, 2001) or phases (Ehri, 2005). In both models, the first stage or phase suggested is one in which the child can recognise familiar words, for example, their own name or an advertising logo. The next stage suggested is an alphabetic stage, when the child begins to learn letters and their corresponding sounds. This is followed by a third stage, when a child demonstrates the ability to read words in connected text. Frith (1985) refers to the first stage as the logographic stage when children can recognise a few significant words. Frith does not consider this to be reading, but simply recognising symbols that represent, for example, a product that is being advertised. The second of Frith's stages is the

alphabetic stage, when children begin to learn letter-sound relationships and use a few letters to recognise words, particularly in the initial and final positions of words. The final stage is orthographic, when children are able to recognise whole words from reading all of the letters.

The model described by Ehri (2005) includes four phases: pre-alphabetic; partial alphabetic; full alphabetic and consolidated alphabetic. In the pre-alphabetic phase, children are thought to recognise words only as a shape, for example, looking at the outline of the word without forming any letter-sound connections. Like Frith, Ehri does not consider this form of word recognition to be reading as it does not appear to involve the alphabetic system; it does not involve making letter-to-sound connections, but invokes connections between visual cues and meaning. It could be argued that there is little clear difference between this pre-alphabetic phase and Frith's logographic stage (Beech, 2005). In the partial-alphabetic phase, children are thought to learn letter names and sounds, but only form connections with some, for instance the initial and final letters. When writing, children in this phase invent partial spellings of words by writing only the more salient sounds and leaving out the medial letters. Full alphabetic phase is considered to have been reached when children learn to recognise words by sight from forming complete connections between letters in spelling and phonemes in pronunciations. Once children have learned the major grapheme-phoneme correspondences, they can decode unfamiliar words. The consolidated phase is considered to have been reached when children perceive recurring letter patterns as larger units. The theory behind the model is that the alphabetic system works as a mnemonic for learning a sight vocabulary. Ehri emphasises that the application of the alphabetic system is not the conscious act of decoding, but an automatic activation of alphabetic knowledge that is used to build a sight vocabulary (Ehri, 2005). Both of the above models to some extent reflect the observed progress children make in reading as a consequence of methods of instruction.

1.3 A Skills-based Approach

For some children, a very structured approach to learning reading would seem to be logical. Children who start school with limited spoken vocabulary, little or no access to books at home and who have not had the experience of being read to from a range of pre-school books are likely to need more support in the early stages of

learning to read than children who have had a wealth of language and book experiences.

The phonics approach is based on the notion that children read words in a linear progression (from left to right in English) and process each symbol to its corresponding sound in sequence. Children are taught to recognise symbols in a specified sequence with gradually increasing complexity. They are taught to blend or 'sound out' the words. There are, of course, some words for which this cannot be done, known as 'common exception words' (the National Curriculum in England, Department for Education, 2014) and these words have to be taught as whole words (these are listed in the 'Letters and Sounds' guidelines, DfES, 2007).

Researchers and educationalists agree that phonics instruction in blending is needed for writing and spelling, but to what extent it is necessary for reading remains in question (NICHHD, 2000; Rose, 2006). The ability to deconstruct words into separate phonemes in order to spell words is important for writing, but to what extent children should decompose words for reading, and at what stage of learning to read, continues to be a subject for debate. Supporters of a skills-based approach advocate teaching phonics as a first step, working from the smallest phoneme to decode the alphabet, and learning the letter-to-sound correspondences. A skills approach does not necessarily imply one particular form of phonics instruction, but may include all the variations associated with phonics teaching (synthetic, analytic, onset-rime, analogy, blending and segmenting).

In the US, the review commissioned by the NICHHD (2000) examined the available research evidence to try to determine which method of teaching children to read was deemed to be the most effective (Ehri, Nunes, Stahl & Willows, 2001a). The authors concluded that systematic phonics instruction had a greater effect on reading ability than no phonics instruction. In addition, they found the effect size for synthetic phonics (d = 0.45) to be greater than the large unit/analytic phonics (d = 0.34). However, there were a number of potentially controversial issues. Of the 38 studies included in the review, only 14 (36%) included normally developing children (not at-risk, low achievers or reading disabled). Only thirteen of the studies were randomised controlled trials; only two trials were both randomised and had a normal population. Only twenty of the studies reported results for comprehension, of which only eight had a normal population. For each of the treatment-control

comparisons, effect sizes across six different outcome measures (decoding regular words; decoding pseudo-words; reading miscellaneous words, spelling words; reading text orally; comprehension of text) were averaged to generate one overall effect size, although not all the trials had results for all six of the outcomes.

One of the studies included had, according to the authors, an atypical effect size (d = 3.71) and "…to limit its influence on the mean effect size for the large-unit phonics category, its effect size was reduced to equal the next largest effect size in the set" (p419). The study in question (Tunmer & Hoover, 1993) compared the use of systematic analytic phonics with no systematic phonics teaching for at-risk six-year-olds. For this large-unit study, the children were taught awareness of visual patterns and shared sounds. The researchers found a significant effect size in favour of the analytic phonics. The sample size was relatively small at sixty four, but fourteen of the other studies in the Review had smaller samples.

The National Reading Panel Report (NICHHD, 2000) concluded from the evidence reviewed in the meta-analysis that synthetic phonics, analytic phonics and other phonics programmes are the most effective type of instruction; however, the report resulted in some controversy. Camilli, Vargas and Yurecko (2003) carried out a reanalysis of the impact of different teaching methods. Their results suggested that tutoring methods and language-rich approaches had as great an effect size as did systematic phonics. The studies used in the Camilli et al (2003) analysis were found to have a smaller effect size for systematic phonics instruction than in the original report. There followed several years of conflicting reanalysis of the evidence that questioned the original effect sizes and some of the inclusion criteria (Camilli, Kim & Vargas, 2008; Stuebing, Barth, Cirino, Francis & Fletcher, 2008). Much of the research, included had been used as evidence for policy changes in the US prior to the 2000 Report, was focused on children with reading difficulties; in addition, the reading gains reported were often in phonemic awareness and pseudo-word pronunciation, but not text comprehension or fluency (Allington & Woodside-Jiron, 1999).

A second large-scale review, conducted at the same time by the National Reading Panel (NICHHD, 2000), looked at the research evidence regarding the teaching of phonemic awareness (Ehri et al, 2001b). Much of what was described as 'phoneme awareness' included blending and segmenting phonemes, blending onset-

rimes and learning to recognise patterns of letter sounds in initial and final positions. There is considerable overlap here with both synthetic phonics and analytic phonics instruction, making it difficult to distinguish differences between the two reviews. The analysis suggested that the effects of greater phoneme awareness were larger for phoneme segmentation and deletion assessments than for blending (synthesising), but overall the conclusion was that phoneme awareness instruction benefited decoding skills and that the effect was greater for at-risk readers and those in the early years. It was also found that there appeared to be an optimum length of study time (5 to 18 hours), after which there were negligible additional benefits. The authors point out that other studies suggest that children can acquire phonemic awareness through learning to read and spell without being explicitly taught, although this may be insufficient for some. There is a large body of research which holds that phonemic awareness is necessary for success in reading, but there is continued debate about how to encourage its development in learners: either through explicit teaching, or allowing it to emerge through both reading and writing (Ehri et al, 2001b; Hatcher, Hulme & Ellis, 1994; Muter, Hulme, Snowling & Stevenson, 2004; Kuppen, Huss, Fosker, Fegan & Goswami, 2011; Vellutino, 1991; Wilson & Colmar, 2008).

In England following the Rose Review (Rose, 2006), the new framework known as the 'Simple View of Reading' was introduced, which places word reading (decoding) and linguistic-comprehension on two distinct dimensions (Gough & Tunmer, 1986). The Simple View of Reading was interpreted by the Review panel (Rose, 2006) to mean that teachers should focus on teaching children how to decode, and comprehension would follow automatically from their understanding of spoken language. The framework has been used to justify giving priority to phonic work as the prime approach to the teaching of reading. However, different skills and knowledge contribute to performance in each of the two dimensions and no specific direction was given as to teaching on the comprehension dimension (Muter et al, 2004). There may be multiple underlying factors on the different measures and there is dissociation across the two dimensions (good word recognition with poor comprehension and poor word recognition with good comprehension) (Nation & Snowling, 1997). In addition, it has been observed that the use of context differs between skilled and less-skilled readers. Less-skilled readers appear to rely on context more for word recognition, whereas skilled readers use context for

comprehension, suggesting context as another dimension that needs to be considered within the framework (Nation & Snowling, 1998).

The Simple View of Reading (Gough & Tunmer, 1986) was originally conceived to help with the assessment of children with reading difficulties. It was intended as a framework to help teachers locate where a child's weaknesses lay, in the two dimensions of either decoding or linguistic comprehension; different factors predict word recognition from those that predict comprehension (Stuart, Stainthorp & Snowling, 2008). However, it is a framework intended to represent individual differences in reading comprehension and is not a developmental model (Protopapas, Simos, Sideridis & Mouzaki, 2012). As such, it is potentially limited in its use as the basis for a teaching curriculum and does not take account of the effects of fluency, speed and expression (Silverman, Speece, Harring & Ritchey, 2013). More recent research has suggested that variation at the classroom level (the teacher, peer group or social setting) has a greater impact on reading comprehension than either of the two dimensions in the Simple View framework (Savage, Burgos, Wood & Piquette, 2015).

During the 1990s, the phonics debate became more focused on the type of phonics instruction used, as well as how much and for how long it would be most effective for beginning readers. The most widely cited research in favour of synthetic phonics, the Clackmannanshire study (Johnston & Watson, 2005), has received extensive publicity (Compton, Miller, Ellenan & Steacy, 2014; Cook, Littlefair & Brooks, 2007; Davis, 2012; Ellis, 2007; Ellis & Moss, 2014; Rose, 2006; Wyse & Styles, 2007). Within this study there were three separate strands. The first of these (1992-3) looked at the pace of teaching phonics; the authors reported gains when teachers began to teach children how to decode three-letter phonically-regular words. They also reported gains for children taught accelerated analytic phonics (the definition of analytic here only involved learning letters in initial positions). The second study of 10 weeks, in 1998, compared synthetic phonics taught at an accelerated pace (two letters a week but in three different positions in a word) with analytic phonics (two letters a week in the initial position only). They concluded from this study that synthetic phonics led to better reading (Ellis, 2007). The third study, begun in 1998, was a 16-week programme, comparing accelerated synthetic phonics with analytic phonics. In the synthetic phonics group children were taught six letter-sounds in eight days in initial, middle and final positions; they

were taught to sound and blend words and how to spell. In the analytic phonics group, children were taught just one letter per week in the initial position. After 16 weeks, the synthetic phonics group were 8 months ahead of the others in spelling and word reading. At the end of the 16-week programme, the children who had been in the analytic phonics group were given synthetic phonics instruction; thus by the end of Primary 1, all the children had been taught synthetic phonics (Johnston & Watson, 2005).

The children were tracked through to the end of primary school (a further 5 years) and substantial gains on phonic decoding skills were reported (on average three and a half years ahead of their chronological age using standardised tests), but in this later phase there was no comparison group and there were no gains reported for comprehension. In addition, there were a number of possible confounding factors: the introduction of new reading schemes and library books, a new literacy programme and new staff development initiatives (Ellis, 2007). Only three of the eight schools were above national average for reading attainment (the largest school, with the lowest number of free school meals was below average for 2002/3) and in the following year (a non-intervention cohort) four of the eight schools were above national average (only one cohort participated in the programme). It has been reported that subsequently the Local Authority as a whole achieved below average scores on Scotland's national reading tests (HMIE, 2006).

As part of another large-scale Scottish study of a literacy intervention in West Dunbartonshire (Mackay, 2007), a supporting study set out to compare the use of synthetic phonics with the existing phonics approach, described as analytic. The authors reported significant gains for the children in the nine participating schools using synthetic phonics over the nine comparison schools using analytic phonics. However, as acknowledged by the author, there were a number of limitations to the study. The assessments used were not standardised, so only raw scores could be used which did not take children's age into account; in Reception classes, children can vary in age from 4 years and 1 month to 5 years. The synthetic phonics condition used "Jolly Phonics" as their reading text, but there was no information regarding the texts or instructional methods being used in the control condition. The schools participating in the intervention were not randomised. Gains were reported for non-word reading and word reading, but comprehension was not assessed. In addition, the children in this supporting study also participated in the whole

intervention, in which the introduction of synthetic phonics was one of ten different measures to improve literacy in the county; hence there were a number of potentially confounding variables, the effects of which cannot be separately assessed. It is thus difficult to draw conclusions as to the actual effect of the different teaching methods.

Synthetic phonics does seem to be a logical approach and for many languages with a transparent or shallow orthography, this approach works extremely well. In these languages letter-to-sound correspondences are consistent. English, however, is considered to have an opaque or deep orthography which has many more patterns and exceptions to learn. About fifty per cent of English words are exceptions to the rules of phonics (Devonshire, Morris & Fluck, 2013). Following a study teaching reading with synthetic phonics, Devonshire et al (2013) concluded that instruction using only synthetic phonics may make it difficult for children to hypothesise about written language that goes beyond sound-to-letter mappings. The authors also suggest that teaching only with synthetic phonics may convey to the children the idea that spellings of words only represent sounds of speech, which, for English, is not true; the morphology within written words also conveys meaning. Where English spelling does represent meaning at the morphological level, spellings are consistent; moreover much of English spelling retains etymological information. Devonshire et al (2013) found evidence that children can use these types of information for learning strategies and recommend that rules of form should be directly taught. Compton et al (2014) go further and suggest that the use of decoding instruction without context may encourage the use of smaller units (phonemes) and promote letter-by-letter reading. They suggest that this approach fails to foster generative reading development and the context-dependent relationships of letters and words.

An alternative skills-based approach to synthetic phonics (sounding out and blending letters) is referred to as analytic phonics. Definitions of phonics in the literature have, at times, been vague, and often contradictory (Torgerson et al, 2006). Analytic phonics in particular has been lacking in a clear definition; referred to variously as onset-rime, large unit, analogy or rhyme analysis. Arguments about the use of phonics (whether synthetic or otherwise) compared to whole-word reading have also been lacking in clarity. Research evidence that has claimed to compare the effectiveness of analytic phonics with synthetic phonics, or phonics

instruction with a whole-word approach has been collected. However, according to Wyse and Styles (2007), very little reported research has made a clear distinction between the various instructional approaches. In addition, out of the 43 studies looked at in the review by Torgerson et al (2006), only 9 were carried out with children aged 5-6 and none for 4-year-olds, the age at which most children begin to learn to read in England.

Analytic phonics has been defined as identification of common phonemes in a set of words (Torgerson et al, 2006). These could be in any position in the word and comprise either a single letter or group of letters that constitute one perceived sound. If the phoneme being analysed is the initial letter, then this approach is not very different from the traditional type of onset-rime teaching. Evidence from eye tracking movements supports the use of this approach by pointing to the fact that eyes fixate on the initial letters of words in the first instance, followed by the final letter and then move to salient features or letter clusters within the word (Rayner, Slattery, Drieghe & Liversedge, 2011). Traditional methods for teaching spelling patterns use what is effectively an analytic approach; for example: boat, coat, goat, moat etc. There are alternative definitions of analytic phonics, such as learning sound-to-symbol relationships within the context of whole-word recognition in which children analyse the common initial phoneme in a set of words, although this also can easily be confused with the concept of onset-rime (Cook et al, 2007). The analytic phonics approach to teaching is based on evidence that children have the ability to use analogy (Cunningham, Perry, Stanovich & Share, 2002; Goswami, 1999; Moustafa, 1995; Moustafa & Maldonado-Colon, 1999; Treiman, Mullennix, Bijeljac-Babic & Richmond-Welty, 1995; Wang, Nickels, Nation & Castles, 2013). Thus, having learned, for example, the words 'boat' and 'coat', they can predict a word such as 'goat'.

Following the introduction of a phonics screening check for six-year-olds in England (technically some children could still be just five), the National Foundation for Educational Research carried out a survey to assess teachers' views on both the phonics screening check and the value of teaching phonics in the early years (Walker, Bartlett, Betts, Sainsbury & Mehta, 2013). Although teachers were, in general, supportive of the use of phonics (not necessarily synthetic), the majority of schools supported an eclectic approach, agreeing that phonics should be taught in the context of meaningful reading, stressing the importance of comprehension. The

introduction of the screening check resulted in one third of the sample initiating changes in teaching by: increasing the time devoted to teaching phonics; introducing nonsense words into their teaching; and using phonetic spelling tests rather than high frequency words. In 2014, the phonics screening check was repeated (Walker, Bartlett, Betts, Sainsbury & Worth, 2014). Findings were very similar to the first report, although there were some concerns expressed. These concerns were: pressure to teach to the test; the attempts made by some children to make pseudo words into real words; an observed adverse effect on spelling; and the perceived 'holding back' of children who were already skilled readers (p47). Again, teachers were generally supportive of using phonics, but emphasised that it should be used alongside other approaches. The final report, after three years of tests, included similar findings. Further, they found no evidence of improvements in literacy standards that could be attributed to the phonics emphasis associated with the test (Walker, Sainsbury, Worth, Bamforth, & Betts, 2015).

In English, there are conditional rules about how consonants may be clustered and the number of vowels in a sequence. These rules facilitate word recognition in skilled readers. There are concerns that teaching children to read using made up words, that do not necessarily conform to these rules, will reduce their reading speed and comprehension (Hempenstall, 1997; Devonshire et al, 2013). A number of studies have indicated that the method of instruction used, has considerable influence on the strategies that children develop in their learning to read. A study to test the assumption that using either synthetic or analytic phonics instruction can lead to qualitative differences in literacy cognition found that children taught using analytic phonics were better at shared rimes in words, and those taught by synthetic phonics were better at blending, but not segmenting (Comaskey, Savage & Abrami, 2009). The authors found that students who are explicitly taught about rimes were able to identify and articulate them. Children were found to use the reading strategies they had been taught, but appeared also to use other strategies in addition (children taught analytic phonics also used a synthetic approach at times and vice-versa); when children were taught to synthesise as their prime approach, they were still able to use analogy. The authors found no clear differences between the two approaches in children's decoding ability.

A similar study of children's reading strategies suggests that the method of reading instruction determines the skills predicting children's initial reading acquisition and development. The study compared an eclectic approach with synthetic phonics. Results showed that for children taught using a mix of synthetic and analytic phonics, their letter sound knowledge, rhyme awareness and oral vocabulary knowledge predicted word reading. For children taught using synthetic phonics, letter sound knowledge, phoneme awareness and short term memory span predicted word reading (McGeown, Johnston & Medford, 2012). In a subsequent study, phoneme awareness was found to be a stronger predictor than rhyme awareness where the method of instruction was synthetic phonics (McGeown & Medford, 2014). In this same study there were inconclusive results comparing synthesis and analysis, despite the teaching focusing on synthesis.

In a series of Canadian studies, beginner readers were taught using either a rhyme analogy approach or a grapheme-phoneme re-coding approach (Ziegler & Goswami, 2005). The results showed that the analogy taught group could also read new words using a phonics approach, whereas the phonics group could not read new words using a rhyme-analogy approach; at follow up, the rhyme-analogy group scored higher than the phonics. A much older study compared reflective and impulsive learners' responses to the different approaches (Readence & Baldwin, 1978). Reflective learners in the synthetic approach performed significantly better than impulsive learners in sight vocabulary, and reflective learners in the analytic approach were significantly better than impulsive learners on comprehension. Taken together, these studies suggest that the skills that young children use when learning to read, although they largely reflect the way they have been taught, also demonstrate an ability to analyse words and choose their own strategies.

1.4 A Meaning-based Approach

The meaning-based, whole-language approach, often referred to as the 'Look and Say' whole-word method of instruction, uses flash cards to build up a sight vocabulary. This approach is associated with a focus on comprehension and engagement with text, but also with a lack of instruction in skills, strategy, text structure and reading for content (information texts). One of the criticisms levelled at a whole-language method of instruction, whereby children rely largely on visual

recognition of word shapes, is that some students fail to develop any strategy for decoding novel words (Hempenstall, 1997).

In a recent small study, where teaching of intensive high frequency words was compared to systematic synthetic phonics only, researchers found that over a five-week period the children taught the high frequency words learned the words rapidly (Watts & Gardner, 2013). These were pupils who were deemed by their teacher to be least able to make the most progress. The learning of high frequency words was also found to improve fluency and accuracy. In addition, evidence from miscue analysis suggested that pupils were not using knowledge of grapheme-phoneme correspondence in their independent reading (Watts & Gardner, 2013). The authors conclude that use of synthetic phonics alone is insufficient for the development of fluent reading. The results of an older comparison of code-based instruction with meaning-based instruction support the view that phonics training may be necessary, but not sufficient for fluent reading (Foorman, Francis, Novy & Liberman, 1991).

According to Vellutino (1991), much of the debate centres on: whether or not automaticity in word recognition is best learned out of context; whether automaticity is necessary for comprehension; and the value of analysing a word's alphabetic structure when learning to read. Vellutino makes a number of generalisations based on his own research findings: firstly that word identification is essential; secondly that fluency is necessary for comprehension; thirdly that word identification can occur out of context; and finally, understanding of the alphabetic principle is necessary, for which phoneme awareness is a prerequisite. The implication being that both an understanding of the alphabetic principle and whole-word identification are equally essential for learning to read.

Further evidence of the importance of a larger unit or visual approach was found in a study by Wang, Nickels, Nation & Castles (2013). The authors found that orthographic knowledge (how letters and letter groups are combined in text - morphology) contributes to orthographic learning (learning new words) beyond phonological decoding skills for both regular and irregular words. In addition, they found that children with better orthographic knowledge were also better decoders. Their results suggest that focusing on orthographic detail is more important than generating the phonology of a word, although knowledge of pronunciation and

meaning was a high predictor for learning irregular words. The importance of orthographic knowledge has been highlighted in some older studies; for example, a study comparing phonics instruction with flash-card drill. Students who were taught a phonic decoding strategy were not found to learn to read words more accurately in connected text than students learning through flash cards (Joseph & Schisler, 2007). In addition, the flash-card drill was the most efficient with regard to oral reading passage fluency. The flash card students also learned to read more words accurately per minute of instruction time than the comparator phonic group. There was no measure of comprehension included in this study. Nevertheless, these studies emphasise the importance of orthographic knowledge.

1.5 Readers with Learning Difficulties

Some researchers have expressed the view that there may be groups of children who are disadvantaged by having to learn to read via any one particular instructional method; for example, children who have an auditory disability (Wedell, 2014). Teachers of children with special needs have observed that pupils have not been making expected progress in response to the current systematic synthetic phonics approach; that there is a proportion of pupils for whom such an approach does not seem to be effective. In a forum for special educational needs teachers (SENCo), the view was expressed that such pupils might need a different approach; there was general agreement that spelling-to-sound correspondences might be acquired at different stages and in different ways, and that whole-word learning could be an initial approach that might be developmentally more appropriate for some (Wedell, 2014). Models of reading (Ehri, 2005; Frith, 1985) suggest a whole-word starting point.

Studies observing children with dyslexia and specific language impairment have shown that both phonological skills and auditory processing are often impaired, and reading interventions that have been effective embed phonics in a wider literacy programme (Duff & Clarke, 2011; Fraser, Goswami & Conti-Ramsden, 2010). As part of a battery of assessments for dyslexia, the Aston Index (Newton & Thompson, 1982) includes a test for auditory sequential memory. This highlights the observation that some children have difficulties with blending letter strings in words (sounding out). Children with Specific Language Impairment have been found to have difficulty in detecting speech segmentation and tone duration due to impaired

phonological processing (Corriveau, Pasquini & Goswami, 2007; Corriveau, Goswami & Thomson, 2010). There are a number of other studies that have associated poor reading with poor auditory processing and difficulty in blending and synthesis (Duff, Hayiou-Thomas & Hulme, 2012; Kuppen et al, 2011; Wallach, 2011).

During speech, even what is generally thought to be the smallest unit, the phoneme, has variations within its articulation. This can depend, for instance, on the intended sound of the phoneme that will immediately follow (for example, if the letter *c* is to be followed by *a* or *o*) or even simple variations in the vocal chords of the individual who is making the utterance. These variations are known as allophones and are not generally perceived. There are, however, some children who do perceive these allophonic variations and for them the result can be slower and less consistent mappings of sounds to graphemes as they have to assign a greater number of sounds to each letter (Goswami, Fosker, Huss, Mead & Szucs, 2011; Hempenstall, 1997).

Another group of children who may benefit from using more than one instructional method are those who may have impaired working memory. For these children, the use of teaching by analogy or onset-rime (analytic phonics) requires less demand on working memory (Baylis & Snowling, 2011; McGeown & Medford, 2014; Wedell, 2014). In a synthetic phonics approach to teaching reading, children are required to retain sequences of letter-sound correspondences in memory to blend together, thereby relying on short-term memory span. The load on short-term memory can be significantly reduced by chunking letters into larger units, by recognising whole words by sight and using whole sentences and situational context (Ferre, 1987).

1.6 A Mixed Approach

A complex mix of results was found in a study that set out to explore the effect of different types of instruction on children of differing levels of oral vocabulary and decoding ability in a normal population of beginner readers (Connor, Morrison & Katch, 2004). Children who began with weaker decoding skills showed greater gains in word reading skills with explicit decoding instruction, whereas children with stronger decoding skills improved less. Children who began with weaker vocabulary skills made greater gains in word reading skills when independent reading and writing skills were minimised, whereas children who began with a stronger

vocabulary achieved greater gains from independent reading and writing. The authors speculated that having stronger oral vocabulary skills may support decoding when encountering unknown words, and that more meaning-based instruction for those children may be associated with stronger gains in word reading. In a separate study, Taiwanese teachers were asked if they found a skills-based or whole language approach more effective for teaching young learners to read English; the results suggested a clear preference for an eclectic approach (Huang, 2014). Wilson and Colmar (2008) recommend a balanced approach which explicitly teaches phonic skills (analytic and synthetic) but maintains the context and point out that it is the *systematic* nature of the teaching of specific skills, whether synthetic, analytic, whole-to-part or part-to-whole which research shows to be key to an effective approach.

The National Reading Panel (NICHHD, 2000), the Torgerson Review (Torgerson et al, 2006) and the Rose Review (Rose, 2006) all emphasised that systematic phonics should be part of a wider language curriculum. Phonics (whether synthetic or analytic) does not produce fluent reading by itself, and young readers need to be able to use semantic and syntactic cues as an aid to word identification and as a corrective (Dombey, 1999). Instead of being the way to learn to read, phonics, which is important for spelling and writing, can be largely learnt through reading (Dombey, 1999). For those who advocate a balanced programme of instruction, with a mix of skills-based and meaning-based approaches, there is the question regarding the optimum combination of these. Successful teaching of reading that leads to understanding in literacy has been associated with an eclectic approach, balancing the direct teaching of skills with contextually grounded activities (Flynn, 2007). A balanced method of instruction would include the teaching of explicit skills and strategies, phoneme awareness, letter-sound knowledge and concepts of print, while retaining the contextualised reading of the whole-language approach.

1.7 Gender

The much-cited Clackmannanshire study (Johnston & Watson, 2005) claimed to have found evidence that boys had made gains compared to girls through the teaching of synthetic phonics, although this was only in word reading, rather than comprehension. Logan & Johnston (2010) have suggested that boys are not

naturally inclined to link phonological and visual information and thus the explicit teaching of this via synthetic phonics is of benefit. There is much reported evidence of higher proportions of boys falling into the poor-reader category (Rutter et al, 2004; Snowling & Hulme, 2012). There is also some evidence of neurological differences in gender (Logan & Johnston, 2010). In addition, more recent research has reported that significantly more girls than boys prefer to use a phonics approach as a strategy to reading unfamiliar words (Beech, 2010). The impact on the gender gap is clearly a significant element in the evaluation of any approach to the teaching of reading to beginners.

1.8 Summary

In the design of any strategy for teaching beginning reading, there must be consideration given to the many and varied aspects of the learning process and the way in which material is presented. This is particularly true in the light of recent research that indicates that the method of instruction has a profound impact on subsequent learning styles, and even on which specific skills are learned more effectively (McGeown & Medford, 2014). In addition, there are many other factors involved in the acquisition of literacy (and reading in particular) which need to be incorporated into any framework for a teaching strategy that combine to form a complex matrix of linguistic, orthographic and lexical skills and knowledge in an environment that fosters generative learning.

Ellis and Moss (2014), suggest that there may be an optimal mix of phonics and whole-language learning which varies for individuals, dependent on, for example, the level of spoken vocabulary or letter sound knowledge with which they start school; it may even depend to some extent on gender or socio-economic status. Chapter Two considers the instructional reading texts that are associated with particular teaching approaches and focuses on the debate between the use of 'real' books and phonically-controlled basal readers.

Chapter Two

Basal Readers versus Real Books

> "Twas brillig, and the slithy toves
> Did gyre and gimble in the wabe;
> All mimsy were the borogoves,
> And the mome raths outgrabe."
> (from Through the Looking-Glass, and What Alice Found There by Lewis Carroll,1871)

Some phonological knowledge is necessary in order to read Lewis Carroll's poem. But it is possible to understand the meaning without knowing either the exact pronunciation of all the words or their exact definition. It is possible to understand this poem because it holds to the rules of English syntax and poetic forms, using words which resemble spoken English and allow the reader to infer meaning (Lucas, 1997). It is this ability, that even beginning readers have, to infer meaning from text without knowing all the words being read, or knowing how to pronounce them, that has led to the 'real books' versus basal readers debate.

The differences in approaches to teaching methods have inevitably created debate about the type of text that beginner readers should be introduced to. Should the text be only what the child can sound out at any given stage, or should a child be encouraged to explore 'real' books written by established children's authors and allowed to infer meaning from context and make use of illustrations. For any reading scheme, there are two issues: effectiveness (how much and what do the children learn) and efficiency (how fast and cost-effective it is). This chapter will examine the literature which looks at the instructional texts being used by teachers.

The 'Great Debate' (Chall, 1967), has been raging for many years amongst those involved in education and it continues: should teachers use reading schemes that gradually introduce words that the children will be able to sound out, or can children learn just as well from books that use natural language, including words that cannot be sounded out such as 'Once upon a time'? The debate has centred mainly on whether children should be taught to read starting with the smallest unit of text in a skills-based phonics approach, or with larger units such as whole words in a meaning-based approach. The perceived advantages for using a phonics-based reading scheme include: that only using words children can sound out guarantees success at reading words aloud for the majority; it is possible to monitor individual

progress; and it is easy to assess group progress. Perceived advantages for using 'real books' include: children developing an understanding of story structure, which aids prediction and comprehension; children learning to use context and grammar; children showing more motivation to read (Coles, 2004).

2.1 Historical context

Early reading schemes were developed by publishers, rather than educationalists or teachers, and based on very little research (Brown, 2000). Most of the early research focused on struggling readers and very little was peer reviewed, randomised or replicated until the work of Jean Chall, in 1967, who concluded that the systematic teaching of phonics would produce better results for all children, and supported the use of instructional texts that taught phonics, but she also emphasised that there was no one solution for all children (Chall, 1967). Reading books with texts that offered repeated practice with a small set of words and controlled vocabulary was thought to assist decoding (word reading) abilities, although there was no attempt at this time to align skills taught with the words to be read (Hoffman, Sailors & Patterson, 2002).

Reading schemes that were published during these years were not based on any particular learning theory; neither were they always systematic in their design or introduction of new sounds. However, research evidence available from analysis of texts found in both children's and adult literature had provided publishers and authors with lists of the most commonly occurring words. Some writers made use of these, known as Key Words to Literacy (McNally and Murray, 1962). In addition, there was on-going research into the development of the spoken language of children and into gender and socio-economic differences, as well as motivation to read, all of which had an influence on published children's reading schemes (Southgate et al, 1981).

Amongst some educationalists, there was a reaction against the development of reading schemes that used a contrived text, rather than more natural language. The work of Kenneth Goodman in the 1960s promoted the view that reading was more about language and comprehension of an author's meaning, than a perceptual process of recognising letters in patterns as words (Pearson, 2004). Advances in sociolinguistics and cognitive psychology, led to challenges to the conventions of reading schemes, and the emphasis on decoding at the earliest

stages of learning to read. In England during the 1970s and 1980s, the skills approach (using phonics and vocabulary control) was sidelined and the use of 'real books' (authentic children's literature) came to the fore. Reading schemes were still being used, but they shifted in their emphasis, in line with the whole-language approach (Pearson, 2004). Even so, teachers differed in their approach to the teaching of phonic skills; either expecting skills to emerge as a result of reading, or choosing to teach these skills explicitly. In the effort to improve the quality of literature, the systematic teaching of decoding skills was largely lost (Hoffman et al, 2002).

The concept of teaching children through a whole-language approach, by encouraging children to use context to predict text, based on their understanding of their own spoken language, was associated with the use of 'real books', rather than with reading schemes. By the mid-1990s in England this approach was widespread in schools, and associated with very little in the way of direct and explicit skills instruction. During the same period, an alternative view emerged, that of teaching the alphabetic principle beyond just initial letters, so that all words could be read as a sequence of sound-to-symbol correspondences.

In New Zealand, the method of choice for many years from the 1980s has been the whole-word, real-books approach, endorsed by consecutive governments, as a result of the perceived success of the Reading Recovery programme (Solar & Openshaw, 2007). The whole language/'real books' remedy for reading failure remains in place, while phonics has been sidelined. By contrast, in 1998, California prohibited the use of books where children could use contextual cues (Pearson, 2004), although, in the US National Reading Panel report (NICHHD, 2000) there was no explicit support for decodable text (texts that could be sounded out using the phonic rules currently in a child's repertoire) (Pearson, 2004). In 2000, the state of Texas introduced new rules for choosing texts for beginning readers. These were required to be decodable according to specific rules: each word had to be decodable (a word was considered to be decodable if all the letter-sound associations in that word had been previously explicitly taught); words had to follow a specified sequence of learning. In this instance, texts were to be analysed according to the number of rimes, repetition of high-frequency words, and the density of the text (Hoffman et al, 2002).

In England, in 2005, the House of Commons Education and Skills committee (House of Commons, 2005) recommended that research be commissioned to compare the use of text which is within a child's current decoding ability, with text which goes beyond it. However, this research was never commissioned and policy decisions were based on existing reviews and consultation. Advocates of the synthetic phonics method argued that using multiple strategies, such as those described in the 'searchlights model' (DfEE, 1998), confuses children, and that synthetic phonics is the only way to ensure effective reading. As a direct consequence, new reading schemes were written to comply with the government's insistence that only words that can be sounded out should be introduced in the beginning stages of learning to read (Marshall, 2011). This was based on the conclusions of the House of Commons Education Select Committee (2005) which states that:

> "Children are only taught to read through texts fully within their current phonological ability. So, although children might encounter words they do not understand, they are not given texts they cannot decode and are therefore not expected to infer words from context or syntax." (HMSO, 2005, p14)

This led to the development of specific criteria for educational publishers. The core criteria that the Department for Education issued for publishers of early reading books include that: children should always apply phonics as their first strategy to reading; children are taught high frequency irregular words; and that they practice reading using texts which are entirely decodable (DfE, 2012b, p1).

Book publishers, since 2010 in England, who wish to be endorsed by the Department for Education, are required to adhere to these criteria: phonics first and fast; daily sessions teaching the grapheme/phoneme correspondences; demonstration of blending from left to right; ensuring that phonic knowledge is used as a first approach even with irregular words; and that texts for reading should be entirely decodable so that the children learn to rely on phonemic strategies (DfE, 2010). Matched funding was provided for schools to aquire recommended texts.

For the primary school setting, the National Curriculum Framework in England (July 2013), statutory from Sept 2014, describes in detail the requirements for the use of the synthetic phonics approach and the reading books that are expected to be used. This includes the opportunity for "speedy working out of the pronunciation of unfamiliar words" (p14); that "pupils need to develop the skill of

blending sounds and establish the habit of applying this skill whenever they encounter new words […] supported by practising their reading with books consistent with their developing phonic knowledge" (p18); and "read aloud accurately books […] that do not require them to use other strategies to work out words" (p20).

As has been discussed in Chapter One, teaching approaches have largely fallen into two camps. These two opposing views can be summarised as: part-to-whole text first (the phonics approach), versus whole-to-part text first (the whole-word approach). For those who advocate a small unit approach (in which words are segmented into individual sounds or phonemes, children are taught to recognise each individual phoneme and to synthesise or blend the sounds in sequence to sound out the word), the books associated with this approach are purposely designed to provide reading material in which the phonemes can be practised (Wilson & Colmar, 2008). The text is therefore carefully controlled, only to contain phonemes that the children should already be familiar with. Where a whole-word approach is advocated, reading books need no such constraints regarding the choice of vocabulary. The whole-word approach is often associated with the use of books written by known children's authors, which stand alone, are not part of a scheme and are referred to as 'real books'.

2.2 Basal Readers

One solution to the problem created by having a large number of irregular words in English, when teaching the synthetic phonics approach, is to create special texts based solely on regular words (Goswami, 2005). Skills-based phonics instruction that focuses on 'small' units only, necessarily restricts the child's access to 'real books', since these are not restricted to regular words, and to decoding words that are familiar rather than extending sight vocabulary (Goswami, 1999).

Skills-based phonics teaching has become associated with the use of decodable texts, although one does not necessarily imply the other. Decodable texts generally emphasise common letter-sound correspondences, spelling patterns and high frequency words within simple sentences and basic story lines, as well as showing some links between the phonics represented in the text and the phonics being taught (lesson-to-text match) (Brown, 2000; Mesmer, 2001). The rationale is that this will reinforce students' current alphabetic knowledge and increase word

identification (Beverly et al, 2009; Mesmer, 2001). The assumption is that texts that follow regular letter-sound correspondences will be read more easily.

The introduction of nationwide testing of standards in literacy has had an impact on the debate (Hempenstall, 1997). It is easier, and certainly quicker, to test single word reading or letter sounds, than to assess levels of comprehension. There are, however, some concerns regarding the use of 'basal readers' or reading schemes (controlled teaching texts), such as: a lack of meaning within the text; a lack of authenticity; and inaccessibility (Coles, 2004; Hassett, 2008; Shannon, 2001). Basal readers have also been criticised on the grounds of having the potential to influence social attitudes and cultural perspectives in the design of the text (Shannon & Crawford, 1997).

There are alternatives to this approach, supported in the literature, such as the use of predictable text. This controls the vocabulary to emphasize: repetition; rhythm and rhyme; simple sentence structures; the inclusion of illustrations to support and extend text; restriction of amount of text per page; a simple plot; and increasing levels of difficulty (Brown, 2000). Although the vocabulary is controlled, it is not restricted to words that are decodable.

2.3 Real Books

An alternative form of text is known as authentic literature, or 'real books' (Pearson, 2004). In authentic literature, word choice and sentence structures are not controlled, and illustrations support and extend the text. There can be a complex plot, and there is generally more text per page than used in predictable or decodable text. In text which uses natural language, there is a presumption that a child's spoken language will assist in word identification (Mesmer, 2001). The 'Early Steps' reading intervention programme, developed in the 1990s, and used successfully with struggling readers, employed natural language reading materials rather than decodable text (Morris, Tyner & Perney, 2000).

There is evidence that the act of reading predictable text increases comprehension, and that children are able to learn new words by reading them in context (Hatcher et al, 1994). Bus and van Ijzendoorn, (1999) provide evidence of the importance of an explicit linkage between phonological awareness and word reading, and that training either skill in isolation is not particularly effective. A

phonological-linkage theory suggests that phonological training and reading skills need to be integrated (Hatcher et al, 1994). The implication is that children will learn more effectively by reading text which has meaning. Moseley (2004), found that young children were able to recognise short words more accurately than long words, regardless of their frequency or regularity, and concluded that there is probably no need artificially to restrict early reading to a small core of high frequency words or regular words.

'Real books' are written by a variety of authors, and there are no planned links between them. The content is neither constrained, nor restricted in its use of phonics, or vocabulary. The expectation is that children will develop skills through repeated exposure to words, in a motivating context (Coles, 2004; Goodman, 1976; Smith, 1973). 'Real books' have the advantage of being authentic text, and often have more illustrations than basal readers. However, because organising 'real books' into some kind of learning sequence is both difficult and time consuming, young readers can find themselves faced with text which comprises inaccessible words and over-complex sentence structures. For struggling readers, to be faced with a book in which very few of the words are accessible, can be highly de-motivating. However, polarising the debate, as 'real books' versus basal readers, is misleading. The question is whether or not a child should be faced with words that are beyond their current decoding ability and should these words be in a meaningful context.

A rarely mentioned, but significant question in the 'great debate', relates to the value given to books by teachers, children and parents. It has been argued that where the main focus of attention is on decoding print, rather than following a story, books can become devalued (Campbell, 2007). In addition, it has been argued that the use of 'real books' is more likely to encourage parents to engage in dialogic (shared) reading with their children, which has been found to have positive observable effects on children's general language development (Hay & Fielding-Barnsley, 2007). Basal readers are clearly designed for teaching text, and parents/carers will see their role as helping the child accurately to decode, whereas 'real books' are more likely to be used to encourage an understanding of text.

The whole-word approach to teaching reading was based on the assumption that knowledge of syntax, semantics and phonology, would allow the reader to make predictions through contextual cues. Whole-language theory suggests that

children can learn to read by reading (Goodman, 1989), when learning to read is always set in the context of natural language. The instructional reading materials used comprise meaningful literature, rather than meaningless texts. Skills were to be inferred from reading authentic text that is characterised by: a rich, non-controlled vocabulary, with complex sentence structures; quality illustrations that support and extend the text; and natural oral language assisting in word recognition (Brown, 2000; Mesmer, 2001). It is assumed that children learn elements of the alphabetic code by association, and that these need not be explicitly taught. According to Wilson and Colmar (2008), this approach involves guessing from context and fails to teach the skills needed for alphabetic decoding. The notion that children can learn to read by being immersed in it is, according to Foorman (1994), a fallacy, since literacy does not 'emerge' in the way that spoken language does. Instead, it is trained through exposure to print and feedback from adults for correct pronunciation or meaning. A reliance on exposure to 'real books' alone may be sufficient for some children but others are likely to need a more structured approach. Nevertheless, proponents of the use of 'real books' have highlighted the importance of meaningful text.

2.4 What the Evidence Says

The main question that arises from this debate is whether or not instructional texts should be used. If yes, then the kind of instructional texts that are most effective needs to be established, as well as which texts are best for whom. It could be argued that, more importantly, what needs to be established, is which type of text is best suited for which type of student and when (Brown, 2000).

The study of the US Reading First programme, which was reported in a review of elementary reading programmes (Slavin, Lake, Chambers, Cheung & Davis, 2009), suggested that an increase in phonics work had small effects on word decoding and no impact on comprehension. The review looked at effect sizes of reading programmes with normal populations of beginning readers. The authors found that overall there were stronger effects for decoding, than comprehension, but an emphasis on phonics did not guarantee positive effects. There were no significant positive effects found from using decodable texts. Instead, the review suggested that programmes should focus on instruction and professional development. The evidence did not support the idea of introducing materials, and

instruction, with a strong emphasis on phonics, and the effects of using phonetic textbooks were found to be very small.

There was a brief mention of the use of decodable texts in the National Reading Panel Report (Ehri et al, 2001a), but none of the trials treated text as an independent variable. Surprise was expressed by the authors on the lack of research in this area, but no specific recommendations regarding texts was made in the report. Instructional texts used in schools can vary between those that emphasize: predictability; high-frequency words; phonetic control (consistent letter-sound correspondences); and 'real books'. Very few studies have attempted to separate the effects of decodable texts from the effects of instruction.

Hoffman et al (2002), made a study of the five sets of basal readers that were accepted by the Texas Education Authority, following their change of policy in 2000. They based their study on a theoretical framework that considered three main aspects: instructional design; accessibility; and engagement quality. Instructional design incorporated the alignment between skills taught in the classroom, and the words to be read in the text, without specifying the skills (analytic or synthetic phonics). Accessibility is described as both decodability (high frequency and phonically regular words), and predictability (context, pictures, and repeated phrases). Engagement quality referred to content and motivational aspects of the text. The authors' analysis of these basal texts, suggested that although they were accessible (as described above), they were less predictable, with less contextual support, and had less engagement quality than texts that had been approved under previous Texas education policies.

A study of at-risk first grade children (Jenkins, Peyton, Sanders & Vadasy, 2004) compared less decodable texts (fewer words could be sounded out and illustrations provided clues), with more decodable texts (the majority of words could be sounded out). The children were taught in tutor groups, and a control group was included in the trial. The children were randomised to intervention condition. In order to control for the teaching variable, there was a concerted effort made to ensure fidelity to a protocol of instruction. All students in both intervention groups experienced the same instructional content. The authors made a detailed analysis of both texts, which were significantly different in terms of decodability. The authors found effects from tutoring, but no significant differences from the use of different

texts. The authors suggested that tutoring had a much greater effect, than the type of texts used, although they pointed out that children had access to other types of texts in their classrooms. The authors concluded that decodable texts did not add value to tutoring programmes for at-risk beginning readers.

In contrast, Foorman, Francis, Fletcher & Schatschneider (1998) found greater gains in word recognition and comprehension for children who were explicitly taught letter-sound correspondences and synthetic phonics, and who used decodable texts, compared to children taught analytic phonics and predictable texts, and a further group, with no phonics instruction and predictable text. However, the reading texts could have been a confounding variable, as these were not the same in all conditions. Neither was it clear which predictable texts were used in the two conditions. In addition, the authors admit that all conditions had a significant literature base, and as it involved at-risk readers, it was not a typical population.

An analysis of two English reading schemes and a selection of 'real books,' carried out by Solity and Vousden (2009), aimed to compare the use of high-frequency words and phonics skills. Their focus was on the number of most-common grapheme-to-phoneme correspondences (64 according to the authors), and most frequently used words. The authors anticipated finding more high frequency words, and phonically regular words, in the reading schemes. The authors found no significant differences in use of the 100 most frequently-used words, between 'real books' and one of the reading schemes (Oxford Reading Tree). There was significantly less occurrence in the other reading scheme (Rhyme World), than the 'real books'. The figures for grapheme-to-phoneme correspondences were more difficult to analyse, and suggested very little difference between the three sets of children's books, although no calculation of significance was given. The authors found that there was consistency in both the reading schemes, which was not apparent in the 'real books'. They made no attempt to compare learning outcomes from the use of the reading schemes with the 'real books', nor was any mention made regarding alignment with reading instruction. However, they did suggest that it is possible to use 'real books', as instructional texts, if carefully selected and appropriately levelled.

Whilst investigating children's motivation for reading, tutoring in phonemic awareness, decoding, fluency building and the reading of controlled text, were not

found to boost motivation in poor readers, even when decoding skills improved (Morgan, Fuchs, Compton, Cordray & Fuchs, 2008). Morgan et al (2008) suggest that targeting skills may be insufficient to remediate the lack of desire to read, which appears to have set in even before starting school. However, they did not make reference to the decodable text being used, which may have had an influence on motivation (Solity & Vousden, 2009). Beginner readers are unaware of whether words are regular or irregular, high or low frequency, or rhyme with a family of other words; it could be the interest of the texts used rather than their graded vocabulary or phonic complexity that is of most importance (Moseley, 2004).

A comparison of the use of phonics-only instruction with phonics plus decodable texts, in a study of young Taiwanese children learning English, (Chu & Chen, 2014), led the authors to postulate that a more meaning-based text might offer greater advantages in this context. The phonics-only group followed a mainly synthetic approach, although there were elements of analytic, or analogy, activities. The phonics plus decodable text group followed a similar approach; however, they had additional activities, using non-decodable, high-frequency, words. The decodable texts conformed to the principles of: consistent letter-sound rules; simple sentences; and matching the content to instruction. Both groups of children improved in their word reading, but there was no significant difference at post-test. However, the authors re-tested the children after two weeks, and found a significant difference in favour of the group that had used the decodable text.

It seems that the language found in 'real' story book literature is both predictable and engaging (Coles, 2004). Young children become familiar with the style of language from an early age, and this familiarity will assist comprehension as they move towards independent reading. Joint book reading exposes children to written language, which is different from spoken language in its structure (Clark, 2014b). Storybook reading is one of the most important activities for developing pre-reading skills. This was the conclusion of the authors, of a meta-analysis of research carried out in 1995 (Bus, van Ijzendoorn & Pellegrini, 1995). Indeed, they suggest that joint book reading, with an adult, is as strong a predictor of reading success (particularly with regard to comprehension) as phonemic awareness.

Results of a small-scale trial comparing the use of decodable texts, with a group who only listened to authentic literature being read aloud (for example The

Tale of Peter Rabbit by Beatrix Potter), produced unexpected results. Children with below average reading, prior to the intervention, made gains in comprehension using the decodable texts. Average readers made greater gains in both word recognition, and comprehension, but only in the literature group. Indeed, the only group to make significant gains on all measures was the literature group. The struggling readers improved in comprehension (although this was not being directly taught) using the decodable texts, but this kind of text appeared to be harmful to the comprehension of average readers. The authors concluded that decodable texts were useful for some beginning readers, but not helpful for more advanced readers (Beverley, Giles & Buck, 2009).

By contrast, Mathes and Torgesen (1999) claim that although there is no evidence from US state-funded research (NICHHD, 2000) advocating either a skills-only approach or the use of decodable texts, the use of decodable text is nevertheless justified. The authors draw attention to the lack of research designed to validate the use of decodable texts, or to examine the balance between decodable text and 'real books'. However, they suggest that this is due to an assumption by researchers, that decodable text would be a logical component in a skills-based programme. They go on to propose that expecting children to read texts they cannot decode is not "relevant reading" (p12).

A review by Mesmer (2001), found only three studies focusing on decodable text in refereed journals. The first study (Juel & Roper-Schneider, 1985), compared a decodable basal (containing more repetitions and a wider vocabulary), with a basal that used high frequency words (fewer repetitions and smaller vocabulary). The decodable group showed advantage over the high-frequency group on a decoding measure, and in addition, the use of a decodable text appeared to affect decoding strategy more than phonics instruction; there were no differences on measures of comprehension or sight word vocabulary. A second study in the review related to at-risk readers, but it was unclear what the two treatment conditions were; these were merely described as code or context treatment (Felton, 1993, cited in Mesmer, 2001). The code group showed advantage on reading pseudo-words. The third study mentioned did not actually compare decodable with non-decodable text. The evidence in this review is not compelling, and yet Mesmer concludes that "phonics lessons and decodable text must be paired" (p136).

As an alternative to decodable texts, some reading schemes have been based on what has been referred to as predictable text. This uses patterns and repetitive language, rhythm and rhyme, whereby children use repeated, and shared reading, and memorise the text (Johnston, 2000). The teaching method associated with these texts is repeated reading. However, Johnston (2000) found that children made greater gains when they learned a bank of sight words, in isolation from the text. In this instance, predicatable text was not linked with a phonics approach.

In a comparison of three different US reading programmes, which all included decodable texts, it was noted that one of these programmes, Reading Mastery, differed considerably in its decodable content (Hiebert, Martin & Menon, 2005). In this programme, the texts were designed to minimize contextual aids (for example having illustrations on the next page), and included nonsense phrases. All three programmes differed in the number of new unique words introduced at any one time, and in the degree of predictability in the text. These differences exemplify the difficulties in trying to isolate decodability as a variable. A more recent study (Mesmer, 2009) made comparisons of reading accuracy and reading rate, in decodable versus predictable text. The results suggested that for poorer readers, decodable text led to greater accuracy, but not for more skilled readers. The reading rate was greater, using predictable text, for all readers in the trial; levelled texts were read at a faster rate. However, there were a number of limitations, acknowledged by the author, such as: lack of randomisation to condition; the effect of differing classroom instruction; and the differences likely to be inherent in using different reading schemes (for example, number of high frequency words, quality and frequency of illustrations). In addition, only two books were used in each group (there were two groups for each condition; a higher level and a lower level), and there was no control condition. Mesmer concluded that different texts provided different advantages, particularly for less-skilled readers.

An intervention, designed for teaching reading to children with learning difficulties, combined both analytic and synthetic phonics, as well as teaching high-frequency words (Wright, Conlon, Wright & Dyck, 2011). Teachers were instructed to model words when the children were unable to blend sounds after several attempts. Irregular words were taught through the use of flash cards; the programme used an eclectic approach to instructional methods. The text to be used was purposely written, and designed to be decodable, predictable, and aligned with

instruction; similar to that of a reading scheme with controlled vocabulary. Although there was no control group, the authors calculated gains to have been made on the basis of clinical significance (reliable change and a minimum standard score, p7) in word and non-word recognition as well as comprehension.

In a small, and much older, study of 32 young readers in the 1960s, children who were already reading fluently before starting school were questioned about their strategies when encountering unknown words (Clark, 1976). Their replies included: guessing, missing words out, and asking someone. Clark suggests that the evidence from these children indicates that the complexity of a text may, in fact, aid understanding. She goes on to suggest that an apparently simplified text may make the reading material not only less stimulating, but also more difficult to comprehend. Looking back on her work, Clark concludes that complexity of text is not necessarily more difficult, and that simplicity is not necessarily easier (Clark, 2014a).

A more recent study that focused on sight vocabulary instruction, which taught words that go beyond basal reader vocabulary, showed significant gains in word reading skills, as well as knowledge of taught vocabulary. This is evidence that the simplistic vocabulary used in basal readers may be too restrictive to generate meaningful vocabulary (Duff et al, 2008). Exposure to vocabulary-rich text may enable children better to understand text, even when used in complex sentence structures (Baumann, Ware, & Carr Edwards, 2007).

2.5 Summary

Two themes seem to emerge from the literature. Firstly, that the most significant predictor of reading ability is the method of instruction, and secondly that this needs to be a complex mix of strategies, especially in a mixed-ability setting. There is no clear picture regarding instructional texts; the effectiveness would seem to depend on the level of integration of the instruction with whichever text type is used and the degree of predictability of the text for meaning. The literature also highlights the need to consider potential confounding variables, such as instructional method, in any comparison of instructional texts. This is taken up in Chapter Three, which details the design of the intervention for this research, in which method of instruction is controlled for, in order better to assess two differing instructional texts.

Chapter Three

Design of the Intervention

> "A child's greatest achievements are possible in play, achievements that tomorrow will become her basic level of real action."
> (Vygotsky, 1978, p100)

The Intervention designed for this research needed to achieve two distinct objectives: to control for variables in the learning context and to provide an opportunity to compare two different sets of vocabulary within meaningful text; to replicate the kind of text found in both 'real' books and basal readers. The children for whom the Intervention was intended were in the Reception classes of schools. Reception is still within the Early Years sector and some children were still only four years of age in the September in which they started school. Therefore, learning materials were designed to be informal play-based group activities. This chapter considers first the role of play and group activities in learning to read. Second, it describes the methods used, and the rationale behind the training and support materials provided for the programme deliverers. Third, it details the materials used by the children and the rationale behind the design of each element, and how these were developed. Finally, the books themselves are described, including the rationale used to select vocabulary, font, word frequency, illustrations and story line.

3.1. The Role of Play and Group Activity in Learning

The Early Years Foundation Stage (EYFS) statutory requirements list some of the Early Learning Goals as: children learning through planned, purposeful play with a mix of adult-led and child-initiated activity; listening attentively and responding appropriately; following instructions involving several ideas or actions; showing awareness of listener's needs; using correct tense and developing narrative; motor control of small movements; handling equipment and tools effectively; confidence to try new activities and to speak in a group; ability to work in a group; understand and follow rules; play cooperatively; take turns; show sensitivity to other's needs (DfE, 2012a). All of these early learning goals are addressed in the design of the ωeebee Reading Programme (Mace, 2014) developed for this research, which focuses on learning through play, alongside the central aim of teaching word recognition. This focus on play allowed for easy adoption of the programme into

typical Reception classes, using familiar teaching approaches. Learning through play is not a new concept. There have been many proponents of this approach over time (Burnett, 2007; Pickett, 1998; Willam, 2009). For children, play is an appropriate learning medium to engage in (Stephen, 2010).

Some structured play activities, such as puzzles, games and constructional materials, appear to engage and stretch children's learning more than free play particularly where there is adult-led enrichment (Sylva, 1984). Part of the training for the weebee Reading Programme involved detailed examples of how the programme deliverers should lead the games and activities, including the kind of language to be used, and how to draw all the children in the groups into the activities. This included training to develop skills for acting as a facilitator (extending children's thinking and strategy); modelling language; joining in; and reiterating children's utterances. This approach, rather than focusing on teaching skills, is more child-centred and focuses on creating opportunities for children to learn through play.

Piaget and Vygotsky both saw play as an essential context for children's learning (Dombey, 2005). Vygotsky considered play to be pivotal for children's development, as the means by which they make meaning of their environment. Indeed, Vygotsky referred to children learning not from being taught, but from discovering skills in play situations (Vygotsky, 1978). This kind of learning has been described as 'tangential learning', whereby children learn through being covertly exposed to knowledge, rather than being actively taught in a structured way. Tangential learning is a mechanism by which a child can learn, by being exposed to things in a context in which they are already highly engaged (Nahachewsky, 2013). Through play, the individual is given the opportunity to map information in a way that is most efficient for them; for example, through verbalisation. Diversifying experiences have been found to benefit cognitive processing (Ritter et al, 2012), as well as providing children with multiple and varying learning experiences, which leads to greater inclusiveness.

There is strong research evidence to support the use of group learning (Abrami, Lou, Chamber, Poulsen & Spence, 2000; Hatcher et al, 2006; Pickett, 1998; Siraj-Blatchford, 2009). In a review of effective reading programmes, clear evidence was found that the most effective programmes were based on cooperative

learning, with young children working together on structured activities in small groups, helping one another (Slavin et al, 2009). The authors concluded that structured peer-to-peer interactions had positive effects, which were consistent with similar findings in older age groups. In cooperative learning, children can benefit from greater motivation, immediate feedback and learning, in a safe space (Tracey, Chambers, Slavin, Hanley & Cheung, 2014).

One of the benefits of group learning is the opportunity for what is referred to as 'joint attention'. Typically, this is when a child and an adult are focused on the same thing, although joint attention also occurs between two children. Joint attention is the name given to the mechanism that infants use to learn language and social behaviour; for example, when a child follows the gaze of an adult reading and pointing to words or pictures. It is primarily the awareness that the experience, being jointly attended to, is a shared experience (Akhtar & Gernsbacher, 2007). It also includes social referencing, pointing, and gaze following. During joint attention there is an appreciation of others' communicative intentions and a desire to imitate. Children will follow an adult's, or other child's, gaze to an object, and thereby learn which cues to focus on. An example would be the adult pointing to words on a page, and indicating, for instance, the first letter of a word during guided reading. The child learns at the same time the properties of books: that print has meaning; that we read left to right and start at the top of a page; that illustrations help to tell the story; that pages need to be turned in sequence and that the book has to be the right way up (Vaughan Van Hecke et al, 2012). These orientation factors have been applied in the design of the group activities and games developed for the *weebee* Reading Programme, whereby children needed jointly to attend to both the adult and the other children to learn the rules of play.

3.2 Training and Support for Programme Deliverers (teachers)

Evidence from existing successful programmes used in schools, suggests that training of programme deliverers, which includes some professional development, increases the effectiveness of the associated interventions. Ideally this should involve initial training followed by regular support visits (Morris et al, 2000; Tracey et al, 2014). Delivery of the *weebee* Reading Programme required only limited planning or assessment, and therefore did not require delivery by a qualified teacher, but did need careful supervision and training. Most Reception classes have

a Teaching Assistant available, who is familiar with the children in the class, and is used to working in group situations. It was felt to be important neither to burden the Class Teacher with extra work, nor to cause the normal curriculum to be affected. The ωeebee Reading Programme was delivered, therefore, mostly by Teaching Assistants, although some schools preferred the use of governors or the class teachers. (Throughout this Thesis, all the programme deliverers are referred to as teachers.)

Training was given to all teachers. This included instructions for the rules of play for each of the games, but also the kind of language to be used, such as modelling the pronunciation of words and drawing attention to initial letters (also included in the teaching Manual). The importance of ensuring that the children should not at any time feel under pressure to read the words aloud during a game was emphasised, but it was also made clear that children should not be prevented from doing so if they chose to. Teachers were asked to verbalise all the words during the games, and if a child struggled visually to match words, to draw their attention to the initial letter shape in the first instance, and then to any other salient features. Emphasis was put on keeping the games moving at a fast pace; for example, by the teacher verbalising the words for the children, in order that the children would perceive the focus to be on playing the game, rather than on reading the words. This served a number of purposes, such as increasing the likelihood of fitting in all the activities into a restricted timetable, maintaining children's motivation, and encouraging a quick response to recognising initial letters and salient word features. As part of the training, teachers were first shown a presentation which detailed a number of principles underlying the development of the ωeebee Reading Programme, and the overarching aims that related to these principles, primarily related to a mixed method approach to teaching beginning reading (as shown below). Both the principles and overall aims on which this programme was developed are an abstraction of the research evidence found in the literature detailed in Chapter One. The research evidence points to a consensus in the literature that there is no single method to suit all children, and that most children appear to benefit from a range of teaching approaches; that no one approach is sufficient on its own and that children benefit from different starting points (Allen, 1998; Morris et al, 2000; Snowling & Hulme, 2012).

Underlying Principles:

- All children can learn to read, but they do not all do it the same way.
- Synthetic phonics is good, but it is not sufficient on its own.
- Some children are unable to blend words.
- Other phonic techniques exist and can also be effective.
- Some children need to begin with a bank of sight words.
- A bank of sight words is good for all children, especially common exception words.

Aims of the Programme:

- To give children the opportunity to learn words in different ways, in addition to synthetic phonics.
- To foster the recognition of letter patterns using 'real' words.
- To build a wide sight vocabulary including many of the 'key words to literacy' (McNally & Murray, 1962).
- To extend oral vocabulary through discussion, songs and stories.
- To motivate children through the use of games.

From a research methodology perspective, in addition to reducing the number of variables present in the reading and learning materials, it was important to limit the differences in delivery of this programme across trial schools. Therefore, a detailed Instruction Manual was prepared and training sessions given, followed by regular support visits. All teachers were given a copy of the Manual, which detailed the programme (see Appendix A). There was an introductory section describing the overriding principles, such as not expecting children to read the words aloud during the activities (other than reading the book), not expecting the children to 'sound out' the words (but allowing it if the children chose to), and encouraging the children to work collaboratively in their groups. The procedure for the activities for each book were described in detail, including a list of resources needed (see Appendix B). These were set out for each term, week and session, for which records were to be kept. Instructions for each session had a detailed introductory whole-class activity to promote discussion and foster the learning of new oral vocabulary using a 'Big Book'. It also had details for the whole-class plenary activity, during which the phonic work from the session was revisited. The resources were all provided by the researcher, including weekly and individual check lists (for pupils

and teachers, see Appendix C). Instructions were also given for the procedure to be followed when hearing the children read the session book after they had completed all the preliminary activities.

3.3 Teaching Resources

In order to address the need for multiple approaches, a range of activities were developed. As all the participating schools taught synthetic phonics, according to the national curriculum statutory requirements (DfE, 2014a), it was assumed that all the children would be exposed to this method as part of their usual curriculum. Therefore, the activities focused on alternative phonic approaches (including teaching single and complex grapheme-phoneme correspondences) and whole-word strategies to build a sight-word vocabulary. Each element of this programme is detailed below (see Section 3.3.1 to 3.3.4) and for each there is a description, followed by the rationale drawn from the research literature. The table below shows the general structure of the weebee Reading Programme and the target skills.

Table 3.1
showing the overall structure of the weebee Reading Programme

Intervention	Conceptual emphasis	Target Skills	Elements
weebee Reading Programme 60 lesson plans	Word reading and comprehension	Vocabulary (spoken)	Songs Big (discussion) Book
		Analytic phonics	Initial/final letter windows Happy word families
		Sight word recognition	Fishing Jig words Bingo Memory game Dominoes Snap Pento Grog's Journey Snakes and ladders
		Comprehension	Sentence matching Reading books

Almost all the teaching materials designed for the programme were group activities or games, reflecting an informal approach to teaching considered to be appropriate for the age group (4-5/6yr olds). There were no games or activities for formal teaching of letters, words or sounds. The games were designed to be played

in pairs or groups, and there were clear instructions to teachers to allow the children to assist each other, and discuss strategies for success. All the games relied on visual recognition, and were intended to encourage visual mapping skills and statistical learning (see Section 3.3.3). A number of the activities were designed to direct the attention to onset-rime or initial letter positions. There were seven games or activities per book, in order to encourage diversification and the generalisation of learning. During the first term in which the programme began, activities focused on word recognition. The second and third term resources continued to build word recognition skills, but also had an additional emphasis on comprehension. Predictors of reading comprehension have been identified as: reading accuracy, listening comprehension and oral vocabulary (Protopapas, Mouzaki, Sideridis, Kotsolakou & Simos, 2013). Therefore, in addition, resources for the second and third terms included activities designed to extend children's oral vocabulary and narrative skills, drama and story-telling (Bowyer-Crane et al, 2008; Clarke, Snowling, Truelove & Hulme, 2010; Fricke, Bowyer-Crane, Haley, Hulme & Snowling, 2012).

3.3.1 Resources Designed to Extend Oral Vocabulary

The two activities included to extend oral vocabulary were: a CD of songs about the story book characters, and a big picture book. These activities also provided background knowledge, in the context of the associated books, to support comprehension.

3.3.1.1 Songs

The opening activity for each session involved the children learning, and joining in with, purposely-written songs about the characters that feature in the books. This was intended to serve as a familiar way to begin each session, to help children engage with a routine in a motivating way. In addition, it was designed to assist their oral vocabulary development and allow the children to learn predictable phrases that would be reflected in the books. The use of songs also addresses one of the Early Learning Goals (for Expressive arts and design), which requires that children sing (DfE, 2012a).

Theorists such as Goswami and others (Goswami, 1999; Goswami et al, 2011; Holliman, Wood & Sheehy, 2010; Imrie, 2008; Leong, Hamalainen, Soltesz & Goswami, 2011; Thomson, Fryer, Maltby & Goswami, 2006; Treiman et al, 1995)

suggest that there are connections between pre-school awareness of rhyme, rhythm and alliteration and later progress in reading. This is partly explained by the use of analogy; that is, recognition of spelling patterns of familiar sounds. Goswami (2005) has found evidence that if children are not directly taught to use rhyme analogies, they will develop the strategy naturally, but slowly, for themselves. It appears that use of rhyme is a valuable natural tool which children use (Hindson et al, 2005). Promoting rhyme is thought to help children to learn more effectively from subsequent phoneme awareness training (Allen, 1998; Hindson et al, 2005). Rhymes in English are often represented by spelling patterns that are consistent, unlike much of English spelling (Goswami, 1999). Traditional nursery rhymes are an ideal technique for teaching recognition of sound patterns. Children who have learned nursery rhymes at the pre-school age have been found to have an advantage in early word recognition (Goswami, 1999). Interestingly, children who are more sensitive to rhyme are also more sensitive to phonemes, and so are better able to learn the grapheme-phoneme correspondence (Goswami, 1999). The songs used in the *weebee* Reading Programme make use of both rhyme and alliteration in the lyrics.

Rhythm also appears to have an impact on early reading. A study of 5 to 7-year-olds, suggested that sensitivity to both speech and non-speech rhythm were independently predictive of reading attainment (Hollimanet al, 2010; Kuppen et al, 2011). Processing rhythm is thought to be the first strategy for segmenting speech used by infants (Goswami et al, 2010). Rhythm in English is linked to the syllabic structure of words, including stressed or non-stressed syllables (Goswami et al, 2010). Studies looking at the effect of using rhythm with young infants suggest that music, singing or other metrical activities can have benefits for phonological development and the ability to recognise segments of speech pattern (Goswami et al, 2010).

There is evidence that stress perception is linked with dyslexia (Leong et al, 2011). This is important for segmenting words and syllables in the speech stream. Leong et al (2011) found metrical perception (rhythm in speech) to be a stronger predictor of reading development than phonological awareness, or other auditory measures in dyslexia. When a child is asked to blend the sounds of a word (as in the synthetic phonics method) they may be unable to reproduce the correct stresses that would normally be associated with that word (Leong et al, 2011). This affects

their comprehension of the word they have sounded out. Research by Vandermosten et al (2010) found evidence to support the hypothesis that the core auditory deficit in dyslexia is the processing of sounds containing rapidly changing temporal cues, suggesting that they might have problems extracting and distinguishing the relevant cues (speech stream) from the environmental acoustic information (background noise). In view of this evidence, all the songs used in this programme are very rhythmic, and a professional singer was used for the recordings to ensure accurate pronunciation (using Received Pronunciation) with correctly stressed syllables.

It was intended that there should also be a motivational element to the songs. Following a study of five-year-old boys, Lever-Chain (2008) suggests an emphasis on formality can negatively affect attitudes of young readers to reading. In a study looking at reading for pleasure, Warrington & George (2014) cite a number of authors who have demonstrated that reading for pleasure, which is determined by motivation, is correlated with attainment. This includes word reading, but also comprehension, grammar and increased sight vocabulary. Children have also been found to be more motivated to read when there is an element of personalisation connected with a text (Kucirkova, Messer & Sheehy, 2014). By introducing the children to the characters in the weebee Reading Programme books before reading them, they had already become familiar with the characters, where they live, and some of their characteristics. This was aimed at creating a personal connection between the children and the characters; eliciting empathy; and increasing curiosity about the books themselves.

3.3.1.2 Big (discussion) Book

The Big Book was designed to be used at the beginning of each session as a discussion activity for the whole group. There was a large illustration, related in some way to the books, and depicted in the same style. The manual gave three or four suggested questions as starting points for the Big Book discussion, but teachers were asked to allow the discussion to flow from the children as much as possible. This was intended to be an opportunity to extend children's spoken vocabulary, put many of the story book elements into a context that the children would be familiar with, and assist with building background knowledge to support inference generation (discussed in more detail later). In addition, the Big Book directly

addressed one of the Early Learning Goals; that of world knowledge, whereby children should have opportunities to talk about facets of their own immediate environment, and make observations of animals and plants (DfE, 2012a). The Big Book was directly related to the story books that centred on animal-like characters in a woodland, rural environment. From book 4 onwards, the Big Book also included a plenary activity to practise the phonic work from the session. The Big Book focused on the oral dimension of language learning (see Figure 3.1 below).

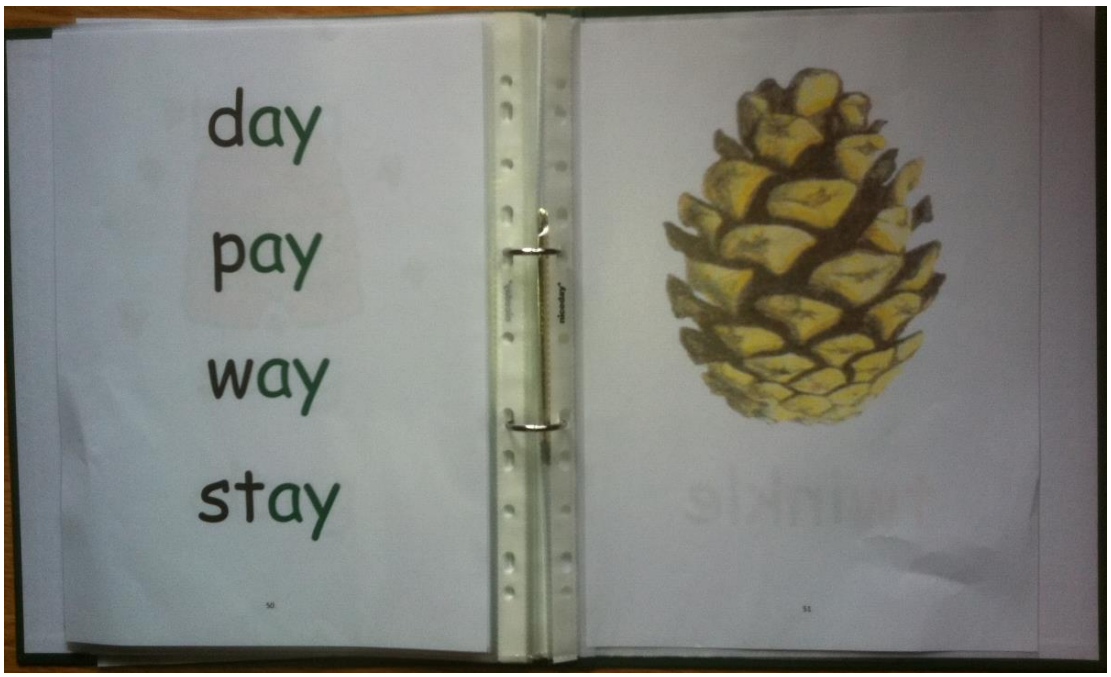


Figure 3.1. Example pages from the Big Book

Clearly, if children have poorly developed oral vocabulary knowledge, they will have difficulty identifying, and assigning meaning, to unknown printed words. Research indicates that oral vocabulary knowledge contributes to the development of decoding skills and real word recognition (Tunmer & Chapman, 2012). In view of the emphasis given to the importance of phonological awareness and phonological skills, in the theoretical literature, there is good reason to place an emphasis on speaking and listening skills, which are associated with comprehension, in the teaching of young children. A number of studies have focused on the inclusion of specific oral training such as vocabulary, inference generation, expressive language and speaking and listening skills. The research evidence suggests that children who receive specific training, in oral language and vocabulary instruction, improve on literacy skills and comprehension (Baumann et al, 2007).

The importance of oral vocabulary knowledge, for reading comprehension, is supported by studies which have demonstrated that as reading develops, there is an increasing role of oral vocabulary, and decreasing role of decoding skill, in the prediction of reading comprehension; talk aids the development of new learning (Protopapas et al, 2013; Silby & Watts, 2015). These studies have consistently found a large, and increasing, proportion of reading comprehension variance attributed to oral vocabulary measures. In addition, having a greater expressive vocabulary has been shown to assist children to benefit more from phonologically-based interventions (Duff et al 2008).

Interventions, that have focused mainly on oral vocabulary, narrative and listening skills, have been found to be effective in improving oral language, and spoken language, and have a significant impact on reading comprehension (Fricke et al, 2012; Muter et al, 2004). Ricketts, Nation and Bishop (2007) found that oral vocabulary skills predicted some word recognition skills, but not others, and suggest that there is a link between oral vocabulary and exception-word reading, whereby meaning-based information has a direct influence in the word recognition process itself. They conclude that oral vocabulary knowledge can aid word recognition, especially when the word has inconsistent spelling-to-sound correspondence. Other studies have reported similar findings which suggest that oral vocabulary knowledge assists recognition of irregular words (Hay & Fielding-Barnsley, 2009; McGeown & Medford, 2014; Muter et al, 2004).

Of course, the ultimate goal of reading is comprehension; extracting meaning from the text. According to Compton et al (2014), there are two kinds of mental representations: text-based (which is explicit), or situation model (using the reader's background knowledge to combine with the text to form a deep representation). This background knowledge can be as little as the title of a story, but will assist with the building of a situation model. Results from a study, by Compton et al (2014), showed that having some background knowledge increased the likelihood of correct answers about the passage. The authors suggest that background knowledge may serve as a compensatory tool, and enhance a reader's ability to recognize the text's most important ideas; that inference generation and comprehension are aided by general knowledge, passage specific knowledge and oral vocabulary knowledge.

An essential element involved in the comprehension of text involves inference generation (Clarke, 2009). Building background, and world knowledge, is fundamental to inference making in reading. It is important to teach children to read between the lines, and this can be done by: introducing new oral vocabulary; reminding children of relevant vocabulary; and discussion about the text before it is encountered. The use of pre-reading activities can build, what Compton et al (2014) refer to as, a micro world. This can include use of prior knowledge, understanding character, discussing possible consequences, and making predictions. These activities can help children to construct a situation model while they are reading. Since oral vocabulary skills and inference generation may be related to comprehension (Ricketts, Bishop, Pimperton & Nation, 2011), explicit teaching of text-related vocabulary will assist in text comprehension, inference generation and construction of situation models (Graesser, Singer & Trabasso, 1994). According to Graesser et al (1994), children may give up the search for meaning if the perceived goal is not to construct a meaningful situation model, but merely to sound out a word as requested by the teacher.

Difficulties in reading comprehension have been found to be associated with a deficient knowledge of word meanings and grammar (Snowling & Hulme, 2011). Children who: learn new vocabulary; study story structure; participate in question generation, and answer questions, have shown improvements in comprehension (Garner & Bochna, 2004; Snowling & Hulme, 2011). Class or group discussion, which increases world or background knowledge by extending vocabulary and knowledge of grammar, is thought to help children to generate inferences from text, and fill in information not directly expressed in the text (Johnston & Barnes, 2008; Williams, 2014). A specific strategy, that Johnston and Barnes (2008) refer to as 'preview strategy' (p129), is used to prime children's general knowledge, associated with a text, in order for them more easily to make predictions about the text. This is very much the role of the Big Book used in this programme.

The Simple View of Reading model (described in Chapter One, suggesting two dimensions of decoding and language comprehension) has recently undergone a revision by one of its original authors (Tunmer, in Tunmer & Chapman, 2012), following research that has suggested that oral vocabulary knowledge has a mediating role to play in reading comprehension. A number of researchers have found that vocabulary contributes to both decoding skills, and real word recognition

(Høien-Tengesadal, 2010; Protopapas et al, 2013). Vocabulary has been found to affect connections between orthography, phonology and the semantic aspects of words (Protopapas et al, 2013). In addition, comprehension depends partly on the degree of ambiguity of pronunciation of a word; for example, the word 'tear' (he shed a tear; there was a tear in his shirt), or from differences in accents or dialect (Davis, 2012). Following further research of their own, the authors of the Simple View of Reading concluded that the original model should be modified to account for the influence of vocabulary on decoding skills. Their recommendation was that the teaching of reading should focus on improving oral vocabulary knowledge as well as alphabetic decoding skills (Tunmer & Chapman, 2012). Since vocabulary knowledge contributes to the learning of decoding skills and real word recognition, as well as comprehension, then efforts to increase the vocabulary of young readers are essential.

Within the ωeebee Reading Programme, the Big Book addresses the early learning goals of the EYFS that states that the children should have the opportunity: to experience a rich language environment; to speak and listen in a range of situations; and should link sounds to letters and use a range of reading materials (DfE, 2012a). The suggested questions in the teacher's Manual to accompany the Big Book, were planned to assist the building of background and world knowledge; to support both inference generation and the construction of situation models, for comprehension.

3.3.2 Phonics and Analogy

Three of the activities focused on either analytic phonics, onset-rime, or word families. These were intended to give explicit instruction in both simple and complex grapheme-to-phoneme correspondences and promote the use of analogy in word recognition.

3.3.2.1 Initial and final letter phonics activity

Each child was given a sheet with a picture of a house or similar illustration with a ωeebee (see Figure 3.2 below). There was a window in the house to be cut out and a number of letters to make up part of a word (or rime) for example _ed. A thin strip, with greyed-out letters, was cut out and the letters written over by the child. These could be drawn through the window to create a number of new words

(onset-rime), for example: red, bed, ted, fed etc. Most, but not all, of the words appeared in the books. The changing letter, or letters, was either in the initial or final position; later in the programme they were located in both positions to allow for quite complex variations. All the words so formed made sense, there were no 'nonsense' words possible.

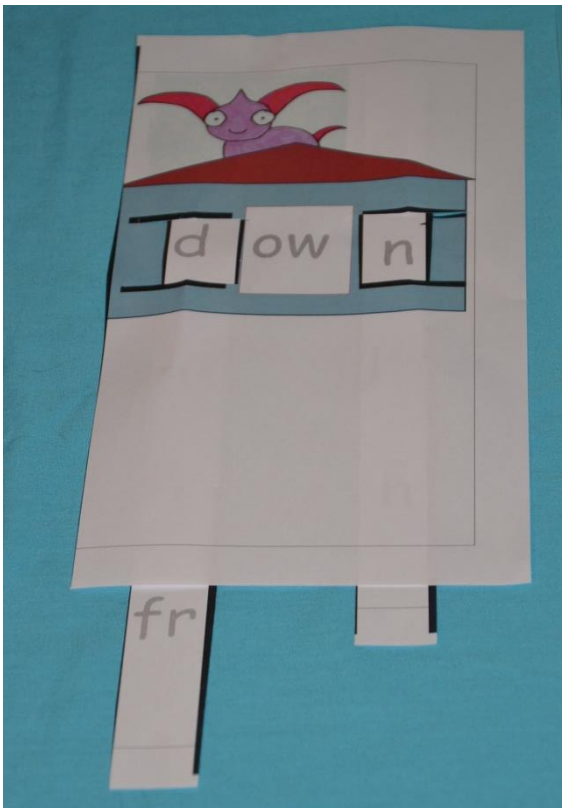


Figure 3.2. Example of a phonics draw-through activity.

There is evidence that much of the information conveyed by written text is redundant for the skilled reader. The skilled reader needs neither to see all the letters in a word, or all the words in a sentence, to be able to read. Furthermore, there is evidence that skilled readers can predict words on the basis of the first three letters (Rayner, Johnson & Perea, 2007). In addition, readers are able to identify words when internal letters have been transposed. In other words, letter recognition is not dependent on their absolute position within the word; however, initial and final letters play a crucial role in visual word recognition (Rayner et al, 2007). Hence the value of giving children activities that draw attention to initial and final letters of words.

This observation is supported by eye-tracking models, which demonstrate that eye movements 'bounce' from initial to final letters, and also any salient

features within a word, but also skip forward several words, as well as backwards, as if searching for context (Blais et al, 2009; Engbert, Nuthmann, Richter & Kliegl, 2005; Pitchford, Ledgeway & Masterson, 2008; Rayner et al, 2011). The results of a study of eye movements between saccades (the jump between fixations) (Blais et al, 2009), led the authors to conclude that, for skilled readers, in addition to only needing three letters to identify a word, regardless of word length, that letters appear to be processed simultaneously (not sequentially). A study by Pitchford et al (2008), suggested that this process, used for identifying letter positions, is adaptive to each particular language and culture. For children learning to read English, this means learning to track from left to right, and fixating on the letters furthest left of any new word in the first instance.

The evidence in the literature, regarding eye movements (Blais et al, 2009; Engbert et al, 2005; Rayner et al, 2011; Wade & Tatler, 2005), demonstrates that eyes do not move smoothly in one direction. For children with dyslexia, these findings may be particularly useful, since they have been found to have a smaller visual attention span, limiting the number of letters that can be processed at any one time (Bucci, Nassibi, Gerard, Bui-Quoc, & Seassau, 2012). Eye movements have been found to show sensitivity to lexical and linguistic variables (Reichle, Reineberg & Schooler, 2010); short words are more likely to be skipped than long ones, and three-letter words in particular have been found to be skipped 67% of the time (Rayner et al, 2011). High frequency words, and words that can be predicted from context, are likely to be skipped regardless of word length (White, 2008). The initial letters of words appear to be processed before they are in full view, and their processing has been found to be not significantly affected by word length (Inhoff, Radach, Eiter & Skelly, 2003). This highlights the importance of encouraging the skill of speedy recognition of initial letters, which is a feature of skilled reading.

The use of an onset-rime teaching approach is clearly supported by the research in eye movements. Children are able to acquire a repertoire of words, albeit small, such as their own name, 'mummy', 'daddy' etc. and from this repertoire they are then able to recognise onset and rime and draw analogies (Dombey, 1999). Dombey (1999) points to significant research evidence demonstrating that children find units of onset and rime easier to access than individual phonemes, and recommends activities, such as the draw-through phonics worksheet, described above, which manipulate onsets for simple words.

The phonological-linkage hypothesis (Hatcher et al, 1994) suggests that phonological training, in combination with the teaching of letter names and sounds is more effective than either one in isolation; children need both phoneme awareness and letter identification knowledge to access the alphabetic principle. The draw-through phonics activity described above has been designed to integrate phonological awareness training with the teaching of reading letters. Manipulating the first and last letters helps children with understanding the encoding of sound into symbol. There is also evidence that this approach can have a positive impact particularly on children with poor decoding skills (Weiser, 2013).

3.3.2.2 Happy word families

This game was for a group of four players. Each child was given a word card, selected at random, with a set of four words from a family (for example with the same onset or ending/rime) highlighted in a colour. The rest of the cards were shuffled and shared out between the players, so that each child had four unseen cards besides their seen word-family card. Each of the unseen cards had one of sixteen words, with a corresponding illustration. The object of the game was to collect all the words from their family. Teachers were asked to help the children to request the cards from the other children. The winner was the first to complete a family.

Figure 3.3. Example of Happy Words Families Game

As far as possible the word-families were selected from words from the books, but did include some additional words. These sets of words were also printed in the 'Big Book' and were part of the plenary whole-group session. One family of words was studied at each session (for example words with ee such as: tree, green, peep, sleep). Attention was drawn to the family characteristic, which was highlighted in a colour (see Figure 3.3 above).

According to psycholinguistic grain size theory (Ziegler & Goswami, 2005), learning to read is fundamentally a process of matching distinctive visual symbols to units of sound, (grain size refers to the size of unit of text or speech that a learner attends to). Children appear to master skills in a hierarchical order: word-level, followed by syllable, followed by onset-rime, followed by phoneme level. The progression described here has similarities with both the stage and phase theories discussed in Chapter One. In English, smaller grain sizes, such as individual phonemes, tend to be less consistent (for example the vowels) than larger groupings of letters, resulting in children showing a preference for using larger, more consistent units, particularly for multi-syllabic words (Ziegler & Goswami, 2005). This theory runs counter to the idea of beginning with phonemes such as in synthetic phonics teaching.

Analogies based on rimes appear to be made more frequently by children at the beginning stages of reading (Allen, 1998; Goswami, 1999), although it has been suggested that phoneme segmentation or sensitivity, rather than use of rime, is more predictive of early reading success (Muter, Hulme, Snowling and Taylor, 1997; Muter et al, 2004; Nation and Hulme, 1997). However, Goswami (1999) points out that for words that have more rime neighbours (word families), children are able to use the strategy of analogy more effectively. A statistical analysis of English spelling (Treiman et al, 1995) showed that consistency was greatest for onset, rime and final consonant, and therefore easier to learn. Goswami (1999) also points out that a phoneme is easier to distinguish, in speech, as onset or rime. As a child's reading vocabulary increases, the ability to use analogy, which is related to the frequency of encountering a particular rime, also increases (Moustafa, 1995). Skilled readers come to realise that position within a word, and surrounding letters, affects the correspondence between orthography and phonology, and this knowledge allows the reader better to match the letter string pronunciations within a known word.

Goswami (2005) concludes, from the evidence of several of her studies, that children use analogy as a reading strategy from the earliest stages of beginning reading, and that the use of analogy will develop faster if explicitly taught.

Part of the development of rapid word recognition involves the construction of what Tunmer and Chapman (2012) refer to as sub lexical, visuo-phonological connections between letter patterns and sounds. As these representations become more familiar, children find it increasingly easy to deduce sound-spelling relationships. Tunmer and Hoover (1993) refer to these letter patterns as phonograms, and suggest that children can learn the syllabic units easily because the vowel sounds within the phonograms are generally stable (consistent).

The self-teaching hypothesis (Share, 1995) postulates that successful orthographic decoding acts as a self-teaching mechanism for encoding specific orthographic representations. Ricketts et al, (2011) demonstrated that children can make use of existing orthographic knowledge, to contribute to further learning, and that the two are inextricably linked. In a study of second grade children, Cunningham et al, (2002) found evidence of orthographic learning under conditions that simulated the kind of self-teaching that might occur in every day reading contexts.

In a case mixing study (mixing upper and lower-case letters), Havelka and Frankish (2010) demonstrated that when multi-letter vowel graphemes (such as 'ea' in steak) were kept in the same case (while the other letters were mixed up) they were read more easily than when mixed (for example 'eA'). They also found that words with multi-letter consonant graphemes were processed faster, even when using mixed-case letters. Multi-letter vowel graphemes were slower to process than consonants, which were felt, by the authors (Hevelka & Frankish, 2010), to be related to the level of sound-to-grapheme consistency; vowels being less consistent.

Because of the inconsistencies in English, there is a preference, among English children, to use larger units of sound than single phonemes (Goswami, 2005). As a result, English children are better able to read non-words that sound like real words, than non-words that have no recognisable patterns, demonstrating the influence of whole-word phonology in reading efficiency (Goswami, 2005). This supports the use of teaching by analogy.

The use of either word families or onset-rime, as described earlier, are both methods that can be described under the umbrella term 'analytic phonics', where children analyse the construction of a word, and note the sound groupings as chunks of information. Teaching children to look at syllables is a similar approach and has demonstrated effectiveness (Diliberto, Beattie, Flowers & Alzozzine, 2008). An analysis of an English database resulted in a calculation that there are 461 possible grapheme-phoneme correspondences (this includes multiple letters representing single phonemes) (Gontijo, Gontijo & Shillcock, 2003). However, many of these occur only infrequently. A further analysis of these letter groupings, found that the most frequently occurring grapheme-phoneme correspondences (GPCs) in texts numbered only 64 (Vousden, Ellefson, Solity & Chater, 2011). Vousden et al (2011) suggest that a 'Simplicity Principle' can be applied to reading, by teaching the most frequently occurring GPCs, beginning with the easiest and most common in children's text. A small study has already demonstrated an effective application of this approach, improving the motivation and reading progress of struggling readers by teaching them common GPCs (Chen & Savage, 2014). The word families (letter groups) used in the games for the weebee Reading Programme begin with the most common GPCs used in the associated books and continue to reflect the vocabulary in the books as they progress to more complex GPCs.

3.3.2.3 Fishing

This game was designed for groups of up to four players. Using magnetic fishing rods, the children 'caught' small fish with letters on. The teacher was to give each child a card with a word on, and the children had to find the fish with the corresponding initial letter. This game focused on speed; the winner was the child who caught the most fish. The intention was that the children would have opportunities to practise looking rapidly for the initial letters of words, in order to deliberately stimulate this technique of skilled reading. Once this game had been played, there was an alternative set of fish with whole words on. The game was played in the same way, but the children were looking for the whole words, and so could use other cues in addition, such as final letters, word length, and salient features (see Figure 3.4 below).

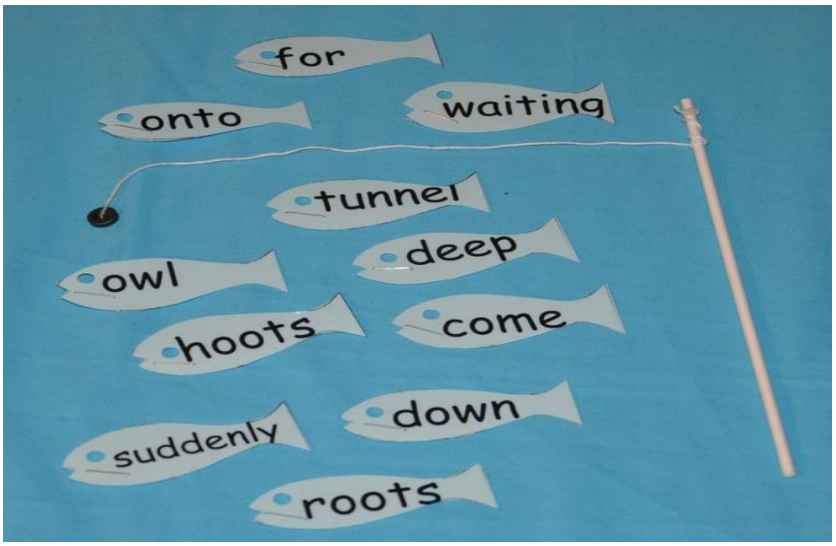


Figure 3.4. Fishing game

There has been considerable debate about the role of a word's shape in its visual recognition (Beech & Mayall, 2005; Lavidor, 2011). Evidence from a study by Beech and Mayall (2005) suggests that the word-envelope hypothesis (the outline shape of a word) is flawed, but that salient external features are nevertheless important. They found that children with a reading age of 8 named words with ascenders and descenders (letter components above the middle third and those below) more accurately than words made up of neutral letters. This suggests that there are more salient features in the outer area of words, which are important for word recognition. Wilkins, Cleave, Grayson & Wilson (2009) argue that letter shape needs to be studied in the context of a word, and that the internal word shape (the visual effect of ascenders and descenders) may be important. Neutral letters are mainly vowels which seem to carry redundant information with regard to whole word shape recognition. The fishing game was designed to draw children's attention to the initial and final letters, as well as salient features in words, in order to practise the behaviours observed in skilled readers: the rapid recognition of and attention to these features.

3.3.3 Games to Establish a Bank of Sight Words

There were eight games included in this programme intended to build and support a bank of sight words to aid fluency and comprehension and reflect some elements of a whole-word approach. They were designed to put into practice the concept of tangential learning, as well as the principles of learning through play and group activities.

A study by Lavidor (2011), suggested that word shape is more likely to affect readers with dyslexia than normally developing readers; for children with dyslexia, there may be an increase in reliance on orthographic features of words, as a consequence of possible phonological deficits. Children with poor phonological skills often rely more heavily on visual memory (Johnston and Morrison, 2007; McGeown et al, 2012). For these children, a teaching strategy that uses more visual cues is likely to be beneficial. Given that the research evidence suggests that children with dyslexia show a preference for using visual cues, Tormanen and Takala (2009) used auditory-visual matching games, which improved the reading-related phonological skills, with dyslexic children. For non-dyslexic children, who have a preference for using visual cues, auditory-visual matching games may be similarly effective. The stage model, proposed by Frith (2001), suggests that children begin with a logographic stage, during which they recognise whole words from looking at shape and even colour.

3.3.3.1 Jig words

Jig words were a selection of single words printed large enough for one word to fit on a single A4-size sheet. Each sheet was divided up into either 4 or 6 jig-saw style pieces cutting across each word (see Figure 3.5 below). The children worked in pairs or groups of three to match up the pieces to find the words. During training, the teachers were asked to help the children complete the puzzles by looking at the shapes of the pieces, rather than trying to reconstruct the word. On completion, the teacher was asked to verbalise the words, but not expect the children to do so. This was because the focus was to be on the task of putting the matching shapes together, not on trying to read the word. The words were printed in black onto an illustrated background, to help with finding the correct pieces. Words were selected from the assigned book. This activity was intended to help the children learn the early vocabulary used in the books, which would be repeated throughout the series. The illustrations allowed the children to put the puzzles together, but the finished image included the word that represented the image. The intention of this activity was to provide an opportunity for tangential learning to occur. At this early stage, the focus was on extending vocabulary, in terms of developing a sight vocabulary to aid fluency in reading the early books.

Figure 3.5 Example of a Jig Word

Fluency is achieved when readers are able to read familiar words rapidly from memory; children are able to read sight words faster than decoding simple consonant-vowel-consonant non-words (Ehri, 2005). More importantly, children have been found to read familiar sight words at the same speed as single letters, suggesting that they read words as whole units. It is more efficient to be able to read words automatically from memory (Allen, 1998). According to Ehri (2005), the learning of sight vocabulary is a connection-forming process; connections are made between spelling, pronunciation and meaning.

3.3.3.2 Bingo

A set of four game cards was provided (with each card illustrated with a different book character). Each card had a selection of twelve out of a possible twenty words (there were twenty new words for each book). The teacher was to keep the words in an opaque envelope and select them at random. As the words were read out they were also to be shown to each child. The teachers were asked to encourage the children to assist each other in looking for the words. After all the cards had been checked, the word was laid clearly visible on the table and then the game proceeded. The winner was the first to cover all the words on their card. The words were printed in black onto four different coloured backgrounds that corresponded with their book character's colour (see Figure 3.6 below).

Figure 3.6. Example of Bingo Cards

As an alternative to the stage model (Frith, 1985) or the phase model (Ehri, 2005) of reading, there may be a continuous on-going process with no fixed end point. The connectionist model describes a process using pattern recognition, guided by parameters and constraints, rather than by a rule-based approach (Foorman, 1994). In this model, learned knowledge exists in the connections between neurons that are made during learning (Rumelhart, 1989). This model represents artificial neural networks and, according to Foorman (1994), explains how the learner can be trained to recognize spelling patterns without the use of phonic rules, or even contextual knowledge. Instead of storing information at the word level, this model suggests what is termed 'parallel distributed processing'; there are parallel representations of orthography, phonology and meaning. This model describes a single processing mechanism, using distributed representations and weighted connections. Multiple exposures to consistent spelling patterns 'push' these weights toward a particular pronunciation. The effects of consistency and frequency result in the adjustment of the weights according to experience, not from the learning of rules.

Connectionist models, supported by computer simulations, provide clues as to the nature of mechanisms that might be used for learning to read (Ashby & Rayner, 2012; Compton et al, 2014). These models mimic human reading

acquisition by using back-propagation of errors as the basic learning mechanism, which adjusts the weighted probabilistic relations between orthography and phonology. During guided reading, children receive feedback from the teacher, from which they can construct a "matrix of correlations among letter patterns" (Foorman, 1994, p43).

In a small study with dyslexic children, evidence was found of altered cortical maps and brain plasticity following a reading intervention, and subsequent improvements in reading ability (Papanicolaou et al, 2001). This evidence is very much in keeping with a connectionist theory of learning as a continuous process. The bingo game provided an opportunity for multiple exposures to written words with repeating patterns, at the same time as hearing the correct pronunciation by the teacher verbally modelling the words.

3.3.3.3 Memory game

In the Memory game a set of nine pairs of words were shuffled and placed upside down in a grid shape. The game was played in pairs. The children took turns to turn over two cards. The teacher was asked to read the words and not expect the children to do so (but not to stop them either). The teacher asked the children if the two words were the same. If they were unsure the teacher was asked to draw their attention to the first letter (none of the words began with the same letter). If they had turned over a matching pair, they could keep the cards and have another go. If they did not match, the cards were turned back over and the children told to try to remember where the words were so that they could try to find them again.

The focus was on matching the words, with attention drawn to the initial letters. The winner was the child with the most pairs at the end. The words were printed in different colours (matching pairs having the same colour so this could be used as an additional cue) onto a pale-coloured background. A few selected words had illustrations on the cards as well as the words for variety, and to assist some of the children (particularly the younger children in the early stages of the programme) (see Figure 3.7 below).

Figure 3.7. Example of Memory Game

There has been a growing interest in the concept of statistical learning. Research into artificial intelligence has led to a greater understanding of how statistical learning might operate. These theories have been applied, particularly in the area of language learning, but there are clear parallels with reading development. They suggest that infants learn visual features based on statistical correlations (recognisable visual patterns). From these correlataions they develop a representation from which they can learn by association (Fiser & Aslin, 2002). Infants learn from frequency and predictability and use probability to develop internal representations of what they observe. Features learned in one context can later be recognised in an alternative context. This is a mechanism for learning more complex features. When attending to spoken language, infants show differential attention to familiar and unfamiliar syllable combinations (Gomez & Gerken, 2000). They are also sensitive to changes in tone sequences, and are able to identify aspects of speech based on predictive syllable relationships. In addition, infants appear to recognise form before the forms have obtained meaning (Gomez & Gerken, 2000). This ability to recognise form supports the use of a whole-word approach to teaching, whereby knowledge of meaning is not a prerequisite of recognition of form. The memory game presents words out of context, but requires attention to the form of the word, the sound of which is also modelled at the same time by the teacher.

Probabilistic learning is discussed by Compton et al (2014) with reference to the frequencies and constraints between the orthography and phonology in English. They recommend the construction of a carefully planned body of words that will train connections between units of various sizes, promoting a parallel growth of decoding skills and word identification skills. The words selected for the Memory game varied in size, but also included words with repeating letter patterns to promote learning through analogy.

3.3.3.4 Dominoes

Cards were provided with two different words on each card (see Figure 3.8 below). The cards were shuffled and placed on the table. The teacher was asked to take a first card (or several if a large group) and lay it on the table to start the game. The game was played in groups; each child taking a turn to take a card and see if they could place it down next to a matching word. If they could not place it down they kept the card. The game finished when none of the players could place a card, or one of the players had used all their cards. The winner was the one with the fewest cards at the end. The words were printed in different colours (matching words having the same colours for additional cues) on white paper, but with a dark-coloured edging/backing. The teacher was asked to verbalise the words needed by each player as the game progressed, but not expect the children to do so, and to maintain a fast pace. The children were encouraged to help each other find the matching words.

Figure 3.8. Example of Dominoes

Observations by Gopnik, Glynmour, Sobel, Banks, Schulz & Kushir (2004) have led to the notion that the connections that children's brains develop, from observing correlations and making predictions, can be represented in the form of Bayes nets (probabilistic graphical model). Figure 3.9 below represents a Bayes net model applied to reading comprehension, showing possible connections between oral vocabulary, and word recognition and oral vocabulary and comprehension.

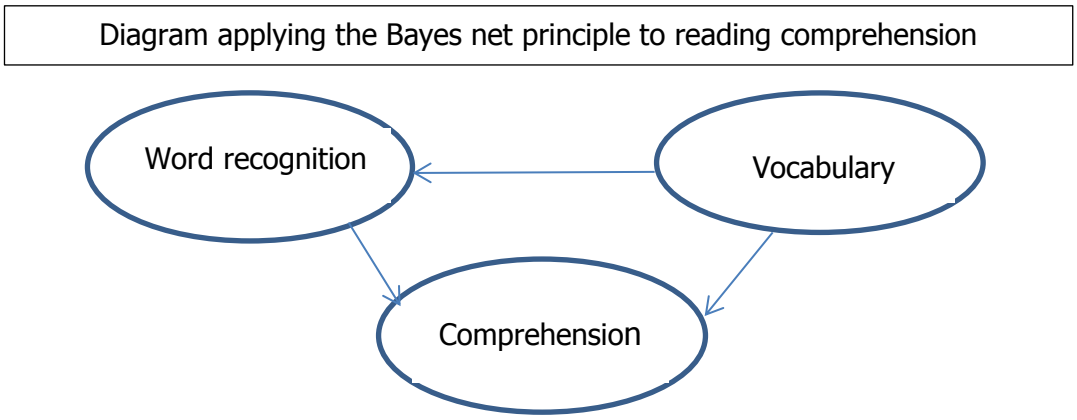


Figure 3.9. Diagram of a Bayes net model of reading comprehension

Children can combine prior knowledge and new observation to discover new relations on the basis of small samples of data. According to Gopnik et al (2004), children are able to use prior learning to underpin future learning. Mosely (2004), has observed that children are aware of relationships and patterns found between letters and sounds, without these being explicitly taught. The research findings of Pitchford et al (2008) led them to conclude that models of reading should incorporate a process of statistical learning, including letter positions, based on their position frequency.

During language acquisition, there is evidence of a critical moment, described as neural commitment (Kuhl, 2004), at which point, the learner's neural network commits to patterns that reflect the language of their home environment. Subsequent learning begins to conform to those patterns, and builds on them but does not change them. In addition, patterns that do not conform are rejected. Words that can easily be mapped to meaning, through an object presented visually, are learned quickly (Romberg & Saffran, 2010). Figure 3.10 below represents an application of the theoretical model (Gopnik et al, 2004) to how children may apply rules and use repeating patterns for letter recognition. For example, learning the

sound of the letter 'm' from familiar words that start with that letter, or the sound of the letter 'y' from familiar words that end with that letter and from words that fall into both groups such as 'mummy'.

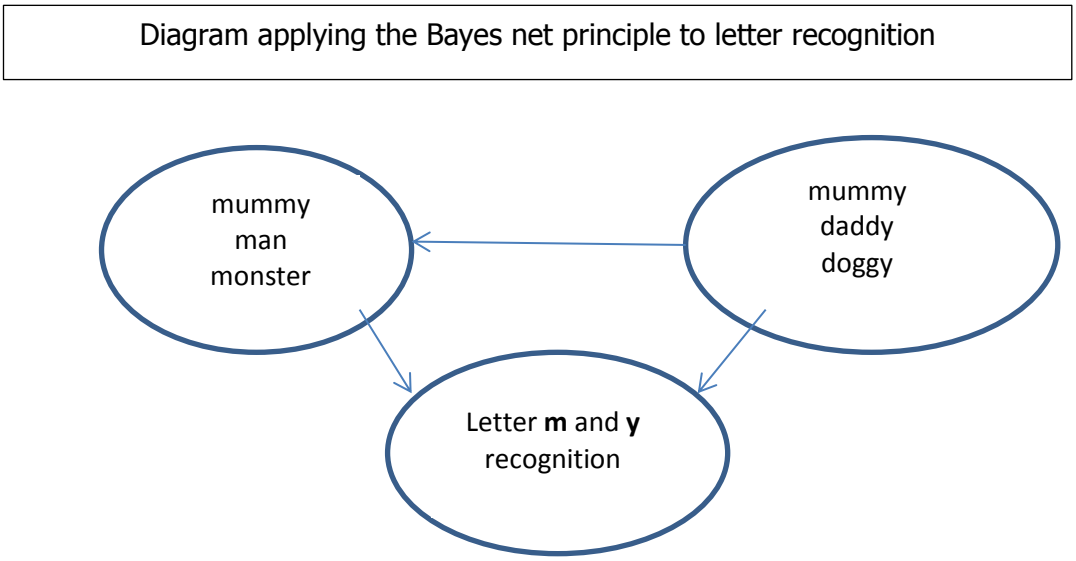


Figure 3.10. Diagram of a Bayes net model of letter recognition

By using mathematical models, such as Bayes nets, it is possible to picture how the human brain maps and stores information (Gopnik et al, 2004). Children learning to speak, and beginning to apply rules of grammar, such as appropriate endings, can be observed to self-correct (Lapin & Shieber, 2007). These same processes apply in reading. Using repeating patterns, the child is able to map letter groups that have consistent sound-to-spelling correspondences. This is similar to the mechanism used by children to learn from the pronunciation corrections that they receive during guided reading. The mathematical models, described in Lapin and Shieber's paper (2007), of machine learning, refer to the need to restrict training samples to a limited number for there to be a convergence of learning. The dominoes game had a specific and limited selection of words intended to foster this convergence of learning.

3.3.3.5 Snap

The twenty new words to be learned for each book were divided into two sets, so that each pair of children played with a different set of ten words. The words were printed in black onto pale coloured backgrounds (see Figure 3.11 below). The focus was on matching at speed. The teacher was asked to supervise fair play, and to

verbalise the words as the game progressed. The winner was the child with the most cards at the end. This game was used to assist in possible statistical learning, as suggested in connectionist theory (see Section 3.3.3.2).

Figure 3.11. Example of Snap Cards

The statistical learning described by Treiman and Kessler (2006) includes recognising probabilistic patterns. When this is constructed as a computer model, it can generate plausible spellings for items not previously encountered. A study of infant behaviour demonstrated that they are able to detect, and attend to, the statistical and distributional properties of the language they are exposed to (Stokes, Kern & Santos, 2012). This is referred to as 'constrained statistical learning', and influences the development of spoken vocabulary (Stokes et al, 2012). As children develop, they need to broaden their learning strategies, in order to be able to include words of lower statistical probability. Words with high neighbourhood density (similar forms such as 'hat' and 'hot') were found to have a far greater relationship to sight vocabulary development than high frequency words (Stokes et al, 2012). These observations were related to spoken language, but it is likely that similar processes would apply to reading development. There is evidence that automaticity has a role in mediating decoding skills (Silverman et al, 2013). Games such as snap encourage speedy recognition of words, which, it is anticipated, would promote automaticity in word recognition. Words were selected to include those with high neighbourhood density (such as 'wall' and 'will').

There is some evidence that if children can infer the meaning of words from context or illustrations, that they will not necessarily attend to all the orthographic detail of a word, which can result in inaccurate guessing. When words are presented completely out of context, children are forced to use the fine details of the print (Johnston, 2000). For example, during a rapid game of snap, the children

can only match words by looking at these orthographic details, particularly where words vary by only one letter (for example hoots and roots).

3.3.3.6 Pento

This is a board game for up to four players; it took its name from the five cards that had to be collected. Each child had to move around the board and collect a set of five words that corresponded to the storybook character that they had chosen as their play figure. In addition, there were places on the board where they could land, which required them to pick up a card telling them to move, either forwards or backwards, by a particular number of moves (see Figure 3.12 below). This involved the children looking repeatedly at a small selection of words to check if they had landed on a word corresponding to their chosen character's set.

Figure 3.12. Example of Pento Game

A whole-word approach to teaching reading is based on the assumption that knowledge of syntax, semantics and phonology allows the reader to make predictions through contextual cues (Ashby & Rayner, 2012). The use of flash cards may no longer be in vogue, but many teachers maintain that a bank of sight words (words recognised at sight, in one go, without the necessity of 'sounding out') is still essential for children to be able to read with fluency and understanding. A bank of sight words, that are a product of automatic word reading, is considered to be a necessary contribution to comprehension (Mesmer, 2009). In many respects, the

weebee Reading Programme uses the same principles as flash cards. However, the presentation of words is through the playing of games, rather than the traditional drill of holding up words to be chanted by the whole class.

3.3.3.7 Grog's Journey

This game is for up to eight players. It is a simple race to the finish, depending on the luck of the dice. Each player travels along the journey via stepping stones with words on. The teacher was to model the words and encourage the children to help each other find their next stepping stone (see Figure 3.13 below). This activity was intended to extend the children's sight vocabulary to aid fluency.

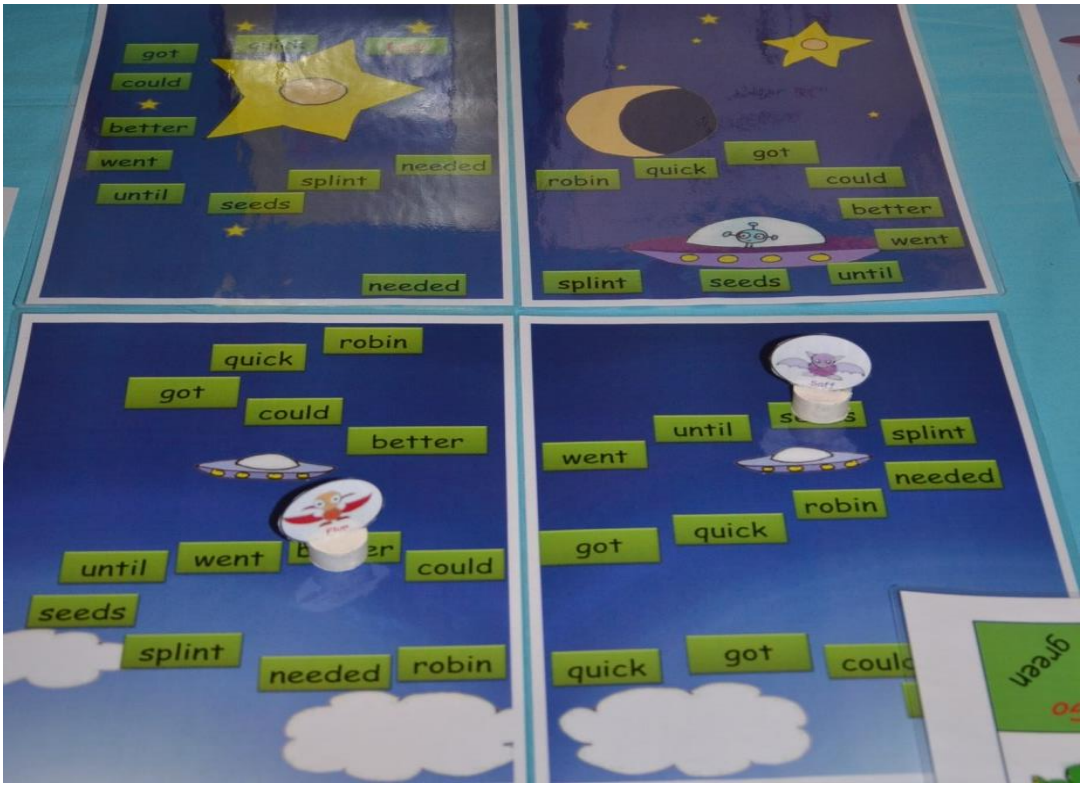


Figure 3.13. Example of Grog's Journey

There is some evidence that there are causal links between reading and writing; that children learn to recognise orthographic patterns and rules from reading and then apply them to spelling. In a study to test this hypothesis, Davis and Bryant (2006) found that spelling scores over a period of two years began to converge with reading levels, suggesting, according to the authors, that children learn to use orthographic rules in reading first, and later apply this knowledge to spelling. This game, used during the second series of books, was designed to

extend sight vocabulary to support the learning of orthographic patterns, using repetition of words that could be applied to spellings.

3.3.3.8 Snakes and Ladders

This is a game for up to four players. It follows the traditional rules of snakes and ladders; the only difference being that children land on words as they move up the board. The teacher was asked verbally to model the words as the game progressed. This game was used to assist the continued building of a sight vocabulary (see Figure 3.14 below).

Figure 3.14. Example of Snakes and Ladders

Evidence suggests that children with learning difficulties are less likely to learn using conventional methods (Lacey, Layton, Miller, Goldbart & Lawson, 2007; Wedell, 2014). Some children appear to learn best from a whole-word approach at first, with phonics introduced at a later stage to supplement the learning (Helenius, Uutela & Hari, 1999; Lacey et al, 2007; Poulsen, 2008) and children with learning difficulties rely more heavily on orthography (Snowling, Gallagher & Frith, 2003). For struggling readers, one benefit from the whole-word approach used in games, such as snakes and ladders, is that it does not require words to be sounded out.

3.3.4 Comprehension

Only one of the activities was designed specifically to address skills of comprehension. It was introduced at the beginning of the second series of books, after the children had been introduced to all eight of the story book characters and

had already built up a significant bank of sight words. By the end of the first series, the children would also have learned a number of songs and accumulated oral vocabulary and background knowledge from the Big Book discussions.

3.3.4.1 Sentence Matching

This activity was designed for individual children; although they were encouraged to work together and discuss their options. Each child had a pack of eight illustrated cards, which were replicas of pages from a book from the first series. For example, when using materials for Book 9, the activity would relate to Book 1; for Book 10 it would relate to Book 2 etc. Each child also had a set of cards with sentences which they had to match to the pictures. These were not the same as they appeared in the original books, but did use words from the first series (see Figure 3.15 below). In order correctly to match the sentences to the pictures, children needed to understand the text, not simply to have remembered the text from the earlier book.

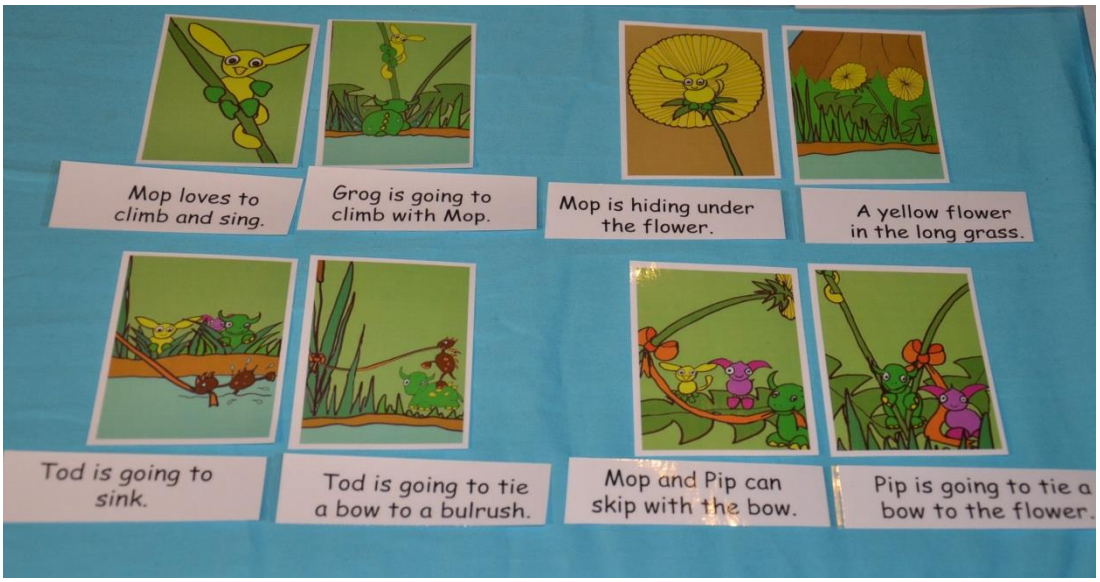


Figure 3.15. Example of Sentence Matching

A number of studies have suggested that because children can perform differentially on two dimensions of skills: decoding and language comprehension, teachers may have to use different strategies for each dimension to support reading development (Kendeou, Savage & van den Broek, 2009; Snowling & Hulme, 2012). Inference generation and poor comprehension appear to be linked; it is thought that poor inference generation leads to poor comprehension, and that poor comprehenders generate fewer inferences (Bowyer-Crane & Snowling, 2005; Clarke, 2009). A study of children with poor comprehension, suggested that the difference

between skilled and less-skilled comprehenders lay in the strategies they used to draw on general knowledge to interpret text, as well as having limited vocabulary (Bowyer-Crane & Snowling, 2005; Elliott & Grigorenko, 2014).

The sentence matching activity was designed to promote inference generation by giving the children ambiguous sentences, from which they had to infer information from the pictures, in order accurately to match the pictures and sentences. In addition, the activity gave teachers an indication of children's comprehension of the text they were being given, as opposed to merely monitoring simple word recognition.

3.4 The Reading Books

All the activities described above preceded the reading of the books, so that the children should have already become familiar with all the new words before reading each respective book. To compare the effect of using different vocabulary on word recognition and comprehension, it was important to control as many variables as possible. Therefore, two parallel sets of activites and books were developed. Each set of books contained the same illustrations and storyline, the same number of new words introduced in each book, and the same number of words per book. (All the learning materials had similarly to be the same, except for the selected vocabulary.)

3.4.1 Illustrations

Throughout the books, full-colour full-page illustrations were used opposite every page of text. The design of these involved simple line drawings with block colour, but included perspective and facial expression to aid inference generation. There was continuity in the backdrop for each of the stories in the series, so that a storybook world was created, and used in the learning materials, as well as the books.

Historically, the earliest form of writing as narrative was in picture form found in cave paintings. Looking at a picture book is part of the process of learning to read at the earliest stages, through which children begin to understand story structure. They are able to make predictions about what will be on the next page. They encounter tenses through illustrations that can portray the past, present and an anticipation of the future. Context comes alive through illustration, and children

can widen their life experience through pictures; for example children who have never seen the sea can 'experience' it through illustration. In addition, by encouraging and allowing detailed discussion of illustrations, vocabulary can be extended; new and previously unknown words can be clearly demonstrated through pictures, for example 'flying saucer' (see Figure 3.16 below).

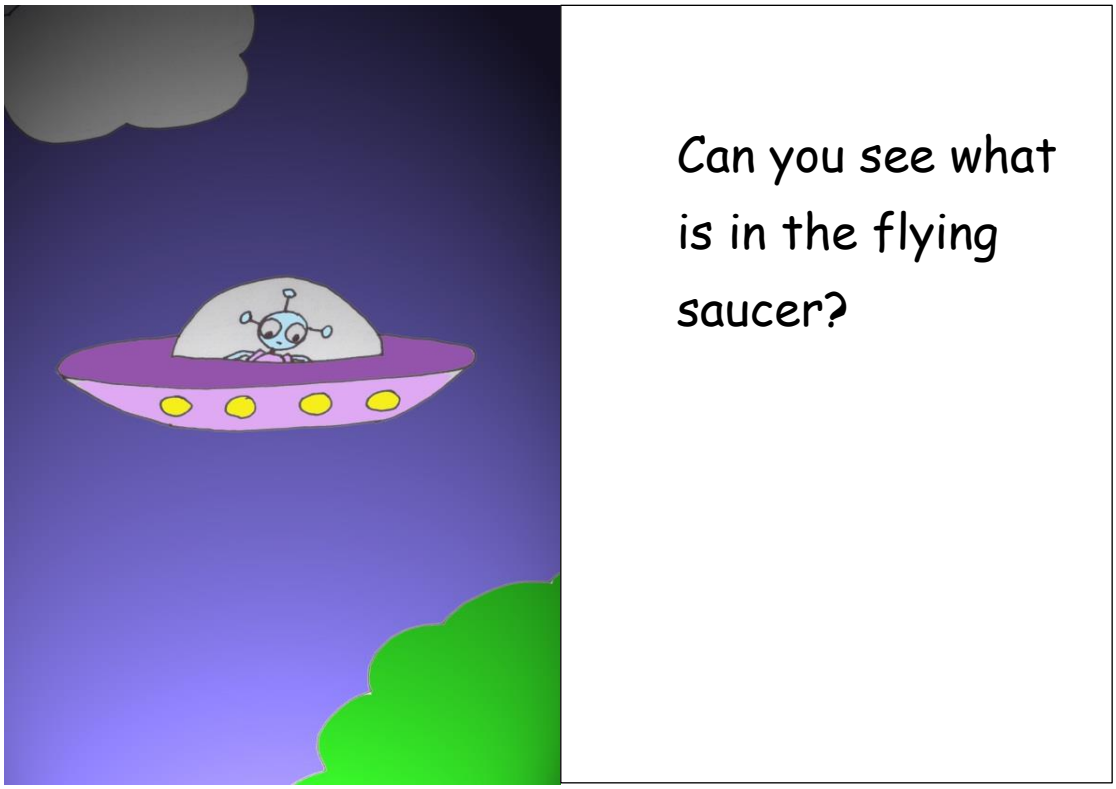


Figure 3.16. Example page with picture demonstrating words in text

Research, from a number of studies, suggests that pictures have a powerful influence on emergent readers, helping them make sense of the text, and evoking an emotional response (Lacey et al, 2007). Illustrations are a significant means for providing access to narratives and ideas that may not be available through text alone, especially to younger children. Higher order reading skills have been shown to develop within students as connections are made between word and image (Arizpe & Styles, 2003). Results, from a study by Walsh (2003), demonstrated children using illustrations to look back at a book, to gain a better understanding, to assist in retelling a story. This retelling included information that was not contained in the text. In other words, the illustrations were embellishing the text. The Big Book discussions also contributed to the construction of micro-world knowledge to supplement the information from the text. With text written for beginning readers, there are, inevitably, significant gaps in information, in terms of textual information,

while the number of recognisable words is still very few. In an older study, the authors reported students remembering significantly more detail from a story that had illustrations compared to no illustrations, even after a five day interval (Haring & Fry, 1979). In addition, illustrations allow the author to introduce emotion and humour without having to use complex text, or even any text (Arizpe & Styles, 2003).

The Dual Coding Theory suggests that there are two systems, by which the reader makes both a verbal representation, and also a non-verbal representation (Sadoski & Paivio, 2001). The authors suggest that information encoded via both routes (verbal and visual) is more readily recalled since it is encoded in different areas of the brain. They also note that beginning readers learn a sight vocabulary much faster when those words are accompanied by a referent picture.

The chosen environment for the story characters was the roots of an old oak tree. All the characters lived either down in the roots of the tree or in its immediate neighbourhood. By using imaginary creatures, roughly the size of insects, it was possible to look at the everyday world from a different perspective. This also allowed for a natural real-life back drop for the fictional characters, in an environment free from technology, or any cultural influence. The fictional characters were situated in reality, making the stories accessible to young children, rather than using abstract concepts. There were no obvious gender differences between the characters although 4 were referred to as female, and 4 as male.

In a study of young children's reading preferences, the majority of children chose books which featured animals and which were fully illustrated (Mohr, 2006). In this study, the topic of animals was a preference for both boys and girls, although girls were found to have wider preferences than boys. In addition, the children preferred books about the natural world than books which would reflect their own lives. All the imaginary creatures featured in the books in the *weebee* Reading Programme were animals except one (which was an alien). Several of them had dinosaur-type features which was a deliberate attempt to appeal to boys. The illustrations used in the Big Book were designed to encourage discussion about the natural world and the habitat of insects and small animals and were directly related to features or events in the books themselves.

The specification given to the concept artist who was asked to design the original line drawings for the characters for the books was that they: should appeal to both boys and girls; should have no obvious gender; should be simple line drawings that could be copied by five-year olds; should have no cultural context and be completely original. The intention was to create characters which would appeal to all children regardless of gender, race, ethnicity etc., in order to avoid those issues. The generic name given to the story book creatures, the '*weebees*', was chosen as a response to a piece of research showing that some children have difficulty discriminating between the sounds 'w' and 'b' when followed by a vowel (Goswami et al, 2011). By deliberately using a word which repeatedly exposed children to these sounds it was hoped that this particular difficulty may be ameliorated in some way.

3.4.2 The Text

The majority of reading schemes (for example the Oxford Reading Tree) begin with picture books and then use single words per page. Gradually the number of words per page is increased. Children read with an adult in 'guided reading' sessions and re-read the books to reinforce learning. There are not usually any additional materials provided directly associated with the books. Because the focus of this research was on the vocabulary being used, the design and use of the books developed for the research was radically different. Before the children even saw the books, they were given activities during which they were exposed to the vocabulary in advance, albeit out of context and in play-learning mode. Thus, right from the first book the programme used a complete storyline. The stories described aspects of the real world and also situated the building of friendships. These two elements reflected one of the four specific areas in the EYFS (DfE, 2012a) 'understanding the world' (talk about facets of own immediate environment; make observations of animals and plants) and one of the three prime areas 'personal, social and emotional development'.

Each book progressively introduced twenty new words, and built on previous words learned. This was based on the successful experience of a remedial reading scheme which introduced twenty new words per book in a similar way (Harris, 1978). The reading of books was on a one-to-one basis with the teacher. There were detailed instructions in the teacher's Manual about the language to be used to

support the reading; when and how to assist if the child did not recognise a word and how to record an unknown word on a tick chart. The emphasis was on drawing the child's attention to the initial letter of the word and re-reading the page for the child, when there had been any hesitation or errors, to aid comprehension and fluency. Teachers were requested not to ask the children to 'sound out' the words, but not to stop them either if they chose to do so as this may have caused confusion, since it is currently the predominant strategy taught in Reception classes.

The original choice of font was 'Comic Sans', but this was revised following specific requests from a number of schools in Study 1 to use 'Sassoon Primary Infant' to allow for consistency with other reading schemes. This font was also used in the control schools as it is the font used in many published phonic-based reading schemes such as Read Write Inc. It is also more consistent with the font used in the York Assessment of Reading for Comprehension (YARC), which was used as an assessment tool in this research. There has, however, been research suggesting that word recognition is slower using a font such as Sassoon Primary Infant, which has a high similarity in shape between neighbouring letters, compared to one with less similarity such as Verdana (Wilkins et.al, 2009). For Studies 2 and 3, the font was changed back to Comic Sans following consultation with a number of teachers, SENCOs and head teachers, who expressed a concern regarding Sassoon Primary Infant font, and in light of the research evidence. The font size used in both the resources and books was larger than that usually used in reading books for 5-year olds. The use of a larger font size is also supported by research which has found that a larger font size increases the average reading age for 7-9 year olds by 4 months on the Salford Sentence Reading Test (Wilkins et.al, 2009).

In the weebee Reading Programme there are two parallel sets of text (P and A). Text P contains only words which should be phonically decodable by children in Reception classes who are following the structured sequence of phonic sounds as set out in the National Curriculum (DfE, 2014a). Text A is intended to replicate the kind of language used in 'real' books, with no constraints as to the use of vocabulary other than being age appropriate. Both texts use repetition and introduce the same number of words per book, have the same length of sentences, and the same number of pages as well as having identical illustrations. There is also an element of lesson-to-text match in both (see Appendix E). These criteria have

been used in a number of analyses of reading texts, including measuring the effects of using decodable texts with young readers (Jenkins et al, 2004; Mesmer, 2009).

3.4.3 Selection of Vocabulary

The selection of words for Text P was largely consistent with the phonic phases as detailed in the National Curriculum (DfE, 2014a). They did not follow in exactly the same way, as this would have limited the ability to produce a predictable text, and achieve a story line that could correspond with the alternative vocabulary. The National Curriculum also details a selection of specific irregular words which are also required to be taught alongside phonics and these were incorporated in the suggested order. Text A used irregular words, with no intention to teach specific sounds or letter groups. The focus of Text A was an authentic naturalistic story-telling language.

In order to control for the use of context and illustrations, predictable text was used for both versions. Selection of the vocabulary for the controlled Text P was made first as this had to conform to a set of pre-determined guidelines, based on the 'Phonics Phases' as detailed in the National Curriculum document 'Letters and Sounds' (DfES, 2007). Repetition of selected words and letter groups was used in a similar way to existing reading schemes. Irregular words were prioritised according to the 'key words to literacy' list that was drawn up following research into the most commonly used words in children's text (McNally & Murray, 1962). Selection of grapheme-to-phoneme correspondences (GPCs) followed the recommended pattern in the National Curriculum (DfE, 2014a), but more complex GPCs were also introduced earlier than recommended, to reflect the number of words being introduced. Although the vocabulary selected for Intervention P was limited by the constraints of the phonics phases, the books were nevertheless written in such a way that children could make some predictions using cues from context, grammar and the pictures. Intervention P used the structured phonic-phases vocabulary in context rather than in a context-free scenario.

The alternative Text A for the 'real books' was not restricted in any way other than to be congruent with Text P. This meant the introduction of the same number of new words per book and keeping to the same number of total words per book which told the same story. (As the texts became longer and more complex this limit was to within ten words. This flexibility was intended to avoid compromising

the story line which is an essential element of authentic text.) Inevitably, some of the vocabulary was the same for both sets of books as they both had to have essentially the same story line and also to fit with the same illustrations. Repetition was used in the same way as for the parallel set of phonically regular books, which acted as reinforcement for particular words and to foster the development of a sight vocabulary; a similar use of repetition as was used in the pre-book activities and games. There is evidence that beginner readers can remember words after four repetitions (Hiebert et al, 2005). Some of the words used were deliberately complex to more closely reflect the stimulating vocabulary found in authentic text (Baumann et al, 2007). There was also greater use of multi-syllabic words (see Appendix F).

3.5 Summary

This Chapter describes the design of the intervention used in all three studies. The rationale for the use of games is detailed in the discussion of the concept of learning through play and the value of group learning. There is an explanation of the training given to teachers and the support provided. The teaching resources have been detailed individually and a rationale given to justify the use of each item. The resources fall into five general categories although there is some overlap. The first category was concerned with extending oral vocabulary (songs, Big Book). The second used analytic phonics (draw-through worksheets, Happy Word Families, Fishing). The third category included the activities designed to build a sight vocabulary (Fishing, Jig Words, Bingo, Memory, Dominoes, Snap, Pento, Grog's Journey, Snakes and Ladders). The fourth category was concerned with comprehension (Sentence Matching); the oral vocabulary and sight word activities all contributed to this. The fifth and final category was the design of the books themselves including both the illustrations and the text including the selection of vocabulary for each set of books. The methodology and design of the trials using the *weebee* Reading Programme for this research are detailed in Chapter Four.

Chapter Four

Methodology

This Chapter details the methodology, and data collection procedures, for the three distinct, but related, studies designed to evaluate the effects of using different kinds of vocabulary, and teaching approaches, used in the weebee Reading Programme (a full protocol is included in Appendix G). Included here are: the aims and objectives of the studies; the design of the trials; recruitment; sample sizes; data collection and analysis; process and implementation fidelity.

4.1. Preliminary Preparation

There were a number of small-scale pilot trials of the various materials to be used in the programme. Firstly, with a small number of individuals from a school local to the researcher, to assess the game designs and for feedback on the original concept art for the story-book characters. Subsequently, the teaching materials (games) were taken into the Reception class of one of the schools which had agreed to participate in Study 1. (These children would have moved into Year 1 by the time the study began.) Groups of children trialled the games and materials, and gave feedback, which led to a number of modifications to both the teaching materials and to the instruction manual for those delivering the programme. This also helped in the planning for training sessions. Changes were made to the choice of font; the thickness of the materials and the designs for three of the games. The children who participated in the pilot were approaching the end of their Reception year and so were confident enough to give some verbal feedback.

In addition, a volunteer child who had been struggling to read at school agreed to trial the first set of books and materials, which made it possible to make an initial assessment of the efficacy of the intended teaching approach. Some modifications to a number of the activities resulted from these sessions. Further modifications were made following the completion of the first term of Study 1 and collection of feedback from class teachers.

4.2. Aim of the Studies

The overall aim of the three studies was to address the two main research questions. First, are there measurable differences between vocabulary that is within a child's existing decoding ability (Intervention P) and vocabulary which is not so

constrained (Intervention A)? Second, are there measurable differences when comparing a synthetic-phonics only approach with a mixed teaching approach? Specifically, will differences be observable in word recognition ability, comprehension and oral vocabulary.

4.3. Objectives of the Studies

The main objective was to attempt to isolate and explore the independent variable, in this case the vocabulary, and control for all confounding variables (illustrations, number of words in text, font, predictable storyline, number of new words per book, and learning materials). The objectives for each individual study related to the assessment of different aspects of the Intervention.

Study 1 (reported in Chapter Five) was an effectiveness trial, to evaluate the impact (in terms of the two research questions) of the ɯeebee Reading Programme and to explore the feasibility of implementation in a 'real world' natural classroom setting. A secondary objective was to assess the teaching materials developed as a response to theories of teaching and learning as described in Chapter Three.

Study 2 (reported in Chapter Six) was an efficacy trial, and sought to trial the ɯeebee Reading Programme in relatively ideal conditions, in order to explore the differences in the use of controlled or language-rich vocabulary more fully.

Study 3 (reported in Chapter Seven) sought to explore the value of the ɯeebee Reading Programme as a short-term intensive targeted intervention for struggling readers.

4.4. Controlling Confounding Variables

All three studies used two parallel sets of books and teaching resources, where both sets were equal in all ways, except in the selection of vocabulary, controlling for illustration, word count and storyline, as detailed in Chapter Three. A second element of the trials related to the teaching methods being used as part of the Intervention. An attempt was made to control for teaching methods as a confounding variable by stipulating the teaching approach for the Reading Programme. It was hoped that there would be continuity across the two Intervention conditions in each study as a result. The chosen approach to teaching was an eclectic one, exploring the effect of mixing phonics instruction with other

methods of teaching, compared to a synthetic-phonics only approach, which is currently the norm in English mainstream state schools. Although all schools are expected to use synthetic phonics, there is evidence that many teachers still use other methods and that there is considerable variation in teaching methods in Reception classes (Walker, Sainsbury, Worth, Bamforth, & Betts, 2015). By deliberately using several different traditional teaching approaches, it was hoped to increase the likelihood of children experiencing similar learning environments in both arms of the Intervention.

By using a Programme Manual, it was hoped that delivery of the programme would be consistent across all schools. Details of session plans and games procedures were provided in step-by-step sequence. The language to be used for introductory and plenary activities was given as scripts. All the materials were provided to the schools in advance. All teachers were given training in the use of appropriate language, and the type of teacher input required, limiting the risk of variation (see Appendix G).

4.5. Design of the Trials

The chosen design for all three studies was a three-armed controlled trial, consisting of Intervention A (non-phonically decodable vocabulary and mixed teaching methods), Intervention P (phonically decodable vocabulary and mixed teaching methods) and a Control condition (synthetic phonics only). It was essential to have a Control group to account for temporal changes and regression to the mean effects (Torgerson & Torgerson, 2008). Interventions A and P were randomised to condition to control for selection bias and to ensure internal validity. Randomisation was assigned independently at the Institute for Effective Education, University of York, by the data manager, to reduce selection bias (Hutchison & Styles, 2010). The Intervention arms were compared to a 'business-as-usual' Control group, following the National Curriculum, which were either matched or self-selected. All three studies lasted for approximately one year, with Studies 2 and 3 beginning two terms later than Study 1. Study 1 was conducted with twelve primary schools in west Worcestershire. Study 2 was carried out in four schools in Herefordshire and Study 3 was conducted in two classes from one school in Hereford. The three arms for each of the trials is shown in Figure 4.1 below.

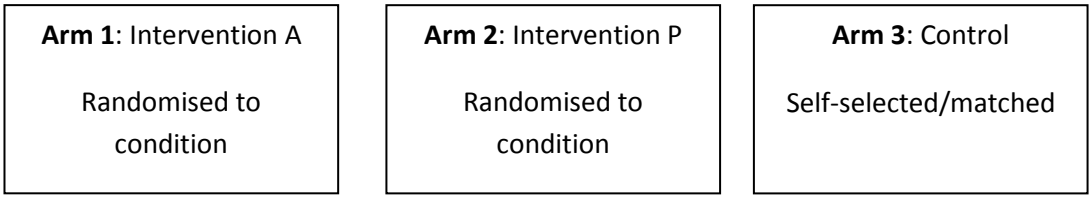


Figure 4.1 Diagram showing the three arms of the trials

4.5.1. Recruitment

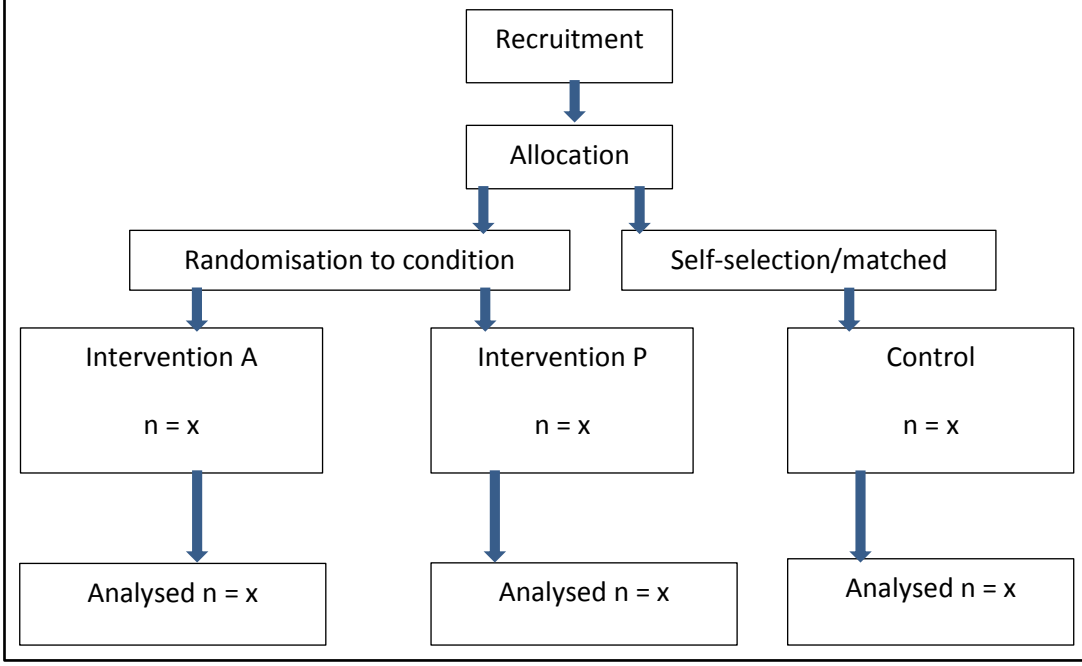
Recruitment for Study 1 began in the spring of 2013. Twelve schools were recruited; they were contacted directly by phone or email in the first instance. Eight of the twelve participating schools agreed to implement the Intervention and were then randomised to condition. The other four schools self-selected as controls, giving a number of reasons for not wishing to participate in the Intervention itself, such as imminent Ofsted inspections and plans to introduce an alternative reading scheme. None of the schools were prepared to accept not knowing in advance whether they would be using the Intervention or carrying on with 'business as usual' as would be the case from using full randomisation, particularly as they would have to plan for training and adjustments to timetabling.

Recruitment for Study 2 began in the spring of 2014. All schools (seventy-two) in a predominantly rural county were sent an invitation to participate, via post, with detailed descriptions of the Intervention and the materials and books that the schools would receive free of charge. Only four schools volunteered to participate. This may well have been due to the imminent changes for the new National Curriculum for September 2014. The small number of schools agreeing to participate in Study 2 could have led to recruitment bias, however, the trial design was intended to control for that possibility by having the two Intervention conditions within-class rather than between-class. A request from one of the schools to focus on struggling readers led to Study 3. This kind of study design, which limits interactions to particular groups, has been found to be an easier way for researchers to gain access to schools (Hill, King, Lemons & Partanen, 2012).

Children from the Reception class of each school participated, up to a maximum of thirty per class. Class teachers were not expected to participate directly unless they particularly chose to, although they were required to facilitate the Teaching Assistants under their immediate direction. The Teaching Assistants

normally assigned to the Reception class by the head teacher were the main deliverers of the Intervention. For Studies 1 and 2, the whole cohort of each Reception class from each school was recruited. For Study 3, the teachers of two classes were asked to select six children who they considered to be struggling readers and who might benefit from a targeted reading intervention (see Chapter Seven). There were small numbers of children with English as an additional language in all of the participating schools, and all the children were able to communicate confidently in English. Schools in all three studies were mainstream state schools that would normally follow the National Curriculum. None of the participating schools had been deemed unsatisfactory by Ofsted. Teachers in all participating classes (both Control and Intervention) were to continue with their normal curriculum in addition to the Intervention. Intervention and Control schools were given copies of the assessment results for their children to use for their own information, which was an incentive for participation. A generic diagram showing the flow of movement through the trials is shown in Figure 4.2 below.

Figure 4.2 CONSORT flow diagram n = children
The diagram indicates the flow of children through the trials



4.5.2 Ethics

Letters providing information regarding the nature, timescale and commitment expected for the studies were sent to all schools as well as consent forms. Copies of

letters that could be sent for parental consent were also attached; this was on an 'opt out' basis (see Appendix H). This was followed by a visit, by which time teachers would have had the opportunity to handle the teaching resources, Manual and books and were able to ask questions. Head teachers, teachers and teaching assistants all signed consent forms prior to the first round of assessments and parental consent was not withheld in any of the schools. The study protocol and other documents were approved by the University of York Education Ethics Committee (April, 2013). It was made clear to schools and parents in the correspondence that it was possible to withdraw from the study at any time (see Appendix H). All schools agreed to allow individual testing at both pre and post-trial stages. All data from schools, teachers and children was anonymised and allocated numeric codes. All raw data, including observations and audio recordings was stored in a secure location. Signed consent forms were stored together with the raw data.

4.5.3. Sample Sizes

The number of schools involved was unlikely to be sufficient for a cluster or school-level design; however on an individual level the sample sizes for Studies 1 and 2 were felt to be sufficient. These are reported in detail for each study in their respective chapters. Effect sizes (which emphasise the difference between groups) were calculated by taking the differences in post-test scores and dividing these by the pooled standard deviation of the post-test scores (Torgerson & Torgerson, 2008).

4.5.4. Data Collection

For all three studies there were two main data collection points: pre-intervention and post-intervention. The pre-test assessments were undertaken on entry into Reception in September for Study 1, and at the beginning of the summer term for Studies 2 and 3. Additional data in the form of audio recordings and fidelity checks was collected at one other point for Study 1 and at two points in Studies 2 and 3, as well as the collection of feedback from teachers at post-test.

All children were assessed individually in a quiet area either within the classroom or just outside (with the door open and the classroom visible). The researcher and child were visible at all times, either through an open door or in a corner of the classroom and the teacher remained both visible and audible. This

meant that there were distractions, but this was consistent for all participants. Assessments, which were carried out by the same researcher for all children, took approximately fifteen minutes per child. A script was followed to avoid researcher bias (Reed, Cummings, Schaper & Biancarosa, 2014). It was possible to test between twelve and fifteen children per day; up to two day visits per class.

The total number of children assessed in Study 1 (at time 1) was 282 over a period of one month. The total number of children assessed in Studies 2 and 3 collectively (at time 1) was 90. It was not possible to control for time of day, which for some of the younger ones may well have had an impact, although this applied across all schools. All the assessments were carried out by the researcher, following a training and practice session with a volunteer child, not connected in any way with the participating schools, to increase the likelihood of consistency. In addition, the researcher was kept blind to condition at both pre and post-test. This was intended to ensure consistency in the use of the published materials in all the standardised tests, in both the forms used and the method of scoring.

Data for assessing fidelity (the extent to which children received the intended learning experience from the reading programme) for Study 1, was collected at one midpoint (audio recordings of group teaching) and at post-intervention (Questionnaires, feedback forms, session record sheets and individual reading records). As a result of general poor fidelity during Study 1 (see Chapter Five), there was more focus on both training and monitoring during Studies 2 and 3. Audio recordings were collected, and observations of lessons made, at two mid-points and performance feedback given. Feedback forms were also collected at these times to monitor enthusiasm and commitment. Following feedback and observations, increased support was provided for two of the schools in the form of modelling sessions and listening to readers. At post intervention questionnaires, feedback forms, session records and individual reading records were collected.

4.5.5. Assessment Tools

The assessment tools chosen for all three studies were the same, although not all the elements were included in each or at both pre and post-intervention due to the ages of the children at the time, and the risk of both floor and ceiling effects. Measures were chosen to assess word recognition and comprehension. Studies have suggested that a large proportion of variance in reading comprehension is attributed

to vocabulary (Protopapas et.al, 2013). More specifically, oral vocabulary has been found to be a strong predictor of semantic learning (Ricketts et al, 2011) and increases in oral vocabulary have been found to have a beneficial effect on reading comprehension (Clarke et al , 2010). Therefore, assessment tools were selected for their ability to measure receptive vocabulary, comprehension, phonic knowledge and early word recognition. Multiple measures were used to gain a broad spectrum of results to give a more accurate indication of reading comprehension (Silverman et al, 2013). Assessments were administered pre-intervention to establish baseline performance, and at post-test. The assessment tools used are shown in Table 4.1.

Table 4.1
Table showing the assessment tools, their reliability, age range and duration

Measure	**Reliability**	**Age Range**	**Time taken**	**Administration**
British Picture Vocabulary Scale (BPVS)	Reliability 0.91 (Cronbach's Alpha)	3 - 15	Approx. 5 minutes	One-to-one
York Assessment of Reading for Comprehension (YARC)	Reliability 0.77 (Cronbach's Alpha)	4 - 7	Approx. 10 minutes	One-to-one
YARC Tests used: • Letter Sound Knowledge (LSK) • Early Word Reading (EWR) • Sound Isolation (SI) • Sound Deletion (SD) • Passage Reading Comprehension (PRC)				

The British Picture Vocabulary Scale (BPVS-III) (Dunn et al, 2009) is norm-referenced and designed to measure the receptive (understood) vocabulary of children aged between 3 and 15 years using pictures. The BPVS is an English adaptation of the Peabody Picture Vocabulary Scale. Each child has to identify one picture item from a set of four which they think represents the word spoken by the administrator of the test. There are fourteen sets of four pictures. Raw scores are converted to standardised scores based on age-related norms. By using pictures, children do not require any pre-existing knowledge or reading skills. As the independent variable in the trials was reading vocabulary, a measure of the children's receptive vocabulary at pre and post-intervention was likely to be informative, particularly as the Intervention included activities designed to extend children's spoken vocabulary, through discussion of aspects of the books, and

relating these to their own experiences. The BPVS III has been standardised in England using a sample of 3,278 children. For children aged 3-5 the sample was 629. For this age group, the norms were based on imputed raw scores using basal and ceiling rules. Raw scores were then converted to standardised scores with a mean of 100 and a standard deviation of 15 (Dunn et al, 2009). The BPVS has a clearly defined protocol with guidance for its use and scoring.

The York Assessment of Reading for Comprehension (YARC) is designed to measure three areas of reading and comprehension: decoding (phonic knowledge), fluency and comprehension (Snowling et al, 2009). The test takes approximately 10 minutes per child. The YARC Early Reading test was selected to assess a number of reading skills including: Letter Sound Knowledge (sound-to-letter correspondence); Early Word Reading (assesses children's ability to read high frequency words from the early stages of learning to read); Sound Isolation (distinguishing the sounds at the beginning and ends of words) and Sound Deletion (the ability to delete single phonemes from a word). The Letter Sound Knowledge and Early Word Reading measures were administered both pre and post-intervention. The Sound Deletion and Isolation measures were not included at pre-test for Study 1 as it was felt that this would be beyond the ability of many of the children, some of whom had only just turned four years of age, thereby resulting in likely floor effects.

At post-intervention, the Passage Reading Comprehension measure was included in the battery of assessments in order assess levels of comprehension. This measure was only administered at post-test as very few children had any significant word recognition at pre-test. The YARC has a beginner passage that is a shared reading task which made it possible to measure comprehension at what was a very early stage of learning to read. A single-word reading measure is used to decide the starting level for the passages, rather than chronological age. The Passage Reading measure is designed to make a distinction between literal understanding, and comprehension by inference.

The YARC was standardised in 2008, in England, using a sample of 1,376 children from ten different regions. For children in Reception class the mean adjusted score was 101.51 (*SD* =14.73, n=157). Accuracy for age equivalence mean score was 5.08 (*SD* =10, n=157) (Snowling et al, 2009). Tests with children across the age groups demonstrated that the comprehension questions were

dependent on information from the passages and therefore considered to be a valid measure. High correlations were found with the BPVS and it has been co-normed with the British Ability Scales (BAS III) (Snowling et al, 2009).

The tests chosen all had clearly defined protocols for use and had been standardised for objective comparison. Standardised scores were used for the majority of measures (the test of Nouns and Reading Speed were non-standardised measures constructed specifically for this set of trials). For very young children, the effect of age is highly significant, given that there can be up to eleven months difference between ages of the children in the same cohort. Standardisation compensates for age, and was therefore an important consideration for the measures used with the young children in these trials.

Nevertheless, there was still a risk of implementation bias from a number of possible factors such as: variation in the environment; changes in word order or emphasis; distractions, or simple mistakes (Reed et al, 2014). The researcher remained blind to allocation of Intervention, but not to Control, at pre and post-test. This could have been compromised by the observation of teaching; however, the number of children being assessed made it unlikely that the researcher would remember which children had been using a particular set of vocabulary. To further minimise variation, the tests were administered in the same order and instructions given in the same way to all children, following a script.

For Study 1, an Intervention-specific word recognition test was added at post-test, to determine if children in both Intervention arms were able to recognise more words, which had been specifically taught, compared to the Control condition by building up a sight vocabulary. The words chosen were thirty two nouns common to both sets of books. The words were presented in the form of a book with eight words per page in large font size, in a similar format and using the same number of words as the Early Word Reading test from the YARC (Appendix I).

As the children in Studies 2 and 3 were two terms older at the time of assessment, it was highly likely that there would have been ceiling effects from using the Nouns test used in Study 1. Therefore, a different book was created, using only words common to both Intervention conditions (A and P). This was designed to look the same as the books in the reading scheme, but without the characters, so as not to disadvantage children in the Control condition. The book

used predictable text and illustrations in keeping with those used in the reading scheme, instead of a simple list of un-connected words as used in Study 1 (Appendix J). The reading was timed, to give an indication of reading fluency (Fuchs, Fuchs & Hosp, 2001; Mesmer, 2009).

For all measures, an analysis of the impact on the gender gap was included. There is consistent evidence in the research literature that girls outperform boys on tests of reading comprehension (DfE, 2013b; DfE, 2014b; Logan & Johnston, 2010; National Literacy Trust, 2012). This gender gap is evident even in Reception classes (DfES, 2009). According to data analysed by the National Literacy Trust (2012), girls of this age enjoy reading more than boys and spend more time reading. The choice of story books characters and setting used in the weebee Reading Programme were intentionally chosen to appeal to boys of this age group to help reduce motivational effects caused by gender (Mohr, 2006). The National Literacy Trust also emphasise evidence that reading and writing are linked and that the gender gap is widest in literacy.

The systematic review of research conducted by Torgerson et al (2006) found no clear evidence to indicate that the use of phonics teaching affected the gender gap. However, the Clackmannanshire study claimed that synthetic phonics benefitted boy's progress, although this was only for word decoding and made no impact on reading comprehension (Johnston & Watson, 2005). Thus, it was felt important to explore which measures, if any, may demonstrate an impact from the use of the Intervention on the gender gap. An analysis of the gender gap was in respect of both research questions (the use of mixed teaching methods and non-decodable vocabulary) and was also used for struggling readers in Study 3.

4.5.6 Intention to Treat Analysis

The decision was made to use intention to treat analysis (ITT) to avoid bias, although there is some debate as to its value in the literature. ITT analysis is the technique of analysing the results of randomised controlled trials by the group that a subject was initially randomised into, regardless of attrition or crossover (Cunningham, 2011). Although the 2010 CONSORT checklist no longer includes intention to treat analysis (ITT), it does recommend that information is provided as to whether an analysis was by original assigned groups, and what numbers were included in the analysis (White, Carpenter & Horton, 2012).

White et al (2012), recommend that analysis using baseline data as a covariate needs only to include those individuals with the outcome observed at follow up; in other words individuals for whom there are both pre-test and post-test scores. Torgesson & Torgesson (2008), acknowledge that although there may be dilution of effect, ITT cannot alter the direction of effect. Cunningham (2011) suggests that it makes no sense to report the results of participants assigned to one treatment but who actually effectively had another. Nevertheless, he goes on to say that by allowing the researcher to take into account non-compliance, ITT enhances a study's external validity. According to Booil, Asparouhov & Higgins (2008), the estimation of ITT effect can be biased in analysis that ignores non-compliance. Participants could be divided into compliers and non-compliers and then compared to each other and to the control group, but this would negate the benefits of randomization (a subanalysis of the impact of fidelity in Study 2 is included in Appendix O).

Poor or non-compliance related to the teaching and delivery of the Intervention (it would not affect control schools). Compliance in these studies varied between total compliance and non-compliance. Figure 4.3 below demonstrates the variety of factors that affected compliance in the studies.

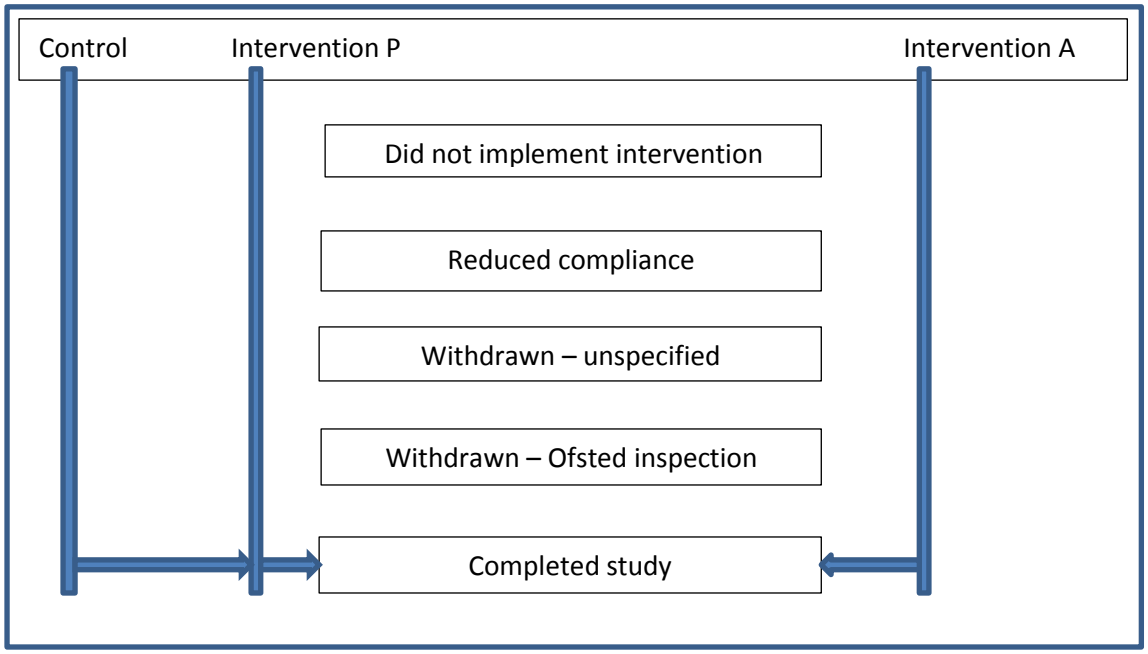


Figure 4.3 Diagram showing the factors affecting compliance

Hollis and Campbell (1999) described ITT analysis as being most useful for trials of effectiveness (such as in Study 1) rather than investigations of efficacy

(such as Studies 2 and 3). They suggest that in some circumstances it may be sensible to exclude non-starters (where the Intervention failed to be implemented, such as in one school in Study 1) as it is unlikely to lead to bias, when the intended effect of an Intervention depends on the occurrence of a subsequent event that cannot be influenced by the randomised allocation. In order to include non-starters and non-finishers, or those lost to follow up, data would have had to have been imputed artificially. Instead, all individuals who had both pre-test and post-test scores were analysed, including children from classes who had not completed the programme, or where there had been poor compliance.

Schools were not allowed to cross over to a different condition, and since using the pre-test scores as a covariate was planned, only pupils with scores at both assessment points were included (White et al, 2012). In spite of the risk of dilution of statistical significance, or effect sizes, by including those who didn't complete the programme (but did provide post-test scores), using intention to treat analysis (ITT) would be less likely to introduce bias, but would still indicate a direction of effect (Ialongo, Poduska, Werthamer & Kellam, 2001; Torgesson & Torgesson, 2008). There was no evidence in any of the studies that missing data was connected to the Intervention and it was therefore considered acceptable to exclude these individuals without causing bias or the need to impute data which was considered to be missing at random.

4.5.7 Analysis of Results

As the main aim of the research reported in this Thesis was to compare the use of different kinds of vocabulary, at an individual level rather than at the school level, it was decided that individual-level analysis would be appropriate for all three studies. A combination of providing a Manual for the Reading Programme, detailed training given to all programme deliverers, and the provision of all teaching resources, will have considerably reduced any clustering effects caused by randomisation by school as units, as occurred in Study 1.

In order to answer Research Question 1, the analysis was planned as a contrast between Intervention A and Intervention P. This analysis were conducted as planned in all three studies. Independent samples t-tests were used to indicate levels of significance in the differences found between groups, using measures at both pre and post-intervention. In addition, effect sizes were to be calculated and

reported in terms of the between-group differences in standard deviation for all comparisons using Cohen's *d*, since this measure is not influenced by differences in group sample sizes (unlike η_p^2). Cohen's *d* indicates the difference between two means. An effect size of 0.1 can be considered to be educationally useful (Hutchison & Styles, 2010). Effect sizes have been calculated for each group comparing means at pre-test and post-test and then the differences between groups have been calculated (Gorard, Siddiqui & See, 2015). Estimates of reading age progress in months, as related to effect sizes were also to be reported, calculated according to the procedure described by Higgins et al,(2013). In order to answer Research Question 2 (comparing the mixed teaching methods used in both arms of the Intervention with the synthetic phonics only used in the Control condition), the analyses were planned as contrasts between Intervention A and the Control condition and between Intervention P and the Control condition using independent samples t-tests to establish group differences and calculating effect sizes to indicate educational value in terms of reading age progress.

Pre and post-test measures were used to make comparisons of relative progress, controlling for pre-test scores. For these measures, a preliminary analysis, evaluating the homogeneity-of-regression assumption, was to be run in SPSS to assess whether analysis of covariance could be run (ANCOVA). This analysis was to be used to control for scores at pre-test, using pre-test scores as covariates, using Bonferroni correction to combat the build-up of errors from repeated tests, and Kolmogorov-Smirnov (KS) to test for normality (Field, 2013). The same analysis was to be used for the non-standardised tests as well as the standardised measures (see flow chart in Figure 4.4).

There is evidence that currently boys as a group have lower scores for the phonics screening check at age 6, and that generally girls outperform boys in reading at this age (National Literacy Trust, 2012; Ofsted, 2012; Walker et al, 2014). Therefore, all three studies were to be analysed for gender differences in outcomes measured and any possible significant differences of gender distribution between each arm of the trials. This was to be done using Pearson's chi-square. The same analysis was used to compare the distribution of children with English as an additional language (EAL) for each arm of the trial, since high levels of EAL in any one arm may well have led to selection bias.

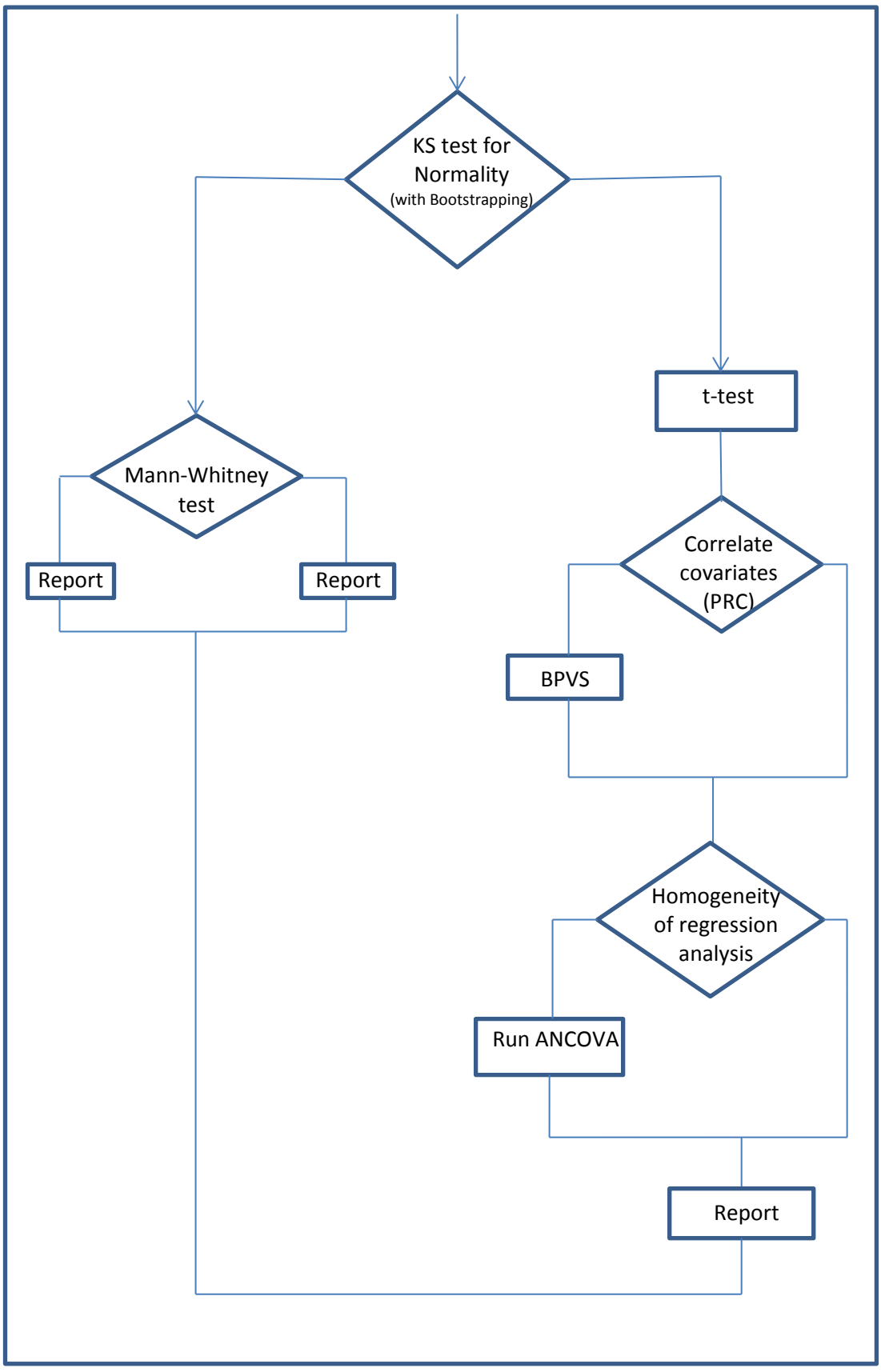


Figure 4.4 Flow chart of procedure for data analysis for all studies

4.6. Process and Implementation Fidelity

All three studies used in this research were conducted in normal classroom environments where the statutory curriculum had to be adhered to; where the usual timetable was not disrupted and where schools were subject to possible Ofsted inspections at any time. In addition, as these were long-term studies (twelve months), they were at risk of reduced compliance due to illness, unexpected events, and staff or children moving schools. It was important, therefore, to monitor compliance, and dosage, and to keep a record of attrition (loss of children's data), in order to assess the extent to which the Intervention was actually delivered and received. Without these kinds of measures (of adherence to an intended intervention design) it is difficult to determine whether outcomes actually reflect a result of the intervention, or why (Mowbray, Holter, Teague & Bybee, 2003).

4.6.1. Teacher's Manual

As stated earlier, the intention in these studies was to attempt to control all the dependent variables, most of which were concerned with the teaching of the specified vocabulary (i.e. the independent variable). Lack of treatment fidelity would have made comparison difficult. In order to try to reduce possible differences, a Manual was devised for teachers to follow in order to increase compliance to the programme protocol (Mowbray et al, 2003). All teachers in the participating schools were provided with a Manual, which included detailed instructions for each of the activities, and lesson plans. Having an easily accessible Manual meant that teachers could refer at a glance, for example, to the rules of a game. It also allowed for cover teachers to take over at short notice when necessary.

Lesson plans were devised to resemble the current practice in schools to increase the likelihood of implementation fidelity (Smith, Daunic & Taylor, 2007). The Manual identified the critical components of the Intervention, such as matching the words rather than sounding them out, and gave detailed guidance on how these should be implemented. Check lists were provided within the Manual for teachers to keep a record of each of the elements of the programme that had been covered for each of the sessions associated with each reading book (dosage). These check lists also acted as a useful tool for reminding the teachers of the requirements of the programme.

4.6.2. Training

All the teachers received a minimum of two hours training in the use of the materials, the language to be used, and the priority focus for each activity, for example, matching the first letter of a word. All teachers (Teaching Assistants, Class Teachers, Governors and other assistants) received the same training. This was school-based and was either for individulas or groups depending on the setting. Although rigid adherence to the programme protocol would have enhanced the internal validity of the trial, teachers were given flexibility in their planning. For example, the time of day was not specified and they were not given a set time by which all the children had to have completed all the activities. Most schools opted to use the Intervention as an afternoon activity. The training for Study 1 consisted mainly of instruction in how to use the materials. However, for Studies 2 and 3, the training was designed to create a rapport with teachers, and included regular supervision by the researcher. It was hoped that if the teachers understood the usefulness of the Intervention, they were more likely to implement it. This would depend on the content of the initial training, as well as on-going supervision and support (regular visits, modelling and performance feedback). There is evidence that interventions achieve higher fidelity of implementation, when they accord with the existing philosophy of a teacher or school (Harn, Parisi & Stoolmiller, 2013). Therefore it was important that there was appropriate training, and monitoring, particularly in relation to teachers' and schools' existing policies regarding their approach to literacy, within the constraints of the National Curriculum. From the outset, the books and activities were designed to look and operate in a similar way to existing resources used in Reception classrooms, including the group activities, for ease of implementation (Smith et al, 2007).

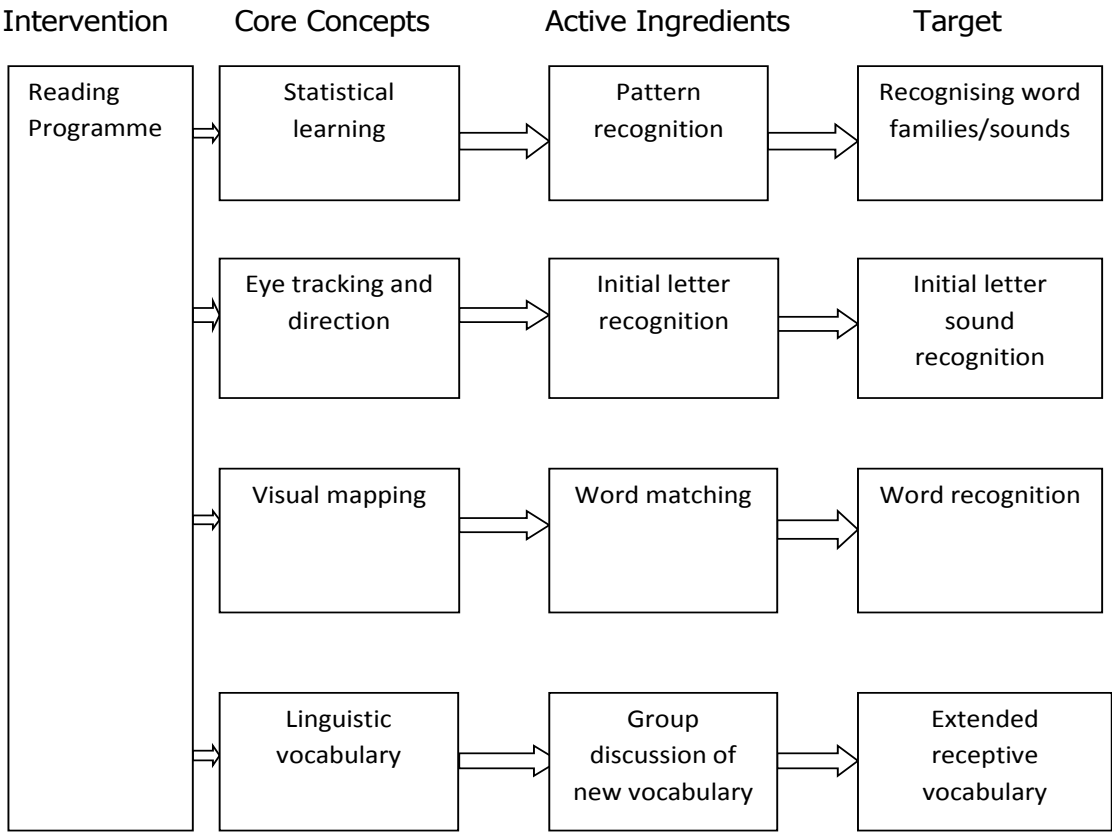
4.6.3. Framework for Assessing Fidelity

A framework for assessing fidelity was established, to identify the core components of the Intervention, and related measures, mainly to assess adherence and dosage (dosage being the amount of time and frequency of activities) (O'Donnell, 2008). As the programme was being implemented in natural settings, it was assumed that there would be some adaptation by teachers to their individual context, such as the number of pupils, the ratio of staff to pupils, the number of pupils with English as an additional language (EAL) and even the layout of the classroom.

Fidelity to the operationalization of the independent variable (and control of confounding variables) needed to be assured in order for outcomes to be attributable to the Intervention (O'Donnell, 2008). Darrow (2013) considers there to be five categories that need to be monitored to ensure fidelity: ideal adherence; quality; exposure; participant response; differentiation. For these studies, the Conceptual Model of core components, designed by Nelson, Cordray, Hulleman, Darrow & Sommer (2012), was used as a guide to a planned analysis of measures to be used to monitor fidelity, shown in Figure 4.5 below.

Figure 4.5 Conceptual Model of core components of the Intervention

A Conceptual Model, showing the core components and active ingredients of the intervention and how they relate to the intervention targets.



The Conceptual Model (Figure 4.5) represents theoretical processes rather than specific activities, and includes only the core components of the Intervention. The Core Concepts represent the theoretical and conceptual aspects drawn from the literature, which guided the design of the materials, as well as the selection of the specific activities to be included. These activities were the active ingredients intended to result in children reaching the desired target outcomes.

Having first constructed a Conceptual Model of the Core Concepts and Active Ingredients, it was possible to construct anticipated logic and change models (Nelson et al, 2012), which served as a guide to determine which outcomes to measure. Darrow (2013) recommends using this type of change model (Figure 4.7 below) when devising measures for fidelity which should represent the intervention, its primary constructs, critical elements and desired outcomes. The 'Target' outcomes column in Table 4.2 represents the anticipated changes which might differ from the Control schools. The tick sheets, observations and discussion groups referred directly to the Intervention and were not concerned with generic teaching skills or approaches. They did, however, ask teachers to detail if and how they had made adaptations to the programme to suit their own context. A number of measures were used to assess fidelity to each facet of the Intervention in order better to assess variance. The components to be assessed and measures used are detailed in the Table below (Table 4.2) for both implementation and process.

Table 4.2
Selected components and measures for assessing fidelity

Focus	Criteria	Validation of criteria	How to measure	Tool used	Target outcomes
Structure (Implementation)					
Big Book	Discussion of set questions	Importance of extending vocabulary	Questions to deliverers Observation	Focus groups Written and aural feedback Audio recording	Increased scoring of received vocabulary (BPVS)
Songs	Learning to sing the songs	Motivation and extending vocabulary	Questions to deliverers	Focus groups Written and aural feedback	Increased interest in the storybook characters and their environment
Plenary	Revision of phonic work	Importance of repetition in learning and value of teacher modelling	Questions to deliverers Observation	Focus groups Written and aural feedback Audio recording	Recognition of patterns in words (word families) YARC
Use of games	Informal learning of sight vocabulary	Statistical learning of word shapes and families	Questions to deliverers	Focus groups Written and aural feedback	Increased score of word recognition (YARC and noun list)
Dosage	Frequency Duration	Importance of all the children using all the activities for full duration of games	Deliverer to record each child's participation in an activity	Tick sheets Participant logs	Completion of the entire programme

Process (Intervention)					
Games	Learning how to play all the games and follow the rules	Group learning	Deliverer to record participation	Tick sheets Participant logs	Learning specific vocabulary as sight words
Books	Reading the books with the learned vocabulary	Reading to a skilled reader who can model correct responses when necessary	Deliverer to record incorrect responses	Tick sheets Participant logs Questionnaire	Reading books with ease, fluency and comprehension
New words	Reading through matching or looking at initial letters	Building up of a sight vocabulary	Observation	Audio recording	Learning 20 new words per book (specific to book)
Response of children	Enthusiasm Engagement	Engagement with activities promotes learning	Questions to deliverers	Focus groups Written and aural feedback	Full engagement in all activities with evident learning
Response of teacher	Enthusiasm Engagement Satisfaction	Valuing the programme promotes positive delivery	Ratings of satisfaction with training and resources	Focus groups Written and aural feedback Likert scale	Valuing the programme and the desire to use the programme in the future.
Character of deliverer	Experience, number or years teaching, level of CPD	Skill in understanding and implementing core concepts	Specific questions	Questionnaire	Relationship of teacher experience to fidelity of implementation

The flow diagram below (Figure 4.6), demonstrates the model as it relates to one specific component, in this case the 'Big Book' used for group discussion. The fidelity measures for all the Intervention components are linked to outcomes (Nelson et al, 2012).

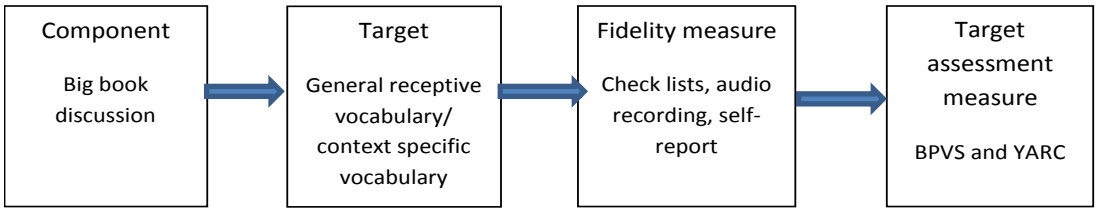
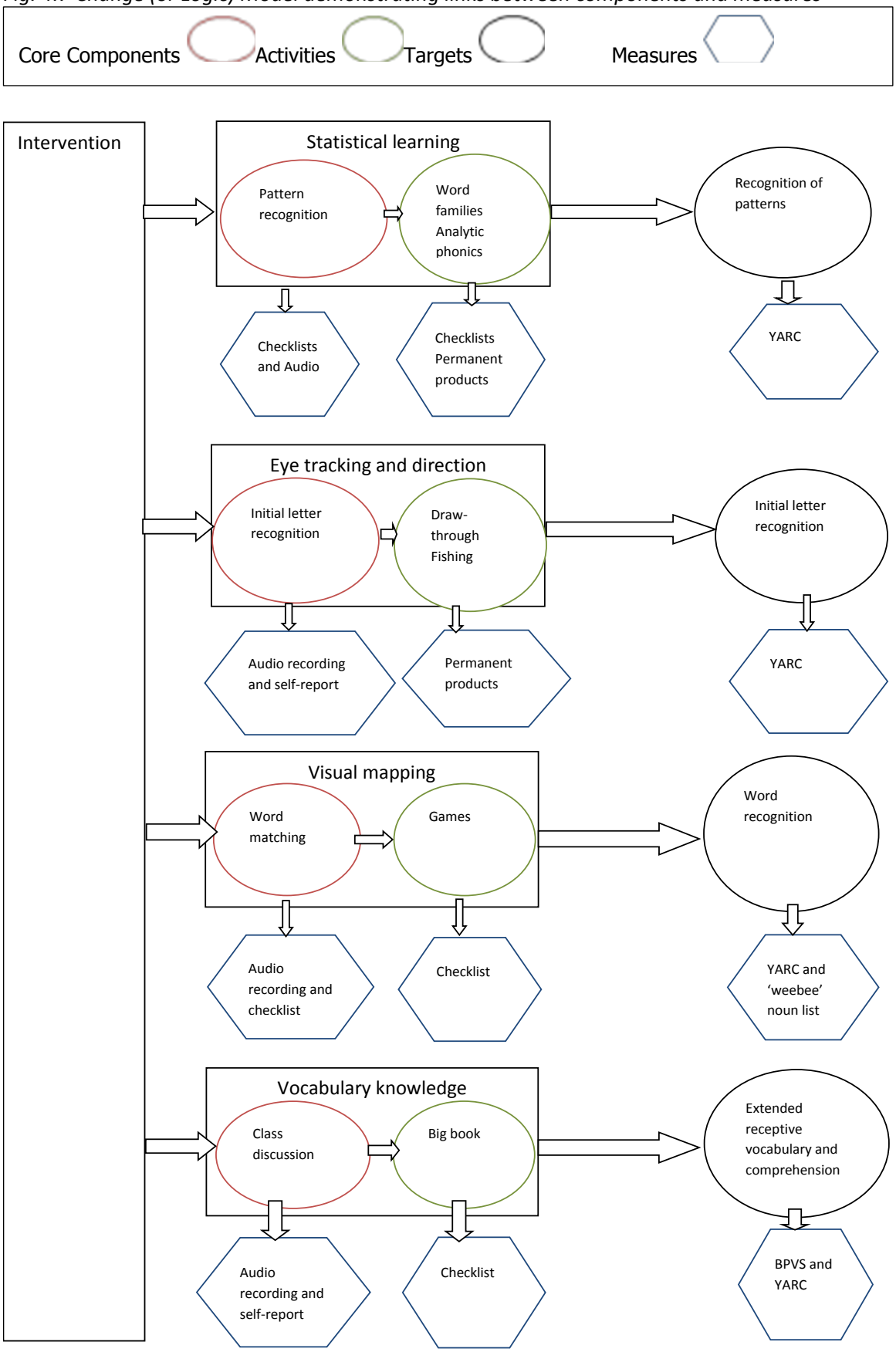


Figure 4.6 Change Model applied to the Big Book component

The change model below (Figure 4.7), demonstrates the links between the Core Components of the Intervention and the measures used to assess fidelity to the programme. It also shows the outcome measures used for each of the activities and targets.

Fig. 4.7 Change (or Logic) Model demonstrating links between components and measures



This Change Model (Figure 4.6) was used to validate the fidelity criteria. Mowbray et al (2003) recommend that the level of adaptation that a setting may legitimately make should be established at the outset, and therefore schools were given flexibility regarding implementation. The only stipulation was that they attempted to complete all the books before the end of the study. The change, or logic model, as described by Nelson et al (2012), represents the specific activities within the conceptual framework, but only includes the core components of the Intervention.

For detailed record keeping of dosage, quality of delivery and changes of circumstances, teachers were provided with a chart which could be filled in with simple tick boxes (see Appendix C) and a full protocol for each week (Appendix B). For the audio recordings, a research assistant was trained (and kept blind to allocation) to increase reliability in coding. The coding used specific criteria for each teaching technique (Mandell et al, 2013). Coding included: gaining child's attention; providing clear and appropriate instructions, such as drawing attention to the first letter of the word and/or matching the shape; using appropriate prompting strategies, such as asking children what they should look for, and asking if they could see which word it is the same as; and use of appropriate correction procedures, again focusing on initial letters and salient features for matching.

Assessment of fidelity to implementation was assessed by requiring the teachers to fill in a tick sheet, for each child, to indicate when they had completed each activity associated with a particular reading book. A record was also to be kept of words which children failed to recognise, while reading the book, having completed all the activities, using record forms provided. Components to be measured were selected prior to the start of the Intervention, and check lists were incorporated into the Programme Manual.

Assessment of fidelity to process (method of delivery according to the programme protocol) was through the audio recording, transcripts and observations using a pre-planned schedule. Rather than trying just to list all the expected behaviours, it was important to consider which undesired behaviours to look for when the observation schedules were being constructed (Mowbray et al, 2003). The observation schedules were intended to measure levels of fidelity, ranging from

non-use to refined use (see Appendix K). The criteria were based on the core components of the fidelity model as shown in Figure 4.4 above.

These fidelity measures were included to help explain why the Intervention may, or may not, have made a difference, by confirming if outcomes being measured were related to the delivery of the Intervention (Darrow, 2013). It was hoped that it would be possible to assess convergent validity by comparing information from the checklists, questionnaires, feedback forms and observations. During Study 1, programme deliverers were only required to state whether an activity had been completed. Collier-Meek, Fallon, Sanetti and Maggin (2013) suggest that a lack of fidelity of implementation in the classroom is often due to lack of time, confusion about components, and an overly complicated intervention. They stress the value of performance feedback, with reviews of progress, and reminders of implementation procedures, and that the children's expected outcomes should be clearly linked to these implementation procedures. For Studies 2 and 3, performance feedback was given to teachers at two time points, following the observations of teaching sessions. This was intended to ensure greater fidelity than had been observed in Study 1. For Studies 2 and 3, the measures included duration of each activity; size of the group; and frequency of sessions (Wolery, 2011). It was hoped that the audio recordings, plus the additional feedback forms, from Studies 2 and 3 would help to give some indication of the balance of delivery across the groups of children. The extent to which there may have been overlap with control conditions also needed to be measured in some way (Nelson et al, 2012). Questionnaires for teachers were designed to assess this (Appendix L).

Implementation monitoring needed to include frequency, intensity and duration and this was assessed through check sheets (Keller-Margulis, 2012). Keller-Margulis (2012) also recommended the use of performance feedback for intervention deliverers. For Study 1 there were no direct observations, only audio recordings of sessions which did not provide an opportunity to give performance feedback. However, in Studies 2 and 3, there were direct observations and the opportunity for feedback at two points, during which teachers were given the opportunity to ask questions, as well as receive confirmation as to correct procedure, and guidance where there were inconsistencies with the intended objectives. It was important to evaluate teachers' fidelity to the programme, to

assess effectiveness in actual use, and this included correct use of instructional language and the materials provided (O'Donnell, 2008; Harn et al, 2013).

Even with careful organisation, training and support, research suggests that quality and quantity of delivery of an intervention is likely to vary in school-based interventions (Wenz-Gross & Upshur, 2012). More experienced teachers have been found to implement lessons with greater fidelity, but also to have more negative attitudes towards new practices (Wenz-Gross & Upshur, 2012). In addition, fidelity to an intervention protocol, has been found to be related to teacher skill level, programme acceptance, morale, and self-efficacy. Teacher attitudes towards implementation of an intervention have also been thought to be affected by the existing philosophical climate in a school, and staff perceptions of an intervention's fit with their own values, practices and timetabling structures (Wenz-Gross & Upshur, 2012).

Higher levels of fidelity are associated with higher level outcomes from the use of an intervention (O'Donnell, 2008). In light of this, for Studies 2 and 3, a questionnaire (Appendix L) was designed to gain an idea of the amount of time the programme deliverers had worked with children, their teaching experience, and what kind of continuing professional development they had received. In addition, they were asked to provide information to assess the degree of overlap of the teaching approach usually used and that used in the programme. This also allowed a more accurate picture to be drawn of the actual differences between the Intervention and Control arms of the trials, although they were few in number (Nelson et al, 2012). Given the high levels of attrition in Study 1 (leading to potential attrition bias) the training and support for Studies 2 and 3 were much more intensive, in the hope of reducing this effect.

The data collected with regard to fidelity of implementation of the programme, included information regarding attrition and compliance. For example, looking at the number of books children had actually read, the number of activities completed by each child and the time spent on activities. Observations designed to assess fidelity of implementation included items that related to teacher behaviours. Each item was rated on a scale of 0 (observed none of the time) – 3 (observed and refined, see Appendix K). The reading record sheets related to child response to the Intervention. Each classroom was to be rated for overall fidelity, following an

analysis of a combination of all the fidelity measures. The observations were also intended to provide performance feedback. Nelson et al (2012) point out that increasing the number of measures of fidelity can increase the statistical power to detect relationships.

4.7 Summary

In summary, there were three separate studies, each using the same materials, but with different objectives. All three were three-armed controlled trials, with participants randomised to Intervention condition for the primary research question (Intervention A compared to Intervention P). Vocabulary was selected as the independent variable, and measures were put in place to control for confounding variables, within the structure and process of the Intervention. In total, sixteen schools were recruited, 372 children, and more than thirty teachers. Data was collected at pre and post-test, as well as at 1 further point for Study 1, and 2 further points for Studies 2 and 3. The majority of the assessment tools were standardised tests (the BPVS III and the YARC), and there were two additional programme-specific measures, which had not been standardised. The decision was taken to use intention to treat analysis (but to exclude non-starters and those lost to follow-up).

Given that the research presented here involved studying activity in naturalistic settings, and asking questions about learning in classrooms in relation to a number of theoretical approaches, it did not seem appropriate to begin with an hypothesis, such as would be expected for a purely experimental design. However, as indicated in Table 4.2, there were anticipated effects on target outcomes, such as differential scores on measures of reading ability. The work presented here is more in the nature of 'design-based research' which is intended to evaluate the Reading Programme in context (Barab, 2014). This context is an integral part of the learning mechanisms being studied and includes the social and political environment affecting the implementation of the Reading Programme.

A Change Model was constructed to represent the Core Concepts of the Intervention, the active ingredients, and the desired target outcomes. From this it was possible to assemble measures of fidelity in terms of both structure and process. To increase the likelihood of fidelity, the rogramme was manualised, and training was given to all teachers. Results of both standardised and non-

standardised tests were to be compared using independent samples t-tests, and effect sizes reported. In addition, where measures were repeated (pre-test and post-test), pre-test scores were to be used as covariates within ANCOVA, in order to take account of differences that already existed at pre-test. The results from Studies 1-3 are reported in Chapters Five – Seven respectively.

Chapter Five

Study 1: Three-armed Controlled Effectiveness Trial

This Chapter includes a brief outline of the specific methodology for Study 1 and then presents the data and an analysis of the results. This includes the aims of the trial, design and evaluation of the trial, limitations of the study design, and pupil characteristics. This is followed by an analysis of the outcomes, including results of assessments, an analysis of measures in respect of gender, fidelity to the programme and teacher feedback. A summary, discussion, and conclusions from the study outcomes follow at the end of the Chapter.

5.1 Aims of the Trial

The general aims for this study were to teach children new words, for both spoken and reading vocabulary, through playing games and other activities, using an eclectic approach to developing word-recognition skills. In addition, the Reading Programme was intended to enhance the children's understanding of narrative, through the reading books and songs created for the programme.

The specific aim for Study 1 was to use an independent groups, one-year design (congruent with the school year), as a trial of the effectiveness of the Reading Programme in a 'real world' classroom setting. The objective was to evaluate both the implementation and impact of the programme in terms of Research Question 1: Intervention P (phonically decodable vocabulary) compared to Intervention A (non-phonically decodable) and Research Question 2: the Intervention (whole-word recognition, analytic phonics plus synthetic phonics) compared to the Control (synthetic phonics only). The desired primary outcome was for improved reading ability, thus the primary outcome measures, chosen for this trial, were word recognition and passage-reading comprehension. The assessment measures were selected on this basis, as discussed in Chapter Four.

Secondary outcome measures were intended to assess the feasibility of implementation of the Intervention in a classroom setting. Measures were constructed for process evaluation, and to assess fidelity to the programme design for the whole trial period. The trial also sought to explore how the Intervention could be implemented by Teaching Assistants in addition to Class Teachers. The trial was carried out in the natural community setting in which it would be expected to

be adopted. This allowed for measurement of the realistic effect of the Intervention, within current school practice (Mandell et al, 2013). The intention was to study the Intervention in its natural setting, by considering the contextual factors as anticipated variables (Dingfelder & Mandell, 2011).

5.2 Design of Study 1

This section includes the structure of the trial, recruitment, sample size, timescale, data collection, and process evaluation procedures.

5.2.1 Three-armed Controlled Effectiveness Trial

The study design was a three-armed trial, with paired randomisation (equal numbers of rural and urban schools), including controls (participants formed into triplets); comparing schools using an intervention with equal numbers of carefully matched control schools (Slavin, 2003), as shown in Figure 5.1 below. The Intervention was compared to a 'business-as-usual' control group, following the National Curriculum. The study was a one-year trial beginning in September 2013. It was set in primary schools in a region with a mix of rural and urban schools. The schools were matched in terms of location, size (single-form entry), and socio-economic group. There were two data collection points: pre-intervention and post-intervention. The three arms were: Control schools (2 urban and 2 rural); Intervention P (2 urban and 2 rural); Intervention A (2 urban and 2 rural).

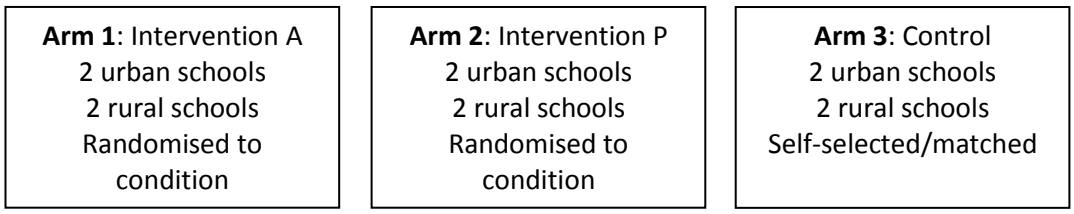


Figure 5.1 Diagram showing the three arms of the trial for study 1.

A matched design was used to reduce bias in terms of school type. Total randomisation in a small sample could have led to chance bias (Torgerson & Torgerson, 2008) whereby, for example, all the intervention schools could have been small rural schools with little or no English as an additional language (EAL) and few children receiving free school meals. The differences in class sizes between these two types of schools would be likely to have an impact on this type of

intervention as it requires considerable teacher input; therefore it was considered important to balance class sizes in each arm of the trial.

The programme was designed to run for three terms, beginning in October, immediately after the first round of assessments, and finishing in May, before the final round of assessments. It was anticipated that each child would spend at least ninety minutes per week on programme-related activities. This included a mix of whole class and group work; designed for groups of up to four children. Teachers were given the freedom to choose how they timetabled the sessions; the majority chose to implement the programme during afternoon sessions because of the need to adhere to the timetabling of morning activities. In some schools, the children were taken out in groups to another classroom by the teacher. In other schools, teachers chose to have the group activities integrated within the classroom.

5.2.2 Recruitment and Training

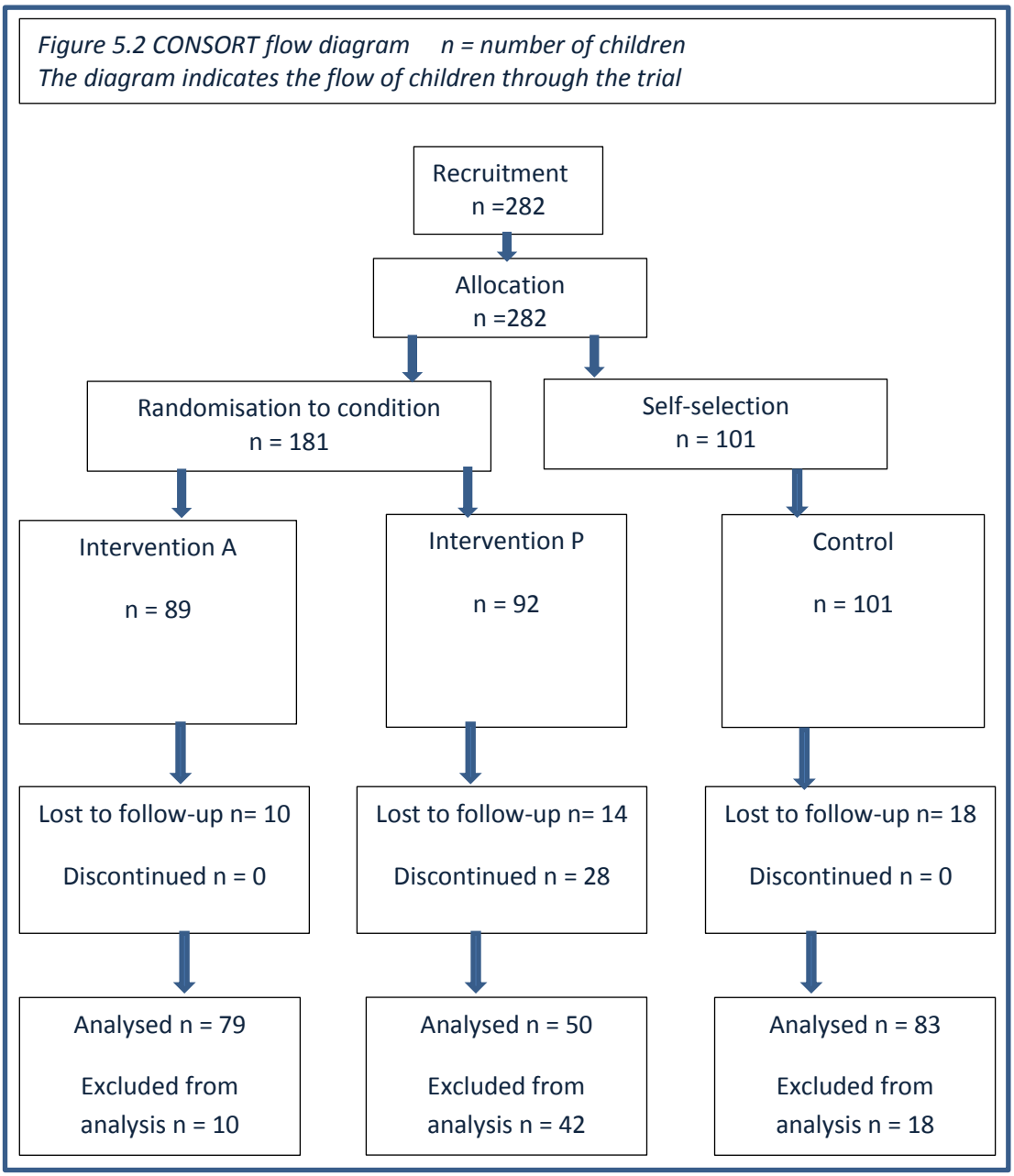
Schools were contacted in the first instance via email. Recruitment was from a largely middle-class rural area, but one which also included a city with a large number of urban schools. The selection criteria were single-form entry with a good, or outstanding, Ofsted designation. The schools were also selected from within the same education authority. In the first instance, twelve schools were approached on the basis of their geographical location; six urban and six rural in closest proximity to each other. Subsequently, schools were self-selecting; choosing to decline the invitation to participate for a variety of reasons, such as imminent Ofsted inspections, existing research projects, and pre-planned reading programmes of their own. The radius was increased until six urban and six rural schools (according to the original selection criteria) had agreed to participate.

All schools were mainstream state schools that would normally follow the National Curriculum. As all rural schools in the local authority had single form entry, urban schools with single form entry were selected to participate in the study where possible, as the style of teaching in large Reception cohorts is often very different from smaller groups. For example, teachers may group the children according to age within the cohort for different sessions (both of which would result in different learning experiences within a class) or some children may spend one part of every day with a different teacher from the others in their class. As the programme was designed to be delivered in small groups, schools that did not have sufficient

members of staff, i.e. a Teaching Assistant, were not recruited. A number of schools preferred to use a school governor to deliver the programme, and they received training alongside the Teaching Assistants in order to take on the role, or for when the Teaching Assistant was unavailable.

A diagram of the flow of children through the trial is shown below in Figure 5.2.

Figure 5.2 CONSORT flow diagram n = number of children
The diagram indicates the flow of children through the trial



The whole cohort of each Reception class from each school was recruited. There were eight Teaching Assistants participating. In addition, two schools asked governors to help with the programme delivery. The number of additional helpers varied from one to four, but they all received the same training. There were eight

participating teachers in the intervention, but only three chose to have the training, in order be a part of the programme delivery. In the other five schools, only the Teaching Assistants received the training. Class teachers were not expected to participate directly, unless they particularly chose to work with the Teaching Assistants (in two schools the class teacher took the lead). The Teaching Assistants, normally assigned to the Reception class by the head teacher, were the main facilitators of the intervention.

Training was given to small groups or individuals in their schools. This consisted mainly of an explanation of how to use the Manual and instructions on how to teach the games. Trainees were given the opportunity to handle the activities and play the games, and question the researcher. No information was given to participants regarding the main research questions, in order to reduce bias. Additional training was offered if requested. Concern was expressed by some at the outset regarding the amount of time that the programme would require, and teachers were therefore given flexibility regarding timescale of delivery. No other concerns were expressed at this time.

During Study 1, the resources for the first term were prepared before the start of implementation. Resources were not prepared in advance for the second or third term, in order to allow feedback from the first term to inform further design. Following mid-term discussion with a number of teachers, it was decided to slow the original pace (this meant not expecting all the activities to be completed in one week). All participating schools were then contacted and requested to slow their pace accordingly. It was anticipated that as the children became more familiar with the games that the pace might increase with time. Many of the children's manipulative skills appeared to be very immature in the early weeks, which meant that some of the activities took longer than anticipated.

5.2.3 Sample Size

The sample size (282 at pre-test) was calculated as having 80% power (probability) to detect a minimum effect size of 0.30 (equivalent to 4 months reading progress). This was calculated using a formula for calculating the Minimum Detectable Effect Size (MDES) recommended by the NFER (Hutchinson & Styles, 2010). Children from the Reception class of each school participated, up to a maximum of 30 per class. At post-test, having lost 70 participants for follow up, the numbers (212 at post-test)

were calculated to have 80% power to detect a minimum effect size of 0.35 (equivalent to 5 months progress).

5.2.4 Timeline

The timeline for Study 1, showing recruitment, training, delivery and assessments, is indicated in Table 5.1 below.

Table 5.1
Timeline for Study 1

Date	Timeline for Study 1	Assessment details
April 2013	Recruitment of schools	
July 2013	TA training for programme delivery	
September 2013	Assessments t1* – for all twelve schools	Assessments t 1* BPVS YARC: Letter Sound Knowledge Early Word Reading
October 2013	Delivery of autumn term resources to eight (intervention) schools	
January 2014	Delivery of spring term resources to eight (intervention) schools	
April 2014	Delivery of summer term resources to eight (intervention) schools Audio recordings Feedback questionnaires	
June 2014	Assessments t2 **– for all twelve schools Collection of other data (records of compliance; completion of activities; reading records)	Assessments t 2** BPVS YARC: Letter Sound Knowledge Early Word Reading Passage Reading Comprehension Intervention specific word (Nouns) recognition Teacher feedback: Questionnaires and discussion groups

*t1 = time 1 (pre-intervention) **t2 = time 2 (post-intervention)

5.2.5 Measures and Data collection

Measures used at pre-test, included the British Picture Vocabulary Scale (BPVS) (Dunn et al, 2009) to measure receptive vocabulary, and two measures from the York Assessment of Reading for Comprehension (YARC) (Snowling et al, 2009): Letter Sound Knowledge (LSK) and Early Word Reading (EWR). At post-test, these measures were repeated. In addition, the Passage Reading Comprehension test from the YARC was administered. This could not have been done at pre-test as the children were too young, and the data would have been subject to floor effects. At

post-test a non-standardised test of thirty-two common Nouns (as used in the two intervention arms) was also administered.

Allocation to intervention condition was kept blind to schools and the researcher, by using a third independent party to label the boxes of resources destined to go into the schools which were then made secure. It was not possible inadvertently to give the intervention to the Control group as the weebee Reading Programme did not exist outside the project. In order to reduce the Hawthorne effect (leading to a type I error: believing there to be an effect when there is none), teachers were told only that it was the method of teaching, i.e. the use of especially chosen/designed games, that was the focus of interest, not the different use of vocabulary, thereby blinding participants to the nature of the trial. Allocation of intervention was kept blind for the assessor at pre and post-test.

5.2.6 Process and Implementation Fidelity Procedures

As detailed in Chapter Four, an evaluation of the fidelity to implementation, was important to the final analysis of this kind of research, which does not have the tightly controlled conditions of a laboratory experiment. As part of the procedure to ensure fidelity to the process of the programme design, a meeting was arranged at the end of the first term to discuss progress, and to ensure that the overriding principles of the project were still being adhered to. Assessment of fidelity to process, in respect of unwanted behaviours (such as encouraging sounding out words) was through audio recording and transcript. For Study 1, this occurred only once, in the final term. Teachers were requested to record one whole class session and three small group sessions. A research assistant was trained to code the audios, in addition to the researcher, for reliability. Each mark awarded was agreed between the two researchers following discussion to establish inter-rater reliability of kappa = 1.0.

Fidelity to the structure of the programme, other than dosage, was assessed only at post-intervention. Information for implementation evaluation was collected through feedback questionnaires, informal discussions, individual child reading logs and session-completion records. The data from the session records and reading logs was used to calculate the intervention dosage. The session records detailed the number of activities actually completed and the number of children who completed them (see Appendix C). Data from questionnaires and informal discussions was only

collected at one time point, at the end of the trial. These questions related to the instruction Manual, the initial training, the resources and children's engagement levels (see Appendix D).

5.3 Pupil Characteristics (at post-test)

Within this sample there were a small number of children with English as an additional language (39 at pre-test and 20 at post-test). The distribution of children with English as an additional language and the distribution of gender across the three arms of the trial are detailed here as they were at post-test (those lost-to-follow-up had been excluded). Table 5.2 below shows the number of children with English as an additional language (EAL) for each arm of the trial.

Table 5.2
Distribution of EAL in Study 1

	Control		Intervention P		Intervention A	
EAL	Number	Percentage	Number	Percentage	Number	Percentage
Yes	9	11	6	12	5	7
No	74	89	44	88	74	93
Total	83	100	50	100	79	100

There was no significant association between the balance of EAL in each arm of the trial as calculated using Pearson's chi-square: χ^2 (2) = 1.46 (less than the critical value of 5.991 and therefore not significant at .05). This distribution of children with EAL is unlikely to have had a statistically significant impact on the outcomes.

Table 5.3 below shows the number of boys and girls for each arm of the trial.

Table 5.3
Distribution of gender in Study 1

	Control		Intervention P		Intervention A	
Gender	Number	Percentage	Number	Percentage	Number	Percentage
Male	37	45	22	44	45	57
Female	46	55	28	56	34	43
Total	*83*	*100*	*50*	*100*	*79*	*100*

There was no significant association between the balance of gender in each arm of the trial as calculated using Pearson's chi-square: χ^2 (2) = 3.15 (less than the critical value of 5.991 and therefore not significant at .05). Across all three arms of the trial, there were a total of 104 boys and 108 girls at post-test. The effect of gender is analysed for each of the separate tests detailed later in the Chapter.

5.4 Impact and Process Evaluation

The outcomes reported here include analyses of both standardised and non-standard tests, using independent samples t-tests and analysis of variance (ANOVA or ANCOVA), written feedback from teachers, and an evaluation of fidelity across the intervention arms using observations and written records.

5.4.1 Impact Evaluation

This section includes an analysis of the results from each of the three standardised assessment measures used at pre-test: British Picture Vocabulary Scale (BPVS) (Dunn et al, 2009) and from the York Assessment of Reading for Comprehension (YARC) (Snowling et al, 2009): Letter Sound Knowledge (LSK) and Early Word Reading (EWR). At post-test the five assessment measures used were: BPVS, and LSK, EWR, and Passage Reading Comprehension (PRC) from the YARC (standardised) and the list of intervention-related Nouns (non-standardised). Pre-test and post-test scores for three of the tests (British Picture Vocabulary Scale, Letter Sound Knowledge and Early Word Reading) and post-test scores for two tests (Passage Reading Comprehension and a list of intervention-related Nouns) are detailed in the table below.

As detailed in Chapter Four, planned contrasts were intended to evaluate measures in respect of Research Question 1 (Intervention A compared to Intervention P) and Research Question 2 (Intervention A compared to the Control condition and Intervention P compared to the Control condition). Significant differences between these comparisons were to be measured using Independent samples t-tests at pre and post-test. Effect sizes (using Cohen's *d*) were calculated and reported as difference (in months). Group means and standard deviations plus significances and effect sizes, between the Control group and each of the intervention arms, for each assessment measure are shown. In addition, a gender analysis is shown for each outcome measure.

5.4.1.1 Comparing Vocabulary (Interventions A and P)

Independent samples t-tests were first conducted on pre-test scores, detailed in Table 5.4 below. Scores were not significantly different from one another at pre-test (BPVS, t (127) = 1.35, p = .177; LSK, t (127) = 1.02, p = .308; EWR, t (127) = .198, p = .843) which means the two groups were comparable. Analysis of covariance was conducted at post-test, using pre-test scores as covariates. For PRC and Nouns, pre-test BPVS scores were used as these correlated best for these measures. A preliminary analysis evaluating the homogeneity of regression assumption indicated that the relationship between the covariate at pre-test and the dependent variable at post-test did not differ as a function of the independent variable for BPVS (F (1,125) = .465, p = .496), LSK (F (1,125) = .012, p = .913), PRC (using BPVS as covariate F (1,125) = .005, p = .941) and Nouns (using BPVS as covariate F (1,125) = .692, p = .407) and therefore ANCOVA could be run. However, EWR (F (1,125) = 8.40, p = .004) was significantly different and ANCOVA could not reliably be run. For EWR, ANOVA at pre-test showed no significant effect of condition F (1,127) = 0.39, p = .843, or at post-test F (1,127) = .264, p = .608.

There was a significant effect of the covariate for BPVS (F (1,126) = 138.067, p < .001, η_p^2 = .523), LSK (F (1,126) = 16.56, p < .001, η_p^2 =.116), PRC (F (1,126) = 37.44, p =<.001, η_p^2 = .229) and Nouns (F (1,126) = 24.1, p < .001, η_p^2 = .161) but no significant effect of the condition after controlling for the covariate for these measures (BPVS, F (1,126) = .003, p = .957, η_p^2 < .001; LSK, F (1,126) = 1.41, p = .237, η_p^2 = .011; PRC, F (1,126) = .519, p = .472, η_p^2 = .004; Nouns, F (1,126) = 1.04, p = .310, η_p^2 = .008). There were no statistically significant differences on these measures (BPVS, LSK, PRC, Nouns) between these groups (A and P). The relative between groups difference in pre-post effect size (see Table 5.5) nevertheless indicates a positive effect for LSK for Intervention A compared to Intervention P. Using standard scores for LSK (to control for age) pre-post effect size difference was d = 0.30, equating to 4 months difference.

5.4.1.2 Comparing Teaching Methods

Comparing Intervention A and Control

Independent samples t-tests were first conducted on pre-test scores. Scores were not significantly different from one another at pre-test (BPVS, t (160) = 1.85, p =

.064; LSK, t (160) = .017, p = .987; EWR, t (160) = 1.32, p = .192) which means the two groups were comparable. Analysis evaluating the homogeneity of regression assumption indicated that the covariate at pre-test and the dependent variable at post-test did not differ as a function of the independent variable for BPVS (F (1,158) = .295, p = .688), EWR (F (1,158) = 1.77, p = .185), PRC (using BPVS as covariate F (1,158) = .238, p = .626) and Nouns (using BPVS as covariate F (1,158) = .025, p = .874) and therefore ANCOVA could be run. However, LSK (F (1,158) = 4.74, p = .031) was significantly different and ANCOVA could not reliably be run. For LSK, ANOVA at pre-test showed no significant effect of condition F (1,160) = 0.00, p = .843, but did show a significant effect at post-test F (1,160) = 4.30, p = .040.

There was a significant effect of the covariate for BPVS (F (1,159) = 295.2, p < .001, η_p^2 = .650), EWR (F (1, 159) = 27.36, p < .001 η_p^2 = .147) PRC (F (1,159) = 62.25, p < .001, η_p^2 = .281) and Nouns (F (1,159) = 36.90, p < .001, η_p^2 = .188). There was no significant effect of the condition after controlling for the covariate for BPVS (F (1,159) = .008, p = .930, η_p^2 < .001), but there was for EWR (F (1,159) = 6.08, p = .015, η_p^2 = .037), PRC (F (1,159) = 6.91, p = .009, η_p^2 = .042) and Nouns (F (1,159) = 8.23, p = .005, η_p^2 = .049).

There were statistically significant differences between these groups for EWR, PRC and Nouns after controlling for the covariate. In addition, the relative between groups difference in pre-to-post effect sizes (Table 5.5) show positive effects for Intervention A for LSK and EWR compared to the Control. Using standard scores (to control for age) these differences were LSK: d = 0.22, equating to 3 months difference and EWR: d = 0.31, equating to 4 months difference.

Comparing Intervention P and Control

Independent samples t-tests were first conducted on pre-test scores. Scores were not significantly different from one another at pre-test for LSK and EWR but were for BPVS (BPVS, t (131) = 3.16, p = .002; LSK, t (131) = 1.09, p = .275; EWR, t (131) = 1.23, p = .220), meaning that the two groups were comparable for LSK and EWR but not for BPVS. Analysis evaluating the homogeneity of regression assumption indicated that the covariate at pre-test and the dependent variable at post-test did not differ as a function of the independent variable for any of the measures (BPVS, F (1,129) = 1.61, p = .205; LSK (F (1,129) = 2.63, p = .107;

Table 5.4

Group means, standard deviations, significance and effect sizes at pre and post-test

	Pre-test scores						Post-test scores					
	A	P	C	A v P	A v C	P v C	A	P	C	A v P	A v C	P v C
	Raw (SD)	Raw (SD)	Raw (SD)	Sig (*p*)	Sig (*p*)	Sig (*p*)	Raw (SD)	Raw (SD)	Raw (SD)	Sig (*p*)	Sig (*p*)	Sig (*p*)
	Standard (SD)	*Standard (SD)*	*Standard (SD)*	(*d*)	(*d*)	(*d*)	*Standard (SD)*	*Standard (SD)*	*Standard (SD)*	(*d*)	(*d*)	(*d*)
BPVS	64.86	68.12	60.02	.177	.064	.002	77.91	80.08	74.16	.306	.123	.010
	(14.47)	(11.1)	(18.34)	(0.25	(0.29)	(0.53)	(14.32)	(9.64)	(16.28)	(0.17)	(0.24)	(0.44)
	104.86	*107.40*	*100.43*	*.196*	*.032*	*.002*	*101.13*	*102.70*	*97.79*	*.422*	*.070*	.012
	(12.00)	*(8.56)*	*(13.90)*	*(0.02)*	*(0.34)*	*(0.60)*	*(11.62)*	*(9.10)*	*(11.71)*	*(0.15)*	*(0.28)*	*(0.46)*
LSK	4.94	5.72	4.93	.308	.987	.275	29.86	29.28	28.45	.439	.038	.358
	(4.07)	(4.31)	(3.29)	(0.18	(0.00)	(0.20)	(3.47)	(5.01)	(4.95)	(0.13)	(0.32)	(0.16)
	105.01	*106.70*	*103.84*	*.546*	*.630*	*.278*	*118.49*	*116.38*	*115.27*	*.328*	*.112*	*.666*
	(16.05)	*(14.33)*	*(14.82)*	*(0.11)*	*(0.07)*	*(0.19)*	*(10.63)*	*(13.71)*	*(14.57)*	*(0.17)*	*(0.25)*	*(0.07)*
EWR	1.02	.92	.469	.843	.192	.220	15.35	14.58	11.62	.608	.006	.056
	(3.33)	(2.17)	(1.78)	(0.03)	(0.20)	(0.22)	(8.25)	(8.45)	(8.62)	(0.09)	(0.44)	(0.34)
	97.86	*97.30*	*95.40*	*.790*	*.142*	*.293*	*113.32*	*112.48*	*107.36*	*.752*	*.007*	*0.56*
	(11.84)	*(11.21)*	*(9.20)*	*(0.04)*	*(0.23)*	*(0.18)*	*(12.41)*	*(14.65)*	*(14.94)*	*(0.06)*	*(0.43)*	*(0.34)*
PRC							18.06	18.08	12.93	.992	.001	.005
							(9.79)	(9.60)	(10.27)	(0.00)	(0.51)	(0.51)
							99.06	*98.66*	*96.07*	*.763*	*.008*	*.042*
							(7.31)	*(7.47)*	*(6.76)*	*(0.05)*	*(0.42)*	*(0.36)*
Nouns							16.87	16.16	11.56	.690	.001	.008
							(10.15)	(9.37)	(9.55)	(0.07)	(0.53)	(0.48)

Intervention A: n = 79; Intervention P: n = 50; Control: n = 83

EWR, F (1,129) = 1.36, p = .244; PRC, (using BPVS as covariate) F (1,129) = .170, p = .681; and Nouns (using BPVS as covariate) F (1,129) = 1.14, p = .287) and therefore ANCOVA could reliably be run.

There was a significant effect of the covariate for all measures (BPVS, F (1,130) = 306.37, p < .001, η_p^2 =.702; LSK, F (1,130) = 23.12, p < .001, η_p^2 = .151; EWR, F (1,130) = 40.58, p < .001. η_p^2 = .238; PRC, F (1,130) = 44.33, p < .001, η_p^2 = .254; Nouns, F (1,130) = 36.10, p < .001, η_p^2 = .218). There was no significant effect of the condition after controlling for the covariate for any of the measures (BPVS; F (1,130) = .003, p = .953, η_p^2 < .001; LSK, F (1,130) = .251, p = .617, η_p^2 = .002; EWR, F (1,130) = 2.15, p = .145, η_p^2 = .016; PRC, F (1,130) = 2.60, p = .109, η_p^2 = .020 and Nouns, F (1,130) = 2.31, p = .131, η_p^2 = .017). There were no statistically significant differences on these measures between these groups. The relative between groups differences in pre-post effect sizes (see Table 5.5) show positive effects for Intervention P for BPVS and EWR but not for LSK. Using standard scores, effect sizes were BPVS: d = 0.33 (4 months difference), LSK: d = 0.08 (1 month) and EWR d = 0.20 (3 months difference).

Table 5.5
Pre-test to post-test effect sizes (Cohen's d)

	BPVS Raw	*BPVS Standard*	LSK Raw	*LSK Standard*	EWR Raw	*EWR Standard*
Intervention A	0.90	*0.31*	6.58	*0.99*	2.27	*1.27*
Intervention P	1.15	*0.53*	5.04	*0.69*	2.21	*1.16*
Difference	0.25	*0.22*	1.54	*0.30*	0.06	*0.11*
Intervention A	0.90	*0.31*	6.58	*0.99*	2.27	*1.27*
Control	0.81	*0.20*	5.59	*0.77*	1.79	*0.96*
Difference	0.09	*0.11*	0.99	*0.22*	0.48	*0.31*
Intervention P	1.15	*0.53*	5.04	*0.69*	2.21	*1.16*
Control	0.81	*0.20*	5.59	*0.77*	1.79	*0.96*
Difference	0.34	*0.33*	0.55	*0.08*	0.42	*0.20*

5.4.1.3 Gender Analyses

Contrasts presented here focused on the impact of the three trial arms on gender rather than comparing boys and girls directly in each arm of the trial. Therefore contrasts were made firstly between girls in Interventions A and P, girls in Intervention P and the Control, and girls in Intervention A and the Control, and secondly between boys in Interventions A and P, boys in Intervention P and the

Control, and boys in Intervention A and the Control condition. Analysis of covariance was conducted at post-test, using pre-test scores as covariates (using BPVS as covariates for PRC and Nouns). Means, standard deviations, significance levels between genders at pre-test and post-test, and pre-to-post effect sizes for BPVS, LSK and EWR are shown in Table 5.6 below. Means, standard deviations and significance levels between genders for PRC and Nouns are shown in Table 5.7.

Comparing Vocabulary: Intervention A and P (Girls)

A preliminary analysis evaluating the homogeneity of regression assumption indicated that ANCOVA could reliably be run for BPVS (F (1,58) = .855, p = .359), LSK (F (1,58) = .100, p = .753), PRC (F (1,58) = .001, p = .971) and Nouns (F (1,58) = .722, p = .399) , but not for EWR (F (1,58) = 5.32, p = .025). For EWR, ANOVA at pre-test showed no significant effect of condition F (1,60) = .221, p = .640, or at post-test F (1,60) = .863, p = .357. There was no significant effect of condition after controlling for the covariate for any of the measures (BPVS, F (1,59) = .389, p = .535, η_p^2 = .007; LSK, F (1,59) = .417, p = .521, η_p^2 = .007; PRC, F (1,59) = .001, p = .979, η_p^2 = .000; Nouns, F (1,59) = .569, p = .454, η_p^2 = .010). The difference in gains in pre-to-post effect sizes for LSK and EWR (d = 1.88 for LSK, d = 0.58 for EWR) indicate a positive effect from Intervention A compared to Intervention P for girls on these two measures.

Comparing Vocabulary: Intervention A and P (Boys)

Preliminary analysis indicated that ANCOVA could reliably be run for all measures (BPVS, F (1,63) = .153, p = .697; LSK, F (1,63) = .190, p = .665; EWR, F (1,63) = 1.72, p = .194; PRC, F (1,63) = .351, p = .555; Nouns, F (1,63) = .421, p = .519). There was no significant effect of condition after controlling for the covariate for any of the measures (BPVS, F (1,64) = .003, p = .953, η_p^2 = .000; LSK, F (1,64) = 1.20, p = .276, η_p^2 = .018; EWR, F (1,64) = .015, p = .902, η_p^2 = .000; PRC, F (1,64) = 1.94, p = .168, η_p^2 = .029; Nouns, F (1,64) = .878, p = .352, η_p^2 = .014). The difference in gains in pre-to-post effect sizes for BPVS and EWR (d = 0.31 for BPVS, d = 0.20 for EWR) indicate a small positive effect from Intervention P compared to Intervention A for boys on these two measures. The difference in pre-to-post effect size for LSK (d = 1.75) indicates a positive effect for Intervention A compared to Intervention P for boys on this measure.

Table 5.6

Means, standard deviations and significance at pre and post-test, and pre-to-post-test effect sizes for gender

Test	Condition	Gender	Raw Time 1 (SD)	*Standard Time 1 (SD)*	Raw Time 2 (SD)	*Standard Time 2 (SD)*	Sig. (*p*) Time 1	Sig. (*p*) Time 2	Effect size (Cohen's *d*) Pre-post
BPVS	Control	M = 37	57.02(17.27)	*97.62(12.97)*	72.45(14.98)	*96.13(10.67)*	.183	.395	0.95
		F = 46	62.43(18.99)	*102.69(14.35)*	75.54(17.30)	*99.13(12.43)*			0.72
	Intervention A	M = 45	62.93(13.12)	*103.73(11.19)*	75.86(15.00)	*100.00(12.46)*	.175	.146	0.91
		F = 34	67.41(15.92)	*106.35(13.01)*	80.61(13.10)	*102.64(10.39)*			0.90
	Intervention P	M = 22	69.50(8.53)	*109.00(7.15)*	81.13(10.32)	*103.59(10.26)*	.444	.498	1.22
		F = 28	67.03(12.90)	*106.14(9.46)*	79.25(9.17)	*102.00(8.20)*			1.09
LSK	Control	M = 37	4.48(3.21)	*102.78(14.68)*	28.51(5.00)	*114.78(14.90)*	.264	.928	5.71
		F = 46	5.30(3.35)	*104.69(15.03)*	28.41(4.97)	*115.67(14.44)*			5.45
	Intervention A	M = 45	3.97(3.74)	*102.62(16.13)*	29.53(4.16)	*118.06(11.84)*	.014	.338	6.46
		F = 34	6.23(4.18)	*108.17(15.60)*	30.29(2.24)	*119.05(8.92)*			7.17
	Intervention P	M = 22	5.22(3.89)	*105.36(14.31)*	28.68(5.87)	*114.50(14.59)*	.480	.460	4.71
		F = 28	6.10(4.66)	*107.75(14.52)*	29.75(4.26)	*117.85(13.05)*			5.29
EWR	Control	M = 37	.62(2.33)	*95.24(9.47)*	12.32(9.73)	*107.43(16.78)*	.183	.395	1.65
		F = 46	.34(1.17)	*95.54(9.08)*	11.06(7.68)	*107.30(13.47)*			1.95
	Intervention A	M = 45	1.06(3.8)	*97.66(11.05)*	13.57(8.16)	*111.48(13.00)*	.900	.027	1.96
		F = 34	.97(2.65)	*98.11(12.97)*	17.70(7.88)	*115.76(11.31)*			2.84
	Intervention P	M = 22	.45(1.40)	*94.95(9.18)*	13.09(8.12)	*110.59(15.75)*	.155	.274	2.16
		F = 28	1.28(2.59)	*99.14(12.43)*	15.75(8.67)	*113.96(13.83)*			2.26

Table 5.7
Means, standard deciation and significance of gender at pos-test

Test	Condition	Gender	Raw Time 2 (SD)	*Standard Time 2 (SD)*	Sig. (*p*) Time 2
PRC	Control	M = 37	13.29(10.07)	*96.08(6.98)*	.778
		F = 46	12.65(10.52)	*96.06(6.67)*	
	Intervention A	M = 45	16.37(9.07)	*98.22(7.04)*	.078
		F = 34	20.29(10.39)	*100.17(7.62)*	
	Intervention P	M = 22	15.36(9.85)	*96.00(6.27)*	.076
		F = 28	20.21(9.00)	*100.75(7.78)*	
Nouns	Control	M = 37	11.59(10.21)		.981
		F = 46	11.54(9.09)		
	Intervention A	M = 45	14.86(10.02)		.043
		F = 34	19.52(9.85)		
	Intervention P	M = 22	14.04(9.68)		.159
		F = 28	17.82(8.93)		

Comparing Teaching Methods: Intervention P and Control (Girls)

Preliminary analysis indicated that ANCOVA could reliably be run for all measures except BPVS (LSK, $F(1,70) = 3.93$, $p = .051$; EWR, $F(1,70) = .476$, $p = .493$; PRC, $F(1,70) = .031$, $p = .862$; Nouns, $F(1,70) = 1.04$, $p = .309$). For BPVS, ANOVA at pre-test showed no significant effect of condition $F(1,72) = 1.27$, $p = .262$, or at post-test $F(1,72) = 1.09$, $p = .299$. There was no significant effect of condition after controlling for the covariate for LSK: $F(1,71) = .801$, $p = .374$, $\eta_p^2 = .011$ or EWR: $F(1,71) = 2.13$, $p = .149$, $\eta_p^2 = .029$. However, there was a significant effect of condition for PRC: $F(1,71) = 9.19$, $p = .003$, $\eta_p^2 = .115$ and Nouns: $F(1,71) = 7.34$, $p = .008$, $\eta_p^2 = .094$. The difference in gains in pre-to-post effect sizes for BPVS and EWR ($d = 0.37$ for BPVS, $d = 0.31$ for EWR) indicate a positive effect from Intervention P compared to the Control condition for girls on these two measures.

Comparing Teaching Methods: Intervention P and Control (Boys)

Preliminary analysis indicated that ANCOVA could reliably be run for all measures (BPVS, $F(1,55) = .545$, $p = .464$; LSK, $F(1,55) = .145$, $p = .705$; EWR, $F(1,55) = .361$, $p = .551$; PRC, $F(1,55) = .477$, $p = .493$; Nouns, $F(1,55) = .398$, $p = .531$). There was no significant effect of condition after controlling for the covariate for any

of the measures (BPVS, F (1,56) = .020, p = .887, η_p^2 = .000; LSK, F (1,56) = .032, p = .860, η_p^2 = .001; EWR, F (1,56) = .229, p = .634, η_p^2 = .004; PRC, F (1,56) = .409, p = .525, η_p^2 = .007; Nouns, F (1,56) = .047, p = .829, η_p^2 = .001). The difference in gains in pre-to-post effect sizes for LSK (d = 1.00) indicate a positive effect of the Control condition compared to Intervention P for boys on this measure. The difference in gains in pre-to-post effect sizes for BPVS and EWR (d = 0.27 for BPVS, d = 0.51 for EWR) indicate a positive effect from Intervention P compared to the Control for boys on these two measures.

Comparing Teaching Methods: Intervention A and Control (Girls)

Preliminary analysis indicated that ANCOVA could reliably be run for all measures except LSK (BPVS, F (1,76) = 2.08, p = .153; EWR, F (1,76) = .677, p = .424; PRC, F (1,76) = .026, p = .873; Nouns, F (1,76) = .000, p = .986). For LSK, ANOVA at pre-test showed no significant effect of condition F (1,78) = 1.21, p = .273, but did at post-test F (1,78) = 4.20, p = .044. There was no significant effect of condition after controlling for the covariate for BPVS: F (1,77) = .500, p = .482, η_p^2 = .006. However, there was a significant effect of condition for EWR: F (1,77) = 11.91, p = .001, η_p^2 = .134; PRC: F (1,77) = 9.22, p = .003, η_p^2 = .107 and Nouns: F (1,77) = 12.85, p = .001, η_p^2 = .143. The difference in gains in pre-to-post effect sizes for LSK and EWR (d = 1.72 for LSK, d = 0.89 for EWR) indicate a positive effect from Intervention A compared to the Control condition for girls on these two measures.

Comparing Teaching Methods: Intervention A and Control (Boys)

Preliminary analysis indicated that ANCOVA could reliably be run for all measures (BPVS, F (1,78) = .134, p = .716; LSK, F (1,78) = 1.02, p = .315; EWR, F (1,78) = 1.00, p = .318; PRC, F (1,78) = .051, p = .821; Nouns, F (1,78) = .006, p = .939). There was no significant effect of condition after controlling for the covariate for any of the measures (BPVS, F (1,79) = .144, p = .706, η_p^2 = .002; LSK, F (1,79) = 1.68, p = .198, η_p^2 = .021; EWR, F (1,79) = .155, p = .615, η_p^2 = .002; PRC, F (1,79) = .498, p = .483, η_p^2 = .008; Nouns, F (1,79) = .827, p = .366, η_p^2 = .010). The difference in gains in pre-to-post effect sizes for LSK and EWR (d = 0.75 for LSK, d = 0.31 for EWR) indicate a positive effect from Intervention A compared to the Control condition for boys on these two measures.

Comparing Intervention A and Intervention P, results suggest that the non-decodable vocabulary may have had a small positive effect for girls but showed little effect for boys in this study. Comparing Intervention P with the Control condition, results suggest that mixed teaching methods had a positive effect for girls and to a lesser extent for boys. Comparing Intervention A with the control condition, results suggest that mixed teaching methods in addition to non-decodable vocabulary had a greater positive effect for girls than mixed teaching methods alone and also to a lesser extent for boys. Results suggest that both boys and girls made greater gains in the intervention conditions compared to the Control condition.

5.4.2 Process Evaluation

The outcomes reported here include feedback from teachers and an evaluation of fidelity across the intervention arms using observations and written records.

5.4.2.1 Feedback

Feedback regarding the Manual was mixed; the majority of teachers reporting that they found the Manual useful, but one school reporting that it was only sometimes useful. The initial training was reported as useful by most but only satisfactory by one school. Questions regarding ease of use of the resources received a mixed response, varying from satisfactory to very easy. All the schools reported that the children enjoyed the games and activities. One school reported that they had received positive feedback from parents regarding the children's enthusiasm for the programme activities. Most of the schools reported that the children enjoyed reading the books, although one school reported only some enjoyment. This school had a very young group, and the teacher reported that for some of these children the books were over long. All schools referred to timetabling constraints, which had resulted in reduced compliance to the original design, such as not ensuring that all the children had completed all the activities, and not completing all the activities before the books were read. Nevertheless, three of the intervention schools reported their intention to run the programme again the following year.

5.4.2.2 Fidelity

At post-test, when data was collected for analysis from participating schools, much of the implementation data was missing; some schools recorded their session completion records in detail, some partially, and others not at all. In addition, there

were significant levels of attrition and non-compliance in the trial, as illustrated in Figure 5.3 below (n = number of schools). Of the twelve schools for which data was collected at pre-test, only data from eleven schools could be collected at post-test; one school (Intervention P) failed to implement the intervention due to loss of staff. One school withdrew following an Ofsted inspection after one term (Intervention A). A further school withdrew after one term, due to staff losses (Intervention A). There was reduced implementation at two schools as a result of staff illnesses.

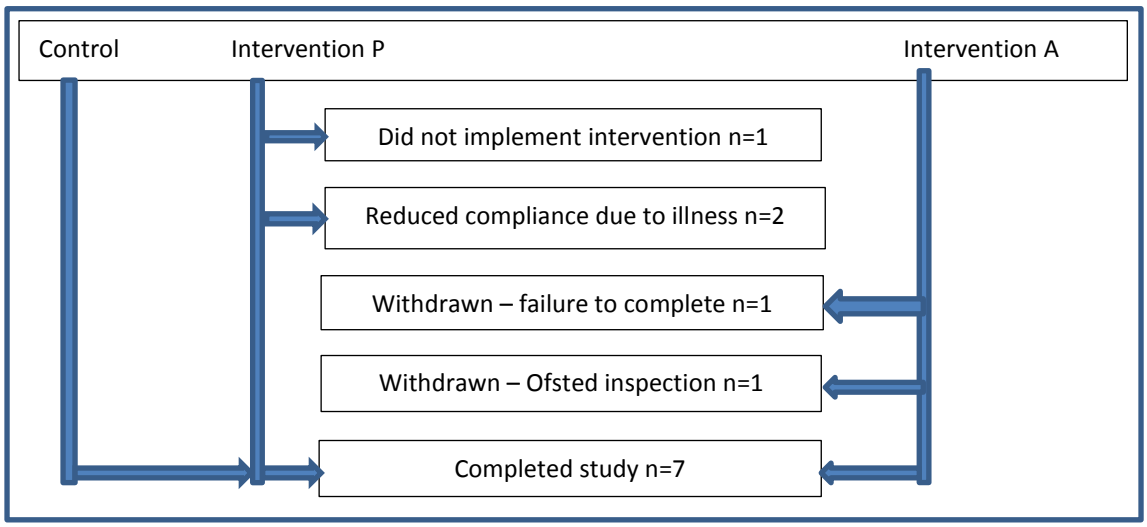


Figure 5.3 Diagram showing factors affecting compliance in Study 1

All schools were provided with twelve sets of books and materials; however none of the schools completed all of these. For the three schools that completed the trial, there was mixed compliance to the original design. One school (Intervention P) completed eight books; one (Intervention A) completed seven books and one (Intervention P) completed five books. Data from the other schools was not provided although frequent requests were made for this, resulting in a lack of data for analysis. Two of the schools who continued with the trial reported using only the first four books although they provided no data to support this.

Of the three (out of eight) intervention schools that provided the detailed session and reading records (see Appendix C), analysis showed a clear link between the number of activities and games used for each book, and the number of words children were subsequently able to read with ease. The mean number of words read for each book, and the number of completed sessions for each book, are shown in Table 5.8 below. The Table shows that for two schools (04 and 01) there were an incomplete number of sessions for books 1 and 2, and a clear difference in mean

scores, between these books and later books, which had completed sessions. By contrast, school 03, which completed all the sessions for all the books, shows a steady increase in scores. School 01 shows higher mean scores from book 2 onwards, although they did not complete as many books. The total percentage of mean words, for the number of books completed (twenty words per book), shows the highest score for school 01, which, although it only completed five books, spent more time in fully completing all the sessions.

Table 5.8
Mean No. of words per book (per school) Study 1

School Code: 04 N = 14 Condition = P			School Code:01 N = 15 Condition = P			School Code:03 N = 15 Condition = A		
Book No	Number of completed sessions (out of 7)	Mean number of words read	Book No	Number of completed sessions (out of 7)	Mean number of words read	Book No	Number of completed sessions (out of 7)	Mean number of words read
1	3	3.57	1	4	4.73	1	7	7.68
2	4	1.85	2	6	8.73	2	7	7.06
3	7	3.07	3	7	11.87	3	7	8.18
4	7	8.42	4	7	16.2	4	7 *	6.92
5	7	12.78	5	7	16.3	5	7	7.46
6	7	8.21				6	7	9.93
7	7	9.14				7	7	12.11
8	7	8.21						
Total mean No. 55.25 Percentage of words to number of books 34.53%			Total mean No. 57.83 Percentage of words to number of books 57.83%			Total mean No. 59.34 Percentage of words to number of books 42.38%		

* High level of absent children

Dosage for the individual schools was calculated on the basis of numbers of sessions completed (58% and 37.5% for Intervention P; 58.3% for Intervention A). The scores, shown in the Table above, suggest that higher implementation fidelity resulted in higher scores in word recognition for the three schools who reported their data. It is worth noting that in these three schools, the intervention was delivered by experienced class teachers. Schools, who provided session records, were working at a rate of approximately three books per term. Before withdrawing,

schools 02 and 07 (both Intervention A), were working at the same rate and are thus likely to have completed up to 25% of the programme. The other two schools, 05 and 06, did not provide any indication of rate of implementation.The dated records show that the teacher in school 01 spent about twice as much time on all the activities, allowing the children time to consolidate their learning, which may have contributed to the higher scores. School 03, although it only completed seven books and was using the non-decodable vocabulary had the highest total mean score and showed a trend of increasing scores.

Audio recordings of teaching sessions, intended to assess process fidelity, were made by the teachers in the same three schools that completed session and reading records. The other two schools (05 and 06) who finished the trial failed to make the audio recordings due to staff illness. Although teachers were given specific instructions, regarding which sessions to record, there was variability in compliance, particularly in the length of time/number of activities. Specific instructions were provided, regarding the exact sessions to record, but only two schools complied. In addition, their sessions lasted different times (between 20 and 40 minutes), depending on the size of groups of children, and time-tabling constraints. The recordings were analysed and coded for both desired and undesired behaviour (see Appendix K). The behaviours were given a rating and the sum of the total scores for each teacher was calculated as a percentage for each of desired and undesired behaviours as shown in Table 5.9 below. Using this calculation as an estimate for process fidelity, the importance of desired behaviours seems apparent, from the scores in Table 5.8 above (school 01 having the highest percentage of mean words learned per book). In addition, school 01 scored 0% for undesired behaviours.

Table 5.9
Observation results for Study 1

School Code	Percentage of desired behaviours	Percentage of undesired behaviours
01	92	0
03	62.96	53.33
04	59.25	46.66

Higher word reading outcomes (as recorded by the class teachers) were associated with higher implementation, and higher process, fidelity. Overall fidelity to the programme was poor, with little or no data available from four of the seven schools.

5.5 Summary and Discussion of Study 1 Outcomes

This section includes a discussion of the outcomes from the standardised and non-standard tests in respect of the two main research questions, firstly, regarding the effect of using non-phonically decodable vocabulary, and secondly, regarding the use of mixed teaching methods. This is followed by a summary of the emergent group characteristics, the effects of gender, and the observed fidelity of the two intervention arms of the trial to the programme protocol.

5.5.1 Limitations of the Study Design

One potential weakness in the design of this study was the risk of variation in implementation of the intervention by different teachers. Although a Manual was provided, and fidelity was observed at a number of time points, the level of engagement with the materials by the teachers was not monitored throughout the trial, due to time and logistic constraints for both schools and the researcher.

Although assessment of children was on an individual basis, teaching was effectively clustered according to teachers. However, randomisation should have ameliorated this effect by removing the risk of selection bias. The assignment of whole classes to a condition, rather than having split clusters, also increased the risk of imbalance arising from attrition. In addition, it increased the risk that any observed effect could be considered to have resulted from the teaching style of a particular teacher rather than the methods or materials being used (Darrow, 2013). The use of a detailed Manual, training, and the provisions of all teaching materials was intended to counteract this issue.

Most teachers felt that the timescales indicated in the Manual were too optimistic; these could have been discussed and planned with practitioners in advance. There was no built in provision for children who needed to progress at a different pace, or needed more repetition, leaving this aspect of delivery open to variation. Training was only given at one time point, and was mainly focused on the use of the materials. There was less emphasis on the kind of language to be used (such as modelling correct pronunciations, emphasising initial letters or salient

features, and encouraging group co-operation), and the importance of encouraging sight-word recognition, as opposed to sounding out words. This led to some teachers not adhering fully to the programme protocol. There was a mix of class teacher, teaching assistants, governors and other helpers amongst the programme deliverers. This resulted in a wide range of experience and training between teachers across the study.

5.5.2 Research question 1: *Are there measurable differences between vocabulary that is within a child's existing decoding ability (Intervention P) and vocabulary which is not so constrained (Intervention A)?*

There were no statistically significant differences between Intervention A and Intervention P on any of the tests. However, there were observable trends. For the BPVS, although the difference between the two intervention conditions was not significant, there was a smaller difference at post-test suggesting a trend towards convergence. This is most likely to be attributable to maturation and unlikely to result from any effect of the intervention. There were no significant differences between the two intervention conditions as measured by LSK either pre or post-test. However, there was a trend towards higher scores and a higher effect size for Intervention A. No differences were observed for the non-standard test of Nouns. Taken together, the test results are not significant, but do indicate a trend towards higher outcomes for Intervention A. This would suggest that there may have been a marginal positive effect on measures of word reading from using non-phonically decodable vocabulary in reading texts for children in this study.

5.5.3 Research Question 2: *Are there measurable differences when comparing a synthetic phonics only approach with a mixed teaching approach?*

When comparing Intervention P with the Control condition, there were no statistically significant differences for any of the measures. The Early Word Reading measure, which was close to a significant difference between groups at post-test, and the pre-post effect sizes for the British Picture Vocabulary Scale, indicate a trend of advantage for the mixed teaching methods with phonically decodable vocabulary on oral vocabulary, word decoding and passage reading comprehension.

When comparing Intervention A with the Control condition, there were statistically significant differences for the Early Word Reading, Passage Reading Comprehension

and Nouns measures. These results suggest that the mixed teaching methods in combination with a non-decodable vocabulary had a significant positive impact on these measures of reading comprehension and word reading.

5.5.4 Group characteristics

For all three arms of the trial, the percentage of children with English as an additional language was similar and small, and showed no significant difference. The percentage of boys to girls was similar in the Control condition and Intervention P, with both having a higher number of girls. In Intervention A, the percentage of boys was higher, which may have contributed to the effect of gender across all measures.

Differences in raw scores between boys and girls at pre-test and post-test indicate a small narrowing of the gender gap for the British Picture Vocabulary Scale in the Control condition and Intervention P and little change for Intervention A. For Letter Sound Knowledge, raw scores at pre and post-test indicate a narrowing of the gender gap in the Control condition and Intervention A, but a widening in Intervention P. For Early Word Reading the gender gap widened in all three conditions.

Mean scores for the Passage Reading Comprehension test suggest there may have been a positive effect from the use of non-decodable vocabulary for boys. Girls in the two intervention conditions scored higher than in the Control condition for Passage Reading Comprehension but with similar scores (see Table 5.7).

For boys, there appears to have been a small positive effect, associated with the use of non-decodable vocabulary and the mixed teaching methods in Intervention A, for Letter Sound Knowledge and Early Word Reading, and from the mixed teaching methods with phonically decodable vocabulary in Intervention P for the British Picture Vocabulary Scale, Letter Sound Knowledge and Early Word Reading. For girls, there was a positive effect observed associated with the use of non-decodable vocabulary in addition to mixed teaching methods across all measures.

5.5.5 Fidelity

Data for assessing both implementation and process fidelity was only provided by three of the participating schools, two of which received Intervention P and the third received Intervention A. Given that one of the schools allocated to Intervention P failed to begin the trial, the two schools represented 66% of the total data analysed from that arm of the trial, whereas the third school only represented 25 % of the total from Intervention A. Imputed scores for the missing schools were calculated based on verbal data (schools claiming to have covered four books) and time spent using the intervention (three books in the first term) to give an estimate of dosage in each of the intervention conditions. This was estimated at 41.6% for Intervention P and 34.5% for Intervention A. Taken together with the results from the audio analyses (shown in Table 5.8) there is evidence that there may have been higher fidelity to the programme in Intervention P, possibly due to the use of decodable vocabulary and its conformity to the National Curriculum guidelines. However, there was insufficient data to draw any firm conclusions.

The high levels of attrition were related in part to perceived time constraints in the classroom. Most teachers commented on the amount of time required to implement the programme, and for those who were under pressure from imminent Ofsted inspections, this was deemed to be unacceptable. Schools that made use of teaching assistants, rather than experienced class teachers, showed lower compliance to programme protocol, as well as reduced implementation. In the schools where class teachers delivered the programme themselves, outcomes were higher in terms of scores on post-intervention tests. School 01, which showed exceptional fidelity to process, also demonstrated better speed of delivery of the games, which had been part of the original training, but was not apparent in the observations in other schools.

5.6 Conclusions

The specific aims of this study were: to evaluate the use of mixed teaching methods for beginning reading; to compare the use of phonically decodable vocabulary with non-phonically decodable; to assess the impact of the intervention on receptive vocabulary, word recognition and comprehension; to evaluate the programme in a classroom environment; and to evaluate the use of Teaching Assistants for facilitation of the Reading Programme.

Part of the evaluation of the impact of the intervention, included an analysis of aspects of the study design and implementation. A number of limitations of the design emerged. A particular problem was caused by attrition, which resulted in an imbalance of numbers between the two intervention arms. This could have been avoided by using a split cluster design, and as a consequence, this method was used for Study 2 (see Chapter Six). There was poor compliance with both implementation and process, which may have been attributable in part to insufficient training and support, particularly in the area of time management. Specific training in the speedy implementation of the activities may have reduced this problem, as well as increasing the number of new words that children would have encountered during the intervention.

The use of Teaching Assistants had the advantage of increasing the likelihood of initial recruitment of schools, however, the evidence from observations and feedback demonstrated that they were less skilled in techniques for enhancing vocabulary development, encouraging group activity, or following the intervention protocol. By choosing to blind participants as to the actual purpose of the trial, as well as to condition, in order to reduce the risk of a type I error, there was the possibility that this led to misunderstandings of the intentions of the programme, and therefore non-compliance to programme protocol.

Attrition and poor fidelity can to some extent be attributable to failures in design, in terms of insufficient on-going support and training. Discussion and feedback was limited to one session at the end of the first term, and a second at post-test. This was insufficient, and early observations would have been more helpful to teachers. Final audio recordings only provided information to the researcher, and gave no opportunity for feedback to the teachers. Teachers were given flexibility in terms of when they delivered the intervention, which was intended to lighten the perceived time burden. Training related to speed of delivery would have been beneficial. Initial training would have benefitted from more detail, and could usefully have been aligned to the observations that were used at the end of the programme for analysis of the audio recordings.

The games, activities and books used in the programme received a positive response. Most teachers reported that they found the resources easy to use, and all the teachers reported that the children enjoyed using the resources, especially the

songs, and that they were very keen on the central characters used in the reading scheme.

Outcomes from the assessments were mixed. Results from the British Picture Vocabulary Scale (BPVS) showed higher mean scores associated with Intervention P. In addition, there was a narrowing of the gender gap in this arm of the trial. There was higher fidelity to the programme from schools in Intervention P, that included a focus on developing aural vocabulary, which may account for this result.

For the other tests, the use of non-phonically decodable vocabulary with mixed teaching methods showed a significant, positive effect for Early Word Reading, Passage Reading Comprehension and Nouns. There were no statistically significant effects from the use of mixed methods with phonically decodable vocabulary.

In addition, results suggested that boys benefitted from using non-phonically decodable vocabulary with mixed teaching methods for Letter Sound Knowledge and Passage Reading Comprehension, and from using mixed teaching methods and phonically decodable vocabulary for the Early Word Reading test. Girls in the intervention conditions showed advantage over the girls in the Control condition on all measures. There was a positive effect for the non-phonically decodable vocabulary with mixed teaching methods for all measures.

Although there were high levels of attrition and poor compliance, the ωeebee Reading Programme (Intervention A and Intervention P) was demonstrated to be a viable reading scheme for use in Reception and Year 1 classes. It was shown to be useable by trained Teaching Assistants without detracting from the normal timetable, although results were better in terms of both fidelity and achievement in classes where children were taught by experienced teachers. In addition, there was clear evidence of impact as measured by word reading and comprehension. Following early observations of attrition and poor compliance in Study 1, Study 2 aimed to evaluate the main research questions in a more intervention-intensive environment with higher levels of fidelity to programme protocol. Study 2 is reported in Chapter Six.

Chapter Six

Study 2: Three-armed, Controlled, Efficacy Trial

This Chapter includes a brief outline of the specific methodology for Study 2, which was informed by the emerging process and implementation issues in Study 1, both its design limitations and successes. Presented in this Chapter are: the aims of the trial; design and evaluation of the trial; limitations of the study design, and pupil characteristics for this study. This is followed by an analysis of the outcomes, including results of assessments, an analysis of measures in respect of gender, fidelity to the programme and teacher feedback. A summary, discussion, and conclusions from the study outcomes follow at the end of the Chapter.

6.1 Aims of the Trial

The general aims for this study were, as for Study 1, to teach children new words, for both spoken and reading vocabulary, using the same games and activities and using an eclectic approach to developing word-recognition skills; to enhance comprehension and narrative skills. Modifications were made only to training and support and the provision of greater quantities of learning resources.

The specific aim for Study 2 was to use a independent groups longitudinal design (1 academic year) for an efficacy trial, in what were intended to be ideal conditions, whereby teaching methods, materials, and the teacher, were all sufficiently controlled that it would be possible to explore the central question of vocabulary. The purpose was to evaluate both the implementation and impact of the programme in terms of Research Question 1: Intervention P (phonically decodable vocabulary) compared to Intervention A (non-phonically decodable) and Research Question 2: the combined interventions (whole-word recognition, analytic phonics plus synthetic phonics) compared to Control (synthetic phonics only). This differed from Study 1 which included an evaluation of the implementation of the intervention. The desired primary outcome was for improved word reading and comprehension. Primary outcome measures, chosen for this trial, were word recognition, phoneme awareness and passage-reading comprehension. Secondary outcome measures were constructed for the process evaluation. Although the trial was carried out in a natural community setting, additional materials were supplied

to schools, with additional training and support to achieve near ideal conditions, and monitoring was more rigorous.

6.2 Design of Study 2

This section includes the structure of the trial, recruitment, sample size, timescale, data collection, and process evaluation procedures.

6.2.1 Three-armed Controlled Efficacy Trial

An early evaluation of the progress of Study 1 led to the design of Study 2. Firstly, there were a number of practical issues with the teaching materials which needed to be addressed and revisions made. This was mainly the result of teacher feedback in the form of questionnaires and discussions at the end of each school term. Modifications were also made to the teaching Manual, to allow for flexibility in time scale, as a result of other constraints on teacher time. A revision of the training sessions and support was also made, in order to reduce the risk of attrition, which had been a feature of Study 1, as well as the risk of non-adherence to programme protocol. The use of clusters in Study 1, for allocation to condition, had both reduced the strength of the study and made it vulnerable to attrition bias (see Chapter Five).

In order to address these issues, a split-cluster design was chosen (Hutchison & Styles, 2010). By having two arms of the intervention in the same classroom, it was possible better to control the dependent variables. In Study 1, the different teaching styles were likely to have impacted on the results. For Study 2, the individual was the unit of allocation, and for each intervention arm within each class, the teacher and teaching style, was a constant. In addition, the trial was less vulnerable to imbalance, in either arm of the trial, from attrition. Each of three classes was randomly divided into two groups of equivalent numbers (see Figure 6.1 below).

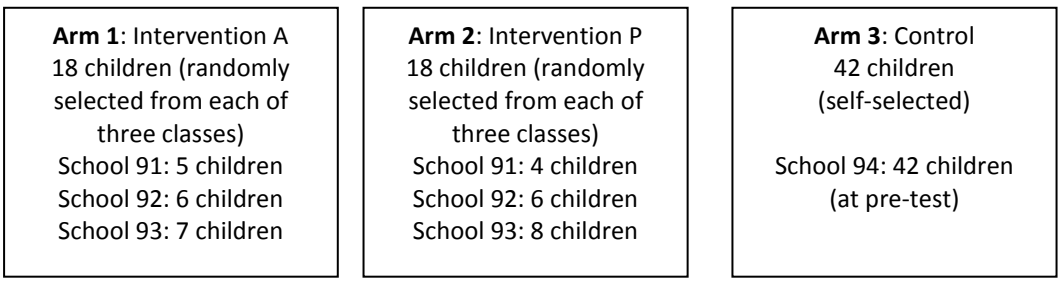


Figure 6.1 showing the three arms of the trial for Study 2.

One advantage of this design lay in the consistency of teaching style and environment in each intervention arm. It also was possible to provide more books and resources per child, and increased the opportunity for training and support to enhance validity. There was a small risk of contamination across conditions and possible teacher bias, which was addressed in the training and monitoring. One disadvantage lay in having the control group from a different school, with possible confounding factors.

The *weebee* Reading Programme was designed to run for three terms, beginning in June 2014, following the first round of assessments and the teacher training, and finishing in March 2015, before the final round of assessments. It was anticipated that each child would spend at least ninety minutes per week on programme-related activities. This included a mix of whole class and group work; designed for groups of up to four children. Teachers were given the freedom to choose how they timetabled the sessions and, as in Study 1, the majority chose to implement the programme during afternoon sessions, because of the need to adhere to the formal timetabling of morning activities. In all the schools children were taken out in groups to another classroom by the teacher. Feedback, from teachers in Study 1, had indicated that during the first term of Reception, many children were adjusting to the new environment, and the progress through the programme was necessarily slower than originally anticipated. None of the schools in Study 1 had completed the first series of books by the end of the trial. Therefore it was deemed reasonable to extend the age band for Study 2 into Year 1.

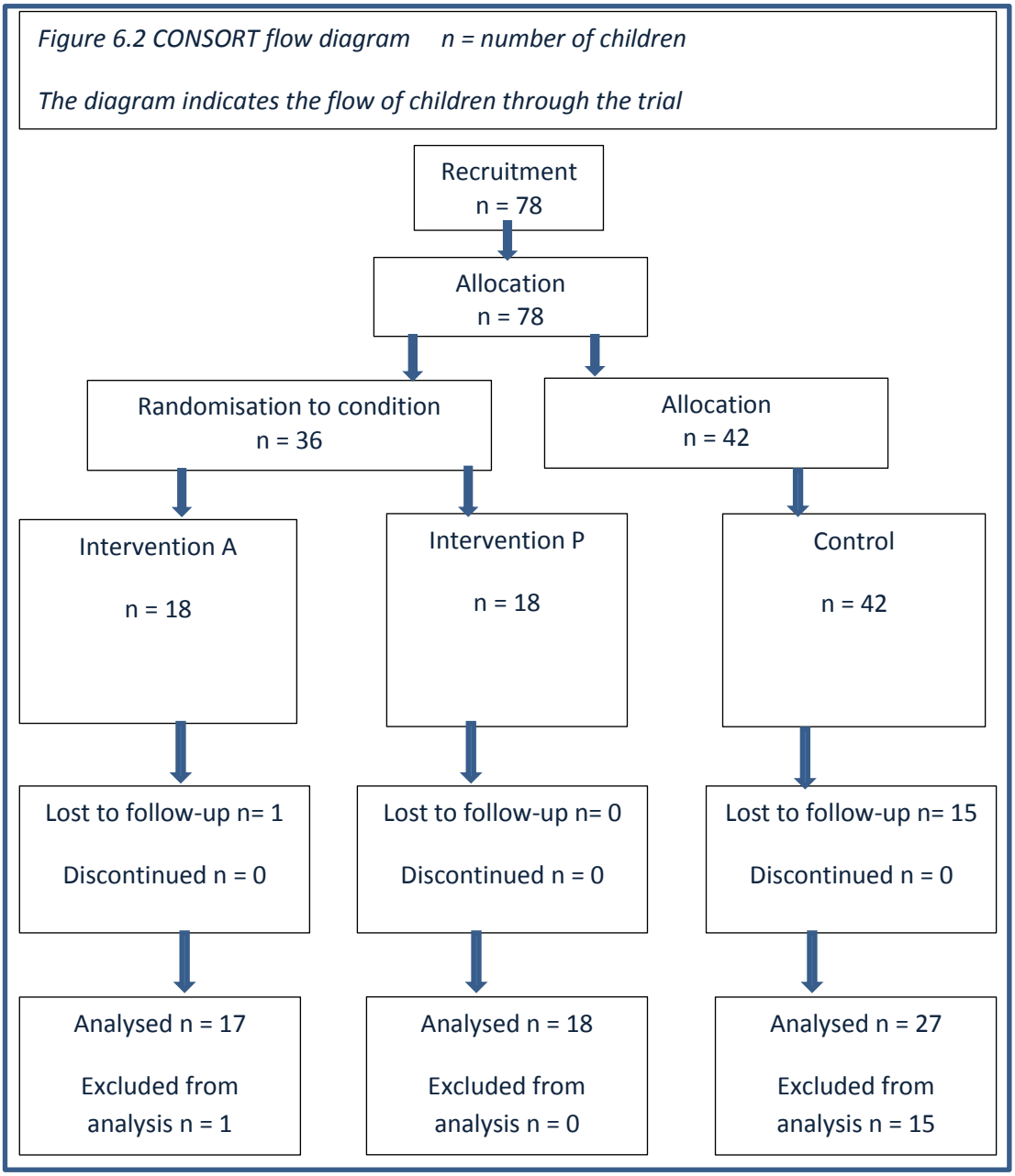
6.2.2 Recruitment and Training

All state primary schools in a mainly rural county were contacted by post. Volunteers were sought for this trial in order to try to reduce the effect of large-scale attrition. Schools who volunteer may show different characteristics from other schools, but for this trial, which sought to use 'ideal conditions,' this was not an issue. Of these, only four responded and volunteered to participate in the study. The perceived constraints of school inspections, seen to have affected schools in Study 1, may have contributed to the low numbers of schools responding to the trial. Three of these schools were single-form entry. One school had two-form entry and although they wished to participate in the research, they felt that their numbers were too large for the intended trial. Instead, a group from these two classes was

selected for participation in Study 3 (see Chapter Seven), and the rest acted as the Control group for Study 2, following the 'business-as-usual' curriculum. All the participating classes had Teaching Assistants available. All schools were mainstream state schools that would normally follow the National Curriculum. In three schools, teaching assistants were used, but in the fourth school the class teacher elected to deliver the programme. The split-cluster design meant that this was not an issue. A diagram of the flow of children through the trial is shown below in Figure 6.2.

Figure 6.2 CONSORT flow diagram n = number of children

The diagram indicates the flow of children through the trial



The whole cohort of each Reception class from each school was recruited. There were four Teaching Assistants participating. The number of additional helpers

(part-time Teaching Assistants) varied from one to two, but they all received the same training. There were five participating teachers in the intervention, who all participated in the initial training, in order to understand the programme and be able to take over the delivery if necessary. Class teachers were not expected to participate directly. The Teaching Assistants normally assigned to the Reception class by the head teacher were the facilitators of the intervention in three of the schools, the class teacher in the fourth school was the main facilitator.

Feedback from the training given in Study 1 informed the planning for training for Study 2. This was given in group seminars, with workshop activities, to permit feedback and questions to build and maintain momentum (Wenz-Gross & Upshur, 2012). Participants in the training were class teachers for Reception and Year 1, plus Teaching Assistants and SENCOS (special needs coordinators). They were first shown a presentation, which detailed the fundamental principles underlying the development of the programme and the overarching aims that related to these principles (see Appendix G). For Study 2 the teacher's Manual had been amended based on feedback from Study 1. These amendments were mainly related to the timing of delivery, but also included some clarification of implementation and more detail regarding process. A copy was provided to each class teacher during training, to allow opportunity for specific questions, and discussion, within the group. The training session was designed such that it was possible to differentiate between novice and expert and to be context specific. This allowed the session to be accessible to both less experienced Teaching Assistants and experienced teachers.

6.2.3 Sample Size

The small sample size (78 at pre-test) was calculated to have 80% power to detect a minimum effect size of 0.60. Because of the small sample size, and the exploratory nature of the study, according to Maxwell & Delaney (2008), the level of statistical significance need only be $p < .10$ to indicate trends when sample size and power are limited and the purpose is to establish evidence upon which further larger scale and more rigorous studies can be based. At post-test, having lost 16 participants to follow up, the sample size (62 at post-test) was calculated to have 80% power to detect a minimum effect size of 0.65 (equivalent to 8 months progress).

6.2.4 Timeline

The timeline for Study 2, showing recruitment, training, delivery and assessments, is indicated in Table 6.1 below.

Table 6.1
Timeline for Study 2

Date	Timeline for Study 2	Assessment details	Year Group
March 2014	Recruitment of schools		
April 2014		Assessments t 1* BPVS YARC: Letter Sound Knowledge Early Word Recognition Sound Isolation Sound Deletion	Reception
May 2014	Initial training for programme delivery		
June 2014	Delivery of first term resources to schools		
September 2014	Delivery of second term resources to schools Audio recordings of teaching Observations Further training		Year 1
November 2014	Teacher feedback Further training		
January 2015	Delivery of third term resources to schools Feedback questionnaires (teacher) Further training		
March 2015	Audio recordings Observations Feedback questionnaires (teacher)		
April 2015	Collection of other data (records of compliance; completion of activities) Teacher feedback	Assessments t 2** BPVS YARC: Letter Sound Knowledge Early Word Recognition Sound Isolation Sound Deletion Passage Reading Comprehension Timed reading of extra book	

*t1 = time 1 (pre-intervention) ** t2 = time 2 (post-intervention)

6.2.5 Measures and Data Collection

Measures for Study 2 were similar to Study 1: British Picture Vocabulary Scale (Dunn et al, 2009) and Letter Sound Knowledge, Early Word Reading and Passage Reading Comprehension. In addition, the Sound Deletion (taking away sounds from words) and Sound Isolation (providing the first or final sounds of a word) tests, to assess phoneme awareness, were also used from the York Assessment of Reading for Comprehension (YARC) (Snowling et al, 2009). As the children had been at school for two terms already, they were by this time familiar with the concept of letters representing sounds in different positions in words. These two tests indicate children's ability to manipulate sounds and phonological skills.

An extra book, made in the style of the *weebee* books, was used post-test to assess speed and fluency by timing the reading (see Table 6.2 below). No new words were introduced for this book (words used were common to both intervention arms of the trial). It included no words that were specific only to the intervention, and could therefore be used with the Control group. In addition, the illustrations used contained only images of generally known animals or objects, to avoid any bias in favour of children participating in the intervention. Speed was calculated as words per minute.

6.2.6 Process and Implementation Fidelity Procedures

Two sets of books and resources were provided to each classroom. The list of children allocated to each arm of the intervention had to be given to the teachers, but the researcher was kept blind to the allocation. The split cluster design for this trial did risk potential contamination between groups, but it was felt that this was less of an issue than the risk of large-scale attrition in one group, of the kind that had occurred during the first school term in Study 1. In order to try to improve rates of compliance to the programme protocol, teachers were given more information during their initial training regarding the aims of the study, than in Study 1, although this increased the risk of a type 1 error (rejection of a true null hypothesis). Allocation of intervention was kept blind for the assessor at pre and post-intervention.

A table of the programme components used in Study 2, the assessment tools, and rationale for their use is shown in Table 6.2 below.

Table 6.2
Showing Programme components for Study 2 linked to assessment tools

Programme components	Objectives	Rationale	Assessment
Spoken vocabulary	Increase in spoken vocabulary	Evidence suggest that increased spoken vocabulary is linked with better literacy	BPVS
Written vocabulary	Increase in word recognition	Key words – easier to read	YARC Early Word Reading
Big book discussion pages and comprehension activities	Understanding of new vocabulary	Word recognition is not sufficient for reading; comprehension is essential	YARC Passage Reading Comprehension
Analytic phonics	Recognition of phonic patterns or 'word families'	Evidence of statistical learning/learning by analogy	YARC Letter Sound Knowledge, Sound Deletion and Isolation
Nouns	Recognition of vocabulary specific to the intervention	To determine how many words recognised may be attributable to the intervention	List of 20 words common to both sets of books (Non-standardised)
Final book of intervention containing no new words	To assess levels of comprehension, speed and fluency	Comprehension is an essential component of reading	Timed reading of book. (Non-standardised)

Meetings with teachers were arranged at two time points: mid-way through the second term, and mid-way through the third term. At both of these time points there was a discussion regarding progress, and observations of a teaching session. The sessions were recorded so that they could be analysed according to a predetermined schedule. A research assistant was trained to code the audios in addition to the researcher, for reliability. Each mark awarded was agreed between the two researchers following discussion to establish inter-rater reliability of kappa = 1.0. The audio recordings were made by the researcher in Study 2, rather than requesting the teachers to do this (as was the case for Study 1), to increase the likelihood of consistency of recording between the intervention schools. Feedback was requested from teachers, and also given to teachers at these times. Feedback to teachers included positive reinforcement of desired behaviours, and attention drawn to areas of non-compliance to programme protocol. In addition, further training was given, where necessary, in delivering the intervention at the correct pace (see Chapter Three).

Monitoring of implementation fidelity was through regular visits to the schools to deliver resources, and using these visits as opportunities to request updates on progress through the programme, as well as providing an opportunity for teachers to ask questions or request support. Further information on implementation fidelity was collected at post-test. This included: feedback questionnaires, informal discussions, individual child reading logs and session-completion records (see Appendix C). The data from the session records and reading logs was used to calculate the intervention dosage. Data from questionnaires and informal discussions related to the instruction Manual, the initial training, the resources and children's engagement levels (see Appendix D).

In order to establish ideal conditions for an efficacy trial, the fidelity requirements were more rigorous than in Study 1, and required continuous monitoring and feedback. During the second term of Study 2, the Manual included individual assessment tasks with follow-up activities. For example, when the children failed to recognise a word when reading the book, the words were divided into nouns and non-nouns and then an extra activity given to help children practice those particular words. The purpose of this procedure was to increase the fidelity of children's response to the intervention (Keller-Margulis, 2012), to make it easier to determine if results were more connected to the implementation rather than the intervention.

6.3 Pupil Characteristics (at post-test)

The distribution of children with English as an additional language, and the distribution of gender across the three arms of the trial are detailed here as they were at post-test. Table 6.3 below shows the number of children with English as an additional language (EAL) for each arm of the trial.

Table 6.3
Distribution of EAL in Study 2

	Control		Intervention P		Intervention A	
EAL	Number	Percentage	Number	Percentage	Number	Percentage
Yes	4	14	0	0	0	0
No	23	86	18	18	17	17
Total	27	100	18	100	17	100

There was no significant association between the balance of EAL in each arm of the trial as calculated using Pearson's chi-square: χ^2 (2) = 5.54 (less than the critical value of 5.991 and therefore not significant at .05). A sub-analysis, with these individuals removed, did not affect the trend of the results (Group means with the four EAL children removed are shown in Appendix M). Table 6.4 below shows the number of boys and girls for each arm of the trial.

Table 6.4
Distribution of gender in Study 2

	Control		Intervention P		Intervention A	
Gender	Number	Percentage	Number	Percentage	Number	Percentage
Male	15	56	10	56	11	65
Female	12	44	8	44	6	35
Total	27	100	18	100	17	100

There was no significant association between the balance of gender in each arm of the trial as calculated using Pearson's chi-square: χ^2 (2) = 0.42 (less than the critical value of 5.991 and therefore not significant at .05). Across all three arms of the trial, there were a total of 36 boys and 26 girls at post-test. The percentage balance for gender was very similar for the Control condition and Intervention P. For all three arms of the trial there were higher percentages of boys to girls. The effect of gender is analysed and reported for each of the measures, and for each arm of the trial detailed later in this Chapter.

6.4 Impact and Process Evaluation

The outcomes reported here include analyses of both standardised and non-standard tests, using parametric tests: independent samples t-tests, and analysis of variance (ANOVA); non-parametric tests where necessary: Mann-Whitney; written feedback from teachers; and an evaluation of fidelity across the intervention arms using observations and written records.

6.4.1 Impact Evaluation

This section includes an analysis of the results from each of the standardised assessment measures used at pre-test: British Picture Vocabulary Scale (BPVS) (Dunn et al, 2009) and from the York Assessment of Reading for Comprehension (YARC) (Snowling et al, 2009): Letter Sound Knowledge (LSK), Early Word Reading

(EWR), Sound Isolation (SI), and Sound Deletion (SD). At post-test the measures used were: BPVS, LSK, EWR, SI, SD, Passage Reading Comprehension (PRC), and Reading Speed (non-standardised).

As detailed in Chapter Four, planned contrasts were intended to evaluate measures in respect of Research Question 1 (Intervention A compared to Intervention P) and Research Question 2 (Intervention A compared to the Control condition and Intervention P compared to the Control condition). Significant differences between these comparisons were to be measured using Independent samples t-tests (or Mann-Whitney). Effect sizes (using Cohen's *d*) were calculated and reported as reading progress (in months, see Higgins et al, 2013). Effect sizes are reported here for all the measures, however, due to the small sample size, only effect sizes of 0.65 or above can be treated as having at least 80% reliability to detect a significant difference.

Where there are no significant differences, trends are reported which is considered to be a legitimate approach (Torgerson & Torgerson, 2008). Pre-test and post-test scores for all measures used in Study 2 are detailed in Table 6.5 below. Group means, standard deviations, significances, and effect sizes, for each arm of the trial, for each test are shown. Between group differences in pre-test to post-test effect sizes, showing relative progression are also reported and detailed in Table 6.6.

6.4.1.1 Comparing Vocabulary (Interventions A and P)

The Kolmogorov-Smirnov test for normality of distribution (Field, 2013) showed non-normal distributions for Letter Sound Knowledge for this sample and therefore a non-parametric test was used in order to assess statistically significant group differences at pre and post-test. The Mann-Whitney test was used for this. Preliminary analysis showed that the data sets for all other measures were suitable for parametric tests.

Independent samples t-tests were first conducted on pre-test scores. Scores were not significantly different from one another at pre-test (BPVS, $t(33) = .568$ $p = .562$; EWR, $t(33) = 1.12$, $p = .268$; SI, $t(33) = 1.42$, $p = .165$; SD, $t(33) = 1.28$, $p = .207$;) which means the two groups were comparable. The Mann-Whitney test showed no significant differences for LSK at pre-test $U = 183.00$, $z = 1.04$, $p =$

.335, r = 0.17, d = 0.35 (Mean Rank P = 16.33, A = 19.76) or at post-test U = 178.5, z = 1.73, p = .405, r = 0.29, d = 0.61 (Mean Rank P = 16.58, A = 19.50).

A preliminary analysis evaluating the homogeneity of regression assumption indicated that the relationship between the covariate at pre-test and the dependent variable at post-test did not differ as a function of the independent variable for any of the measures (BPVS, F (1,31) = .388, p = .538; EWR, F (1,31) = 3.92, p = .056; SI, F (1,31) = .018, p = .893; SD, F (1,31) = 3.47, p = .072; Nouns, F (1,31) = 1.76, p = .194; PRC, using BPVS as covariate F (1,31) = .215, p = .646; RS, using BPVS as covariate F (1,31) = 1.27, p = .267) and therefore ANCOVA could be run.

There was a significant effect of the covariate for all measures (BPVS, F (1,32) = 86.48, p < .001, η_p^2 = .730; EWR, F (1,32) = 30.83, p < .001, η_p^2 = .491; SI, F (1,32) = 4.66, p = .038, η_p^2 = .127; SD, F (1,32) = 18.25, p = .001, η_p^2 = .363; Nouns, F (1,32) = 8.33, p = .007, η_p^2 = .207; PRC, F (1,32) = 17.07, p < .001, η_p^2 = .348; RS, F (1,32) = 10.53, p = .003, η_p^2 = .248).

Except for PRC (using BPVS as covariate, F (1,32) = 5.14, p = .030, η_p^2 = .138), there was no significant effect of the condition after controlling for the covariate for the other measures (BPVS, F (1,32) = 1.96, p = .171, η_p^2 = .058; EWR, F (1,32) = 4.11, p = .051, η_p^2 = .114; SI, F (1,32) = 3.51, p = .069, η_p^2 = .099; SD, F (1,32) = .517, p = .477, η_p^2 = .016; Nouns, F (1,32) = 1.10, p = .301, η_p^2 = .033; RS, F (1,32) = 1.95, p = .172, η_p^2 = .057).

Passage Reading Comprehension (using BPVS pre-test scores as covariate) was the only measure which was significantly different between the two intervention arms after controlling for pre-test scores (d = 0.75, 9 months difference at post-test). EWR was close to significance (p = .051). The between groups relative differences in pre-post effect sizes (see Table 6.6) indicate a positive effect for Intervention A for BPVS, EWR, SI and SD compared to Intervention P.

6.4.1.2 Comparing Teaching Methods

Comparing Intervention A and Control

Independent samples t-tests were first conducted on pre-test scores. Scores were significantly different from one another at pre-test for all measures except SD (BPVS, t (42) = 3.0, p = .005; EWR, t (42) = 3,07, p = .004, SI, t (42) = 2.67, p =

.004 ; SD, t (42) = 1.96 , p = .056) thus the two groups were only comparable at pre-test for SD. The Mann-Whitney test for LSK showed a significant difference between groups at pre-test: U = 86.5, z = 3.55, p = < .001, r = 0.53, d = 1.26 (Mean Rank C = 17.20, A = 30.91) but not at post-test: U = 178.5, z = 2.06, p = .067, r = 0.31, d = 0.65 (Mean Rank C = 20.61, A = 25.5).

Analysis evaluating the homogeneity of regression assumption indicated that the covariate at pre-test and the dependent variable at post-test did not differ as a function of the independent variable for BPVS (F(1,40) = .219, p = .642), PRC (using BPVS as covariate F(1,40) = .650, p = .425) and RS (using BPVS as covariate F(1,40) = 2.61, p = .114) and therefore ANCOVA could be run. However, EWR (F(1,40) = 14.11, p = .001), SI (F(1,40) = 4.66, p = .037) and SD (F(1,40) = 17.42, p < .001) were significantly different and ANCOVA could not reliably be run.

For EWR, ANOVA at pre-test showed a significant effect of condition F(1,42) = 9.45, p = .004, but not at post-test F(1,42) = 2.81, p = .101. For SI, ANOVA at pre-test showed a significant effect of condition F(1,42) = 7.17, p = .011, and at post-test F(1,42) = 6.64, p = .014. For SD, ANOVA at pre-test showed no significant effect of condition F(1,42) = 3.85, p = .0.56, but did show a significant effect at post-test F(1,42) = 11.86, p = .001. PRC was significantly different between the groups at post-test (p < .001, d = 1.18, 14 months difference) as was RS (p = .035, d = 0.64, 7 months difference).

There was a significant effect of the covariate for BPVS (F(1,41) = 111.48, p < .001, η_p^2 = .731), PRC (F(1,41) = 38.03, p < .001, η_p^2 = .481) and RS (F(1,41) = 10.39, p = .002, η_p^2 = .202). There was a significant effect of the condition after controlling for the covariate for BPVS (F(1,41) = 5.47, p = .024, η_p^2 = .118), but not for PRC (F(1,41) = 3.63, p = .064, η_p^2 = .081) or RS (F(1,41) = .691, p = .411, η_p^2 = .017).

There were statistically significant differences between these groups only for BPVS (d = 1.28, 16 months difference at post-test). The relative difference in between groups pre-post effect sizes (see Table 6.6) indicate large and positive effects for BPVS, LSK, EWR, SI and SD for Intervention A compared to the Control.

Comparing Intervention P and Control

Independent samples t-tests were first conducted on pre-test scores. Scores were not significantly different from one another at pre-test for EWR (t (43) = 1.41, p = .163), SI (t (43) = 1.36, p = .180), and SD (t (43) = .43, p = .664) but were for BPVS (t (43) = 2.34, p = .023), meaning that the two groups were comparable for EWR, SI, SD and Nouns but not for BPVS. The Mann-Whitney test for LSK showed a significant difference between groups at pre-test: U = 133.0, z = 2.61, p = .008, r = 0.38, d = 0.84 (Mean Rank C = 18.93, P = 29.11) but not at post-test: U = 222.0, z = .697, p = .462, r = 0.10, d = 0.20 (Mean Rank C = 22.22, P = 24.17).

Analysis evaluating the homogeneity of regression assumption indicated that the covariate at pre-test and the dependent variable at post-test did not differ as a function of the independent variable for BPVS (F (1,41) = .016, p = .901), LSK (F (1,41) = 3.66, p = .063), PRC (using BPVS as covariate F (1,41) = .169, p = .683) and RS (using BPVS as covariate F (1,41) = .304, p = .584) and therefore ANCOVA could reliably be run. However it did differ significantly for EWR (F (1,41) = 5.25, p = .027), SI (F (1,41) = 8.66, p = .005) and SD (F (1,41) = 6.60, p = .014) and therefore ANCOVA could not reliably be run for these measures. For EWR, ANOVA at pre-test showed no significant effect of condition F (1,43) = 2.01, p = .163, or at post-test F (1,43) = .073, p = .788. For SI, ANOVA at pre-test showed no significant effect of condition F (1,43) = 1.85, p = .180, or at post-test F (1,43) = 2.87, p = .097. For SD, ANOVA at pre-test showed no significant effect of condition F (1,43) .191, p = .664, but did show a significant effect at post-test F (1,43) = 5.03, p = .031.

There was a significant effect of the covariate for all the measures for which ANCOVA could be run (BPVS, F (1,42) = 132.9, p < .001, η_p^2 =.760; PRC, F (1,42) = 40.79, p < .001, η_p^2 = .493; RS, F (1,42) = 14.0 , p = .001, η_p^2 = .250). There was no significant effect of the condition after controlling for the covariate for any of these measures (BPVS, F (1,42) = 1.28, p = .263, η_p^2 = .030; PRC, F (1,42) = .015, p = .904, η_p^2 = .493; RS, F (1,42) = .422, p = .520, η_p^2 = .010).

There were no statistically significant differences between these groups. The between groups relative differences in pre-post effect sizes (see Table 6.6) show a positive effect for Intervention P for LSK, EWR, SI and SD compared to the Control condition.

Table 6.5

Group means, standard deviations, significance and effect sizes at pre and post-test

	Pre-test scores						Post-test scores					
	A	P	C	A v P	A v C	P v C	A	P	C	A v P	A v C	P v C
	Raw (SD)	Raw (SD)	Raw (SD)	Sig (*p*)	Sig (*p*)	Sig (*p*)	Raw (SD)	Raw (SD)	Raw (SD)	Sig (*p*)	Sig (*p)*	Sig (*p*)
	Standard	*Standard*	*Standard (SD)*	(*d*)	(*d*)	*(d)*	*Standard*	*Standard*	*Standard*	(*d*)	(*d*)	(*d*)
	(SD)	*(SD)*					*(SD)*	*(SD)*	*(SD*)			
BPVS	79.35	76.72	64.18	.562	.005	.023	97.11	91.88	78.55	.222	.000	.011
	(10.87)	(15.20)	(18.90)	(0.19)	(0.98)	(0.73)	(9.88)	(14.42)	(17.82)	(0.42)	(1.28)	(0.82)
	106.94	*103.05*	*95.44*	*.247*	*.003*	*.038*	*105.29*	*99.94*	*91.55*	*.102*	*.000*	*.017*
	(10.10)	*(9.39)*	*(13.00)*	*(0.39)*	*(0.98)*	*(0.67)*	*(8.49)*	*(10.17)*	*(11.75)*	*(0.57)*	*(1.34)*	*(0.76)*
LSK	31.0	30.33	25.74	.335	.000	.008	32.00	31.83	30.25	.405	.067	.462
	(1.27)	(2.16)	(8.20)	(0.35)	(1.26)	(0.84)	(0.0)	(.383)	(4.88)	(0.61)	(0.65)	(0.20)
	124.82	*124.05*	*114.74*	*.782*	*.038*	*.049*	*121.52*	*119.11*	*114.55*	*.046*	*.049*	*.187*
	(8.04)	*(8.23)*	*(18.27*)	*(0.09)*	*(0.71)*	*(0.65)*	*(2.78)*	*(3.98)*	*(13.98)*	*(0.70)*	*(0.69)*	*(0.44)*
EWR	11.52	8.83	6.25	.268	.004	.163	25.35	21.05	21.70	.028	.055	.788
	(6.64)	(7.46)	(4.72)	(0.38)	(0.91)	(0.41)	(3.78)	(6.91)	(8.46)	(0.78)	(0.55)	(0.08)
	114.88	*106.94*	*106.66*	*.095*	*.095*	*.953*	*112.23*	*104.22*	*106.88*	*.035*	*.220*	*.532*
	(13.93)	*(13.36)*	*(16.41)*	*(0.58)*	*(0.54)*	*(0.01)*	*(10.57)*	*(10.96)*	*(15.55)*	*(0.74)*	*(0.40)*	*(0.19)*
SI	9.29	8.00	6.44	.165	.004	.180	11.70	10.94	9.48	.025	.003	.054
	(2.02)	(3.19)	(4.06)	(0.48)	(0.88)	(0.42)	(.587)	(1.21)	(3.51)	(0.80)	(0.88)	(0.55)
	121.47	*113.94*	*108.96*	*.095*	*.022*	*.371*	*120.29*	*111.33*	*107.44*	*.011*	*.012*	*.416*
	(10.53)	*(14.85)*	*(19.97)*	*(0.58)*	*(0.78)*	*(0.28)*	*(10.18)*	*(9.53)*	*(18.47)*	*(0.90)*	*(0.86)*	*(0.26)*

SD	5.52	4.16	3.74	.207	.056	.664	9.05	8.05	5.77	.179	.000	.031
	(2.76)	(3.43)	(3.04)	(0.43)	(0.61)	(0.12)	(1.47)	(2.64)	(3.73)	(0.46)	(1.15)	(0.70)
	110.17	*99.50*	*98.74*	*.094*	*.060*	*.902*	*113.41*	*107.88*	*97.92*	*.166*	*.003*	*.062*
	(16.89)	*(19.66)*	*(20.39)*	*(0.58)*	*(0.61)*	*(0.03)*	*(8.88)*	*(13.57)*	*(19.08)*	*(0.48)*	*(1.04)*	*(0.60)*
PRC							12.82	10.44	8.44	.034	.000	.127
							(2.55)	(3.66)	(4.55)	(0.75)	(1.18)	(0.48)
							112.52	*103.05*	*99.40*	*.007*	*.004*	*.374*
							(10.33)	*(9.31)*	*(15.41)*	*(0.96)*	*(1.00)*	*(0.28)*
RS							74.29	54.38	48.96	.136	.035	.572
							(45.45)	(30.55)	(31.73)	(0.51)	(0.64)	(0.17)

Intervention A: n = 17; Intervention P: n = 18; Control: n = 27

Table 6.6

Pre-test to post-test effect sizes (Cohen's d)

	BPVS Raw	*BPVS Standard*	LSK Raw	*LSK Standard*	EWR Raw	*EWR Standard*	SI Raw	*SI Standard*	SD Raw	*SD Standard*
Intervention A	1.70	*0.17*	1.11	*0.54*	2.55	*0.21*	1.62	*0.11*	1.59	*0.24*
Intervention P	1.02	*0.31*	0.96	*0.76*	1.69	*0.22*	1.21	*0.20*	1.27	*0.49*
Difference	0.67	*0.14*	0.15	*0.22*	0.86	*0.01*	0.41	*0.09*	0.32	*0.25*
Intervention A	1.70	*0.17*	1.11	*0.54*	2.55	*0.21*	1.62	*0.11*	1.59	*0.24*
Control	0.78	*0.00*	0.66	*0.01*	2.25	*0.01*	0.80	*0.07*	0.59	*0.04*
Difference	0.92	*0.17*	0.45	*0.53*	0.30	*0.20*	0.82	*0.04*	1.00	*0.20*
Intervention P	1.02	*0.31*	0.96	*0.76*	1.69	*0.22*	1.21	*0.20*	1.27	*0.49*
Control	0.78	*0.00*	0.66	*0.01*	2.25	*0.01*	0.80	*0.07*	0.59	*0.04*
Difference	0.24	*0.31*	0.30	*0.75*	0.56	*0.21*	0.41	*0.13*	0.68	*0.45*

6.4.1.3 Gender Analyses

Contrasts were focused on the impact of the three trial arms on gender. Where parametric tests could be run, analysis of covariance was conducted at post-test, using pre-test scores as covariates (using BPVS as covariates for PRC and Reading Speed). Scores between genders at pre-test and post-test (and pre-to-post effect sizes) for BPVS, LSK and EWR are shown in Table 6.7 below. Scores between genders at post-test for PRC and Reading Speed are shown in Table 6.8.

Comparing Vocabulary: Intervention A and P (Girls)

Tests for normality showed a non-normal distribution for LSK and SI. The Mann-Whitney test showed no significant differences for LSK at pre-test U = 34.00, z = 1.34, p = .228 (Mean Rank (MR): P = 6.25, A = 9.17) or at post-test U = 30.0, z = 1.27, p = .491 (MR: P = 6.75, A = 8.50). The Mann-Whitney test for SI showed no significant difference at pre-test U = 28.50, z = .587, p = .573 (MR: P = 6.94, A = 8.25) or at post-test U = 32.50, z = 1.29, p = .282 (MR: P = 6.44, A = 8.92). Preliminary analysis indicated that ANCOVA could reliably be run for BPVS, F (1,14) = .564, p = .470; EWR, F (1,14) = .971, p = .348; SD, F (1,14) = .953, p = .352; PRC, F (1,14) = 1.16, p = .307 and Reading Speed, F (1,14) = .440, p = .522. There was no significant effect of condition after controlling for the covariate for any of the measures (BPVS, F (1,14) = .027, p = .873, η_p^2 = .002; EWR, F (1,14) = 5.31, p = .096, η_p^2 = .232; SD, F (1,14) = 301, p = .594, η_p^2 = .027; PRC, F (1,14) = .393, p = .543, η_p^2 = .034; Reading Speed, F (1,14) = 1.93, p = .191, η_p^2 = .150). The difference in gains in pre-to-post effect sizes for SD (d = 0.54) indicates a greater positive effect from Intervention P. However, for BPVS, EWR and SI (d = 0.64 for BPVS, d = 2.58 for EWR, d = 0.37 for SI), differences indicate a greater positive effect from Intervention A for girls.

Comparing Vocabulary: Intervention A and P (Boys)

Tests for normality showed a non-normal distribution for LSK and Reading Speed. The Mann-Whitney test showed no significant difference for LSK at pre-test U = 57.50, z = .191, p = .863 (Mean Rank (MR): P = 10.75, A = 11.23) or at post-test U = 60.50, z = 1.04, p = .705 (MR: P = 10.45, A = 11.50). The Mann-Whitney test for RS showed no significant difference at post-test U = 68.50, z = .951, p = .349 (MR: P = 9.65, A = 12.23).

Table 6.7

Means, standard deviations and significance at pre and post-test, and pre-to-post-test effect sizes for gender

Test	Condition	Gender	Raw Time 1 (SD)	*Standard Time 1 (SD)*	Raw Time 2 (SD)	*Standard Time 2 (SD)*	Sig. (*p*) Time 1	Sig. (*p*) Time 2	Effect size (Cohen's d) Pre-post test
BPVS	Control	M = 15	55.46(20.30)	*89.06(12.84)*	70.53(18.82)	*85.66(10.73)*	.005	.004	0.76
		F = 12	75.08(9.33)	*103.41(8.07)*	88.58(10.12)	*98.91(8.56)*			1.38
	Intervention A	M = 11	79.0(12.30)	*105.27(12.33)*	99.0(10.93)	*105.90(10.09)*	.863	.303	1.71
		F = 6	80.0(8.60)	*110.00(2.19)*	93.66(7.17)	*104.16(4.95)*			1.72
	Intervention P	M = 10	82.10(11.60)	*107.20(5.59)*	95.90(14.37)	*103.30(9.53)*	.114	.196	1.05
		F = 8	70.00(17.18)	*97.87(10.89)*	86.87(13.71)	*95.75(9.91)*			1.08
LSK	Control	M = 15	23.13(10.14)	*110.0(21.45)*	29.13(6.36)	*112.26(17.08)*	.047	.149	0.70
		F = 12	29.00(2.69)	*120.66(11.61)*	31.66(.887)	*117.41(8.62)*			1.32
	Intervention A	M = 11	30.81(1.47)	*122.63(9.20)*	32.00(0.0)	*121.27(3.13)*	.444	NA	1.14
		F = 6	31.33(.816)	*128.83(2.85)*	32.00(0.0)	*122.00(2.19)*			1.16
	Intervention P	M = 10	30.80(1.31)	*124.20(9.47)*	31.90(.316	*119.60(3.06)*	.322	.426	1.15
		F = 8	29.75(2.91)	*123.87(7.01)*	31.75(.462)	*118.50(5.07)*			0.95
EWR	Control	M = 15	5.06(4.21)	*103.26(17.41)*	19.40(9.41)	*102.20(15.90)*	.146	.116	1.96
		F = 12	7.75(5.08)	*110.91(14.66)*	24.58(6.34)	*112.75(13.51)*			2.92
	Intervention A	M = 11	11.09(7.72)	*112.18(15.53)*	24.18(3.68)	*109.09(11.76)*	.725	.066	2.16
		F = 6	12.33(4.54)	*119.83(9.64)*	27.50(2.34)	*118.00(4.38)*			4.20
	Intervention P	M = 10	8.80(7.16)	*106.40(14.00)*	20.80(7.23)	*103.10(11.25)*	.984	.867	1.66
		F = 8	8.87(8.32)	*107.62(13.42)*	21.37(6.96)	*105.62(11.17)*			1.62

SI	Control	M = 15	4.80(4.24)	*100.00(20.88)*	8.66(4.11)	*102.80(19.83)*	.016	.183	0.92
		F = 12	8.50(2.81)	*120.16(11.89)*	10.50(2.35)	*113.25(15.48)*			0.77
	Intervention A	M = 11	9.18(2.04)	*118.81(12.31)*	11.63(.674)	*119.27(12.00)*	.768	.527	1.61
		F = 6	9.50(2.16)	*126.33(2.87)*	11.83(.408)	*122.16(6.08)*			1.49
	Intervention P	M = 10	7.90(2.60)	*113.80(12.27)*	10.60(1.42)	*108.10(6.10)*	.887	.185	1.28
		F = 8	8.12(4.01)	*114.12(18.50)*	11.37(.744)	*115.37(11.79)*			1.12
SD	Control	M = 15	3.20(3.12)	*95.00(21.39)*	4.66(3.97)	*92.26(19.21)*	.312	.084	0.40
		F = 12	4.41(2.93)	*103.41(18.90)*	7.16(3.01)	*105.00(17.09)*			0.92
	Intervention A	M = 11	5.18(2.60)	*108.36(14.82)*	9.18(1.60)	*113.63(8.61)*	.500	.657	1.85
		F = 6	6.16(3.18)	*113.50(21.06)*	8.83(1.32)	*113.00(10.19)*			1.09
	Intervention P	M = 10	4.30(3.33)	*100.60(19.31)*	7.50(3.20)	*104.90(15.56)*	.860	.334	0.97
		F = 8	4.00(3.77)	*98.12(21.34)*	8.75(1.66)	*111.62(10.35)*			1.63

Table 6.8
Means, standard deviations and significance of gender at post-test

Test	Condition	Gender	Raw Time 2 (SD)	*Standard Time 2 (SD)*	Sig. (*p*) Time 2 only
PRC	Control	M = 15	6.13(4.10)	*91.66(12.03)*	.002
		F = 12	11.33(3.33)	*109.08(13.91)*	
	Intervention A	M = 11	12.90(2.73)	*111.81(11.07)*	.859
		F = 6	12.66(2.42)	*113.83(9.66)*	
	Intervention P	M = 10	10.80(3.29)	*103.90(7.82)*	.659
		F = 8	10.00(4.27)	*102.00(11.38)*	
RS	Control	M = 15	38.80(24.36)		.061
		F = 12	61.66(36.17)		
	Intervention	M = 11	69.18(55.26)		.547
	A	F = 6	83.66(18.46)		
	Intervention	M = 10	54.10(31.87)		.966
	P	F = 8	54.75(30.99)		

Preliminary analysis indicated that ANCOVA could reliably be run for all other measures (BPVS, $F(1,21) = .726$, $p = .406$; EWR, $F(1,21) = 3.93$, $p = .064$; SI, $F(1,21) = .595$, $p = .451$; SD, $F(1,21) = 3.91$, $p = .064$; PRC, $F(1,21) = .005$, $p = .945$). There was no significant effect of condition after controlling for the covariate for four of the measures (BPVS, $F(1,22) = 3.46$, $p = .079$, $\eta_p^2 = .162$; EWR, $F(1,22) = 1.41$, $p = .250$, $\eta_p^2 = .073$; SI, $F(1,22) = 2.78$, $p = .112$, $\eta_p^2 = .134$; SD, $F(1,22) = 2.15$, $p = .159$, $\eta_p^2 = .107$). However there was a significant difference for PRC ($F(1,22) = 5.11$, $p = .036$, $\eta_p^2 = .221$). The difference in gains in pre-to-post effect sizes for BPVS, EWR, SI and SD ($d = 0.66$ for BPVS, $d = 0.50$ for EWR, $d = 0.33$ for SI, $d = 0.88$ for SD) indicate a greater positive effect from Intervention A compared to Intervention P for boys on these measures.

Comparing Teaching Methods: Intervention P and Control (Girls)

Tests for normality showed a non-normal distribution for LSK and SI. The Mann-Whitney test showed no significant difference for LSK at pre-test $U = 56.00$, $z = .332$, $p = .851$ (Mean Rank (MR): C = 9.83, P = 11.50) or at post-test $U = 45.00$, $z = 1.04$, $p = .705$ (MR: C = 10.75, P = 10.12). The Mann-Whitney test for SI showed no significant difference at pre-test $U = 48.50$, $z = .039$, $p = 1.00$ (MR: C = 10.46, P = 10.56) or at post-test $U = 49.00$, $z = .085$, $p = 1.00$ (MR: C = 10.42, P

= 10.62). Preliminary analysis indicated that ANCOVA could reliably be run for all measures except SD (BPVS, F (1,16) = .117, p = .737; EWR, F (1,16) = 1.94, p = .183; PRC, F (1,16) = 1.83, p = .194; RS, F (1,16) = .560, p = .465). For SD, ANOVA at pre-test showed no significant effect of condition F (1,18) = .077, p = .785, or at post-test F (1,18) = 1.81, p = .114. There was no significant effect of condition after controlling for the covariate for BPVS: F (1,17) = .379, p = .546, η_p^2 = .022, EWR: F (1,17) = 4.42, p = .051, η_p^2 = .206, PRC: F (1,17) = .174, p = .682, η_p^2 = .010, or RS: F (1,17) = .052, p = .822, η_p^2 = .003. The difference in gains in pre-to-post effect sizes for BPVS, LSK and EWR (d = 0.30 for BPVS, d = 0.37 for LSK, d = 1.30 for EWR) indicate a greater positive effect from the Control condition for girls on these measures. However, for SI and SD (d = 0.35 for SI, d = 0.71 for SD), differences indicate a greater positive effect from Intervention P.

Comparing Teaching Methods: Intervention P and Control (Boys)

Tests for normality showed a non-normal distribution for LSK and SI. The Mann-Whitney test showed no significant difference for LSK at pre-test U = 122.50, z = 2.72, p = .007 (Mean Rank (MR): C = 9.83, P = 17.75) or at post-test U = 89.50, z = 1.13, p = .421 (MR: C = 12.03, P = 14.45). The Mann-Whitney test for SI showed no significant difference at pre-test U = 109.00, z = 1.90, p = .062 (MR: C = 10.73, P = 16.40) or at post-test U = 88.00, z = .745, p = .485 (MR: C = 12.13, P = 14.30). The Mann-Whitney test for RS showed no significant difference at post-test U = 91.50, z = .916, p = .367 (MR: C = 11.90, P = 14.65).

Preliminary analysis indicated that ANCOVA could reliably be run for all measures (BPVS, F (1,21) = .535, p = .473; EWR, F (1,21) = 2.49, p = .129; SD, F (1,21) = .945, p = .342; PRC, F (1,21) = .094, p = .763). There was no significant effect of condition after controlling for the covariate for BPVS, EWR and PRC (BPVS, F (1,22) = .411, p = .528, η_p^2 = .018; EWR, F (1,22) = .883, p = .358, η_p^2 = .039; PRC, F (1,22) = .092, p = .764, η_p^2 = .004). However there was a significant difference for SD (F (1,22) = 4.31, p = .050, η_p^2 = .164). The difference in gains in pre-to-post effect sizes for EWR (d = 0.30) indicate a positive effect of the Control condition compared to Intervention P for boys on this measure. The difference in gains in pre-to-post effect sizes for BPVS, EWR, SI and SD (d = 0.29 for BPVS, d = 0.45 for EWR, d = 0.36 for SI, d = 0.52 for SD) indicate a positive effect from Intervention P compared to the Control for boys on these measures.

Comparing Teaching Methods: Intervention A and Control (Girls)

Tests for normality showed a non-normal distribution for LSK and SI. The Mann-Whitney test showed a significant difference for LSK at pre-test U = 58.00, z = 2.12, p = .041 (Mean Rank (MR): C = 7.67, A = 13.17) but not at post-test U = 42.00, z = 1.02, p = .616 (MR: C = 9.00, A = 10.50). The Mann-Whitney test for SI showed no significant difference at pre-test U = 43.50, z = .711, p = .494 (MR: C = 8.88, A = 10.75) or at post-test U = 46.50, z = 1.17, p = .335 (MR: C = 8.62, A = 11.25).

Preliminary analysis indicated that ANCOVA could reliably be run for BPVS, PRC and RS (BPVS, F (1,14) = .068, p = .799; PRC, F (1,14) = .000, p = .999; RS, F (1,14) = .881, p = .364) but not for EWR or SD. For EWR, ANOVA at pre-test showed no significant effect of condition F (1,16) = 3.46, p = .081, or at post-test F (1,16) = 1.15, p = .298. For SD, ANOVA at pre-test showed no significant effect of condition F (1,16) = 1.34, p = .263, or at post-test F (1,16) = 1.63, p = .219. There was no significant effect of condition after controlling for the covariate for any of the measures for which ANCOVA could be run (BPVS: F (1,15) = .259, p = .619, η_p^2 = .006; PRC: F (1,15) = .603, p = .430, η_p^2 = .039 and RS: F (1,15) = 1.36, p = .260, η_p^2 = .084. The difference in gains in pre-to-post effect sizes for LSK (d = 0.16) indicate a small positive effect from the Control condition. The difference in gains in pre-to-post effect sizes for BPVS, EWR, SI and SD (d = 0.34 for BPVS, d = 1.28 for EWR, d = 0.72 for SI, d = 0.17 for SD) indicate a greater positive effect from Intervention A for girls on these measures.

Comparing Teaching Methods: Intervention A and Control (Boys)

Tests for normality showed a non-normal distribution for LSK and SI. The Mann-Whitney test showed a significant difference for LSK at pre-test U = 136.50, z = 2.89, p = .004 (Mean Rank (MR): C = 9.90, A = 18.41) but not at post-test U = 104.50, z = 1.81, p = .259 (MR: C = 12.03, A = 15.50). The Mann-Whitney test for SI showed a significant difference at pre-test U = 133.00, z = 2.64, p = .008 (MR: C = 10.13, A = 18.09) and at post-test U = 128.50, z = 2.53, p = .015 (MR: C = 10.43, A = 17.68).

Preliminary analysis indicated that ANCOVA could reliably be run for BPVS, PRC and RS (BPVS, F (1,22) = .004, p = .949; PRC, F (1,22) = .079, p = .782; RS,

F (1,22) = 1.93, p = .178) but not for EWR or SD. For EWR, ANOVA at pre-test showed a significant effect of condition F (1,24) = 6.55, p = .017, but not at post-test F (1,24) = 2.52, p = .125. For SD, ANOVA at pre-test showed no significant effect of condition F (1,24) = 2.93, p = .100, but did at post-test F (1,24) = 12.57, p = .002. There was a significant effect of condition after controlling for the covariate for BPVS and PRC (BPVS: F (1,23) = 6.03, p = .022, η_p^2 = .208; PRC: F (1,23) = 7.19, p = .013, η_p^2 = .238) but not RS (RS: F (1,23) = .077, p = .784, η_p^2 = .003). The difference in gains in pre-to-post effect sizes for BPVS, LSK, EWR, SI and SD (d = 0.95 for BPVS, d = 0.44 for LSK, d = 0.20 for EWR, d = 0.69 for SI, d = 1.45 for SD) indicate a greater positive effect from Intervention A for boys on these measures.

Comparing Intervention A and Intervention P, results suggest that the non-decodable vocabulary may have had a small but mixed effect for girls, but for boys there was a significant effect on the PRC measure. Comparing Intervention P with the Control condition, results suggest that the mixed teaching methods had little effect for the girls, but for the boys there was a significant effect for the SD measure. Comparing Intervention A with the Control condition, results suggest that mixed teaching methods in addition to non-decodable vocabulary had little effect for girls, however, there was a significant effect for boys for BPVS and PRC.

6.4.2 Process Evaluation

The outcomes reported here include written feedback from teachers and an evaluation of fidelity across the intervention arms, using observations and written records.

6.4.2.1 Feedback

Feedback regarding the Manual was mixed. One school reported that they found it satisfactory in the second term but only sometimes useful in the third term. This school also showed the lowest levels of implementation. One school reported that they found it only sometimes useful in the second term, but useful in the third. This particular school initially showed low levels of implementation, but this improved later in the trial. The third intervention school reported that the Manual was very useful at both time points. This school showed the highest fidelity to implementation.

The initial training was reported as useful by two schools, but only satisfactory by one school, this latter was the school which showed the lowest implementation levels. Support was rated between 'satisfactory' and 'very useful' (Appendix D). Use of the resources was consistently reported as easy across schools and time points. All the schools reported that the children enjoyed, or enjoyed a lot, all the games, activities, and books. Two schools noted that time was an issue but that they had noticed the benefit from using the intervention. One of the intervention schools reported their intention to run the programme again the following year.

A brief outline of the core reading programme in each school was obtained from class teachers in order better to assess how the intervention related to the way reading was being taught outside the intervention (Hill et al, 2012). Schools 93 and 91 had adopted a new synthetic-phonics based reading scheme which had very little emphasis on developing a sight vocabulary. School 92 used a mix of old and new reading schemes. The class teacher in this school referred to the overlap between the school's approach and the *weebee* Reading Programme in her written feedback. These differences were distributed evenly across the two intervention arms and so will not have caused an imbalance in results. School 94, the Control school, was using a synthetic-phonics based reading scheme alongside an older reading scheme which used guided reading, and included some words which would have been beyond the child's current decoding ability. There may have been some overlap with the intervention as a result of using of the older-style reading scheme, however none of the non-phonically decodable vocabulary was directly taught.

6.4.2.2 Fidelity

At post-test, when data was collected for analysis from participating schools, some of the implementation data was missing as some schools recorded their sessions more accurately than others. In addition, there were significant levels of attrition in the Control condition. This was caused by a high number of children coming from families connected with the military who moved away from the area. Since these children were excluded from the final analysis and the remaining children were from stable families, it is unlikely that this will have caused bias; there were no significant differences between those who moved away and those who stayed based on scores of BPVS, EWR, SI and SD at pre-test. Two of the schools reported a reluctance fully

to implement the intervention, because of concern regarding a possible Ofsted inspection. In addition, one school reported peer pressure, from other members of staff, to limit the implementation, because of concerns that it may impact on future inspections, although the head teachers in all schools had given permission for the intervention to go ahead.

Table 6.9
Mean number of words per book (per school) for Study 2

School Code: 92 Number of children = 12			School Code: 91 Number of children = 9			School Code: 93 Number of children = 15		
Book No	Mean No of words P	Mean No of words A	Book No	Mean No of words P	Mean No of words A	Book No	Mean No of words P	Mean No of words A
1	15.83	17.33	1	16.25		1		
2	16.0	17.66	2	13.75		2		
3	13.83	15.16	3	16.25		3		
4	13.5	15.5	4	13.75		4		
5	11.0	14.83	5	9.0				
6	7.16	13.16	6	4.75				
7	9.83	14.5	7	5.5				
8	8.66	13.66	8	5.25				
9	13.5	15.66	9	16.75				
10	12.83	14.5	10	16.0				
11	10.66	12.5	11	14.73				
12	11.16	13.33	12	14.5				
Total	143.96	177.79	Total	146.48		Total		

None of the schools returned the session records. School 91 provided reading records which suggested that all twelve books had been used; however there were reading records only for Intervention P. This school failed to return reading records for Intervention A, despite requests for these items, meaning that this data could not be included in the analysis. School 92 provided reading records

for all twelve books and for all children. School 92 reported spending three hours completing all the sessions for each book, including both arms of the intervention. School 91 estimated the time spent as approximately 45 minutes per group, for each book; not completing all the sessions for each book. The mean number of words read for each book, where provided, is shown in Table 6.9 above.

School 93 failed to provide any records at all, but did report using four books, and completing all the sessions for these, but provided no data to support this. However, during the second observation at this school, the resources for book 4 were indeed being used and the children were evidently familiar with the relevant book characters. Due to lack of data, it was not possible accurately to calculate implementation fidelity based on session completions for any of the schools. However, due to the split cluster design, variation in fidelity was equal across the two intervention conditions.

In addition to observations of teaching sessions, intended to assess process fidelity, audio recordings were made by the researcher at two time points: mid-way through the second term and midway through the third term. These were made at the same time as an observation of the sessions. The data from these two sources (written observation and audio recordings) were both analysed and coded for desired and undesired behaviour (see Appendix K). The behaviours were given a rating and the sum of the combined total scores for each teacher was calculated as a percentage for each of the desired and undesired behaviours as shown in Table 6.10 below.

Table 6.10
Observation results for Study 2, showing percentage ratings from combined audio and written observations at time 1 and 2

School Code	% of desired behaviours time 1	% of desired behaviours time 2	% of undesired behaviours time 1	% of undesired behaviours time 2
91	13	31	20	0
92	68	62.5	0	6.6
93	56	30.5	20	13

School 92 shows fairly consistent process fidelity. School 91 improved following the feedback after the first observations. The reduced fidelity of school 93 reflects the peer pressure to reduce compliance to the programme from staff members who were concerned that the intervention may compromise an Ofsted inspection report.

Information from questionnaires (Appendix L) provided background information of teacher experience. One Class Teacher had 10 years' experience teaching in Reception and Years 1-3 and had been on training courses for teaching synthetic phonics. The Teaching Assistants in the other three schools had between 5 and 12 years' experience. One had attended a synthetic phonics workshop but the others had no specific literacy training. In one school two Teaching Assistants shared the delivery which may have led to some inconsistency, although they had all received the same training.

6.5 Summary and Discussion of Study 2 Outcomes

This section includes a discussion of the outcomes from the measures used, in respect of the two main research questions, firstly regarding the effect of using non-phonically decodable vocabulary, and secondly regarding the use of mixed teaching methods. This is followed by a summary of the emergent group characteristics and the effects of gender, and the observed fidelity of the two intervention groups to the programme protocol.

6.5.1 Limitations of the Study Design

One potential weakness in the design of this study was the risk of cross contamination between the two intervention arms due to having split clusters. By having both sets of books and resources in the same classrooms, teachers could make errors, and it would be relatively easy for resources to get muddled between the two sets. The monitoring for Study 2 was far more thorough than in Study 1, but still did not ensure that any replacement teachers received the correct training. Organised training was only given at one time point due to time constraints in the schools, but informal training was given at various times throughout the trial. Although training was given on the correct protocol for delivery of the intervention at pre-test, there was not enough emphasis on the importance of this to the

research, which may have led to the poor compliance shown by two of the teachers, who continued to request children to sound out all the words.

A significant potential weakness of this study was the small sample size, which reduced the power of the design. In order to be able to measure a reliable effect size for this sample, it would need to be an effect size of at least 0.6. Difficulties with recruitment made this aspect unavoidable.

6.5.2 Research Question 1: *Are there measurable differences between vocabulary that is within a child's existing decoding ability (Intervention P) and vocabulary which is not so constrained (Intervention A)?*

Unlike Study 1, there were a number of measures that showed a statistically significant difference between Intervention A and Intervention P in Study 2. In addition, there were observable trends indicating differences between the three arms of the trial. For the British Picture Vocabulary Scale (BPVS), although there was no significant difference between the two intervention arms, there was an increased effect size in favour of Intervention A for both raw and standardised scores. For Letter Sound Knowledge (LSK) there was little observable difference between the two arms of the intervention and the results were clearly influenced by ceiling effects.

After controlling for the covariate, the Early Word Reading (EWR) scores showed close to a significant difference and a large pre-post effect size difference, suggesting a difference between these two arms for this measure in favour of Intervention A. For the Sound Isolation (SI) and Sound Deletion (SD) measures, there were similar moderate differences in pre-post effect sizes (shown in Table 6.6) in favour of Intervention A.

After controlling for the covariate (BPVS at pre-test) the Passage Reading Comprehension measure showed a significant effect of condition. There was a significant difference between the two intervention arms and a large effect size for both raw (d = 0.75, equivalent to 9 months difference) and standardised scores after controlling for age (d = 0.96, equivalent to 12 months difference) in favour of Intervention A.

The Reading Speed test, although not standardised is evidence of fluency in reading. A comparison between the two intervention arms showed no significant

difference, but there was a moderate effect size (d = 0.51, equivalent to 6 months difference).

In all the reported measures, Intervention A shows an advantage over Intervention P for the children in Study 2. For most of the measures these are trends. However, after controlling for the covariate at pre-test and for age using standardised scores, Early Word Reading scores were close to a significant difference and for Passage Reading Comprehension the difference was statistically significant.

6.5.3 Research Question 2: *Are there measurable differences when comparing a synthetic phonics only approach with a mixed teaching approach?*

When comparing Intervention P with the Control condition, there were no statistically significant effects of the condition for any of the measures after controlling for the covariate using ANCOVA. Nevertheless, the Sound Deletion measure was not significantly different between these two conditions at pre-test but was at post-test (p = .031) and Sound Isolation, also non-significant at pre-test was close to significance at post-test (p = .054). Pre to post effect size differences shown in Table 6.6 indicate a trend of advantage for the mixed teaching methods with phonically decodable vocabulary for oral vocabulary, word decoding and phoneme awareness.

When comparing Intervention A with the Control condition, there was a statistically significant effect of condition after controlling for the covariate using ANCOVA only for BPVS. Nevertheless, the Sound Deletion measure was not significantly different between these two conditions at pre-test but was at post-test ($p < .001$) and both the Passage Reading Comprehension and Reading Speed measures were significantly different at post-test. Pre to post effect size differences indicate a trend of advantage for the non-decodable vocabulary and mixed teaching methods over the other two conditions for oral vocabulary, word decoding, phoneme awareness and comprehension.

6.5.4 Group characteristics

The percentage of children with English as an additional language although not statistically significantly different between the control group (14%) and both arms of the intervention (0%), may have had an impact on the outcomes measured;

group means with the EAL scores removed were calculated and are shown in Appendix M. These scores show reduced differences between the control and intervention conditions but did not affect the overall trends and were not significant. The percentage of boys to girls was similar in the all three conditions of the trial, with boys having a larger percentage in each arm, which may have contributed to the effect of gender in the measures reported.

The effect of gender on British Picture Vocabulary Scale in Study 1 was minimal in all three arms of the trial. The results for Study 2 are strikingly different. For BPVS in the Control group, there was a significant difference at both pre-test (p = .005) and at post-test (p = .004). For Intervention A, although there were no significant differences at either time point, standard scores controlling for age indicate a narrowing of the gender gap. For Intervention P, there were no significant differences at either time point, but as was the case for Intervention A, standard scores indicate a narrowing of the gender gap.

Scores in all three arms of the trial were influenced by ceiling effects for the Letter Sound Knowledge measure. For the Control condition there was a significant difference for gender at pre-test (p = .047) but not at post-test (p = .149) indicating a narrowing of the gender gap. There were no significant differences in the other two arms of the trial for this measure but both raw and standard scores indicate a narrowing of the gender gap.

Similarly to results in Study 1, there were no significant effects of gender for any of the three arms of the trial in Study 2 as measured by Early Word Reading. In all three arms there was a widening gender gap for both raw and standard scores, although this effect was least in Intervention P.

There were two extra measures, included in Study 2, to assess phoneme awareness. The children had been in school for two terms at the start of this trial and so it was felt that they could manage these extra measures. The first of these, Sound Isolation, showed a significant effect of gender in the Control condition at pre-test (p = .016) but no significant effect at post-test (p = .183). For Intervention A there were no significant differences at either time point. Raw scores indicated little change from pre to post-test, but standard scores indicated a narrowing of the gender gap. For Intervention P there were no significant differences at either time point, however, both raw and standard scores indicated a widening of the gap.

The second of these measures, Sound Deletion, showed no significant effects of gender at either time point. For the Control condition both raw and standard scores indicate a widening gender gap over time. For Intervention A scores showed a narrowing of the gender gap effect over time, but for Intervention P there was a widening of the gender gap. For boys, Intervention A showed most advantage for the Sound Deletion test.

Gender effects on the Passage Reading Comprehension test were significantly different between the arms of the trial. For the Control condition, the effect size in favour of girls (d = 1.39) represented 16 months difference. By comparison, neither of the intervention arms showed a significant difference, and only small effect sizes.

The Reading Speed test, although not a standardised test, showed a similar pattern regarding gender to the Passage Reading Comprehension test. Although not statistically significant, there was a large effect size showing a gender gap in favour of girls in the Control condition (d = 0.74, equivalent to 9 months difference). By comparison, the combined intervention arms showed little effect of gender on Reading Speed.

6.5.5 Fidelity

Although some of the paper work was not returned, there was still more data available for the analysis of fidelity for Study 2 than for Study 1. Data for calculating dosage was missing from all the participating schools; however some level of implementation could be deduced from the reading records and observations. These indicated that two of the schools had completed all twelve books by the end of the trial and the third school had only completed four books. Because of the split cluster design, although there was poor compliance in one school and reduced compliance in another, this was the same for both arms of the intervention and should not have caused bias. It will, however, have diluted the observable impact of the intervention.

The data on process fidelity was collected from all participating schools, making it possible to estimate the percentage of adherence to the programme protocol. There was evidence that some of the Teachers understood the objectives of the trial, but there was also evidence of resistance to these objectives in one of

the participating schools. The increased levels of training, support and feedback for Study 2 were also evident from the observations. Nevertheless, the study would have benefitted from a third observation earlier in the trial in order to have reinforced the initial training sooner, as a number of unwanted behaviours had already become apparent by the first observation. As in Study 1, higher levels of compliance were associated with higher levels of training and experience.

6.6 Conclusions

The specific aims for Study 2 were to evaluate the impact of using non-phonically decodable vocabulary in predictable text, and of using a mix of teaching methods, rather than synthetic phonics only, in ideal conditions. Ideal conditions might be possible in laboratories, but not in classroom settings, nevertheless the quantity of resources made available, and the small numbers of children led to higher levels of implementation than in Study 1. In addition, the extra training, observation, and support provided led to higher levels of process fidelity than Study 1. The significant differences observable between Intervention A and P in Study 2, are likely to be the results of higher levels of compliance.

As part of the evaluation of the impact of the trial, there was an analysis of the study design and implementation. A number of limitations emerged. The split cluster design reduced the risk of attrition bias, but introduced the risk of cross-contamination. There was no evidence of this happening, but in the school which failed to provide reading records, there was no data available to refute this possibility. Higher fidelity may have been achieved from more training and support and at least one further observation. The small sample size was the main weakness of the trial, but this was not by design, but a failure to recruit more schools.

Assessment of implementation fidelity was made difficult as a result of missing data. However, the data that was provided did support the evidence from Study 1, indicating that higher levels of fidelity led to higher word-reading outcomes for the programme books. The feedback from teachers indicated that, on the whole, they were satisfied with their training and support, and all reported that the children enjoyed the games, activities and books used in the trial.

There were some consistent patterns from the assessment outcomes. The Control group did not show advantage over either of the intervention arms on any

of the measures except for Early Word Reading over Intervention P. Intervention A (non-phonically decodable vocabulary) showed advantage over Intervention P on all reported measures. Intervention P (mixed teaching methods with phonically decodable vocabulary) showed no advantage for four measures (Letter Sound Knowledge, Early Word Reading, Passage Reading Comprehension and Reading Speed) but showed advantage for three (British Picture Vocabulary Scale, Sound Isolation and Sound Deletion). The intervention mixed teaching methods with non-phonically decodable vocabulary showed no advantage for Letter Sound Knowledge but did show advantage for all other measures (EWR, BPVS, SI, SD, PRC and RS). Results for Study 2 indicate that there was an observable effect in respect of both research questions. In addition, results suggested that boys showed advantage from using both non-phonically decodable vocabulary and mixed methods in respect of British Picture Vocabulary Scale, Letter Sound Knowledge, Sound Isolation, Sound Deletion, Passage Reading Comprehension and Reading Speed.

Results, from both Study 1 and Study 2, indicate that mean scores were higher for children using the intervention compared to mean scores for children in the Control conditions. For both of these studies, the weebee Reading Programme was used as a general reading scheme for all the children in the class, there was no differentiation according to ability. Therefore, a third study was designed, to use the programme as an intensive intervention for children deemed, by their class teachers, to be falling behind their peers in reading progress. Study 3 is reported in Chapter Seven.

Chapter Seven

Study 3: An Intervention for Struggling Readers

This Chapter includes a brief outline of the specific methodology for Study 3. Schools, with children who attract Pupil Premium funding, are concerned with value for money in selecting catch-up interventions (Gorard et al, 2015). This study, working with small groups of children, was intended to evaluate the Reading Programme when used as an intensive intervention for struggling readers, to add to the body of knowledge concerned with evaluating catch-up reading programmes. Presented here are the aims of the trial; design and evaluation of the trial; limitations of the study design and pupil characteristics. This is followed by an analysis of the outcomes, including results of assessments, an analysis of measures in respect of gender, fidelity to the programme and teacher feedback. A summary, discussion and conclusions from the study outcomes follow at the end of the Chapter.

7.1 Aims of the Trial

The general aims for this study were to teach children, who had been identified by their class teachers as struggling readers, word decoding skills through playing games and other activities using an eclectic approach, in addition to their usual synthetic phonics, and to enhance comprehension and narrative skills.

The specific aim for Study 3 was to evaluate the Reading Programme in terms of Research Question 1: (Intervention A compared to Intervention P) and Research Question 2: (the combined intervention methods compared to the Control using synthetic phonics only) when working with struggling readers. The desired primary outcome was for improved reading ability, word-attack skills and comprehension. Thus the primary outcome measures chosen for this trial were in respect of word recognition, phoneme awareness and passage-reading comprehension; the same measures as used in Study 2.

There is evidence in the literature, regarding children with special needs (including dyslexia), that synthetic phonics is not necessarily the best approach for some learners, and that using a number of different methods can be more beneficial (Wedell, 2014). This was considered to be sufficient justification for trialling the

programme as a catch-up intervention. Secondary outcome measures were constructed for process and implementation evaluation for the whole trial period.

7.2 Design of Study 3

This section includes the structure of the trial, participant selection, sample size, timescale, data collection, and process evaluation procedures.

7.2.1 An Intervention for Struggling Readers.

This was a three-armed trial with paired randomisation, including controls, to evaluate the use of the weebee Reading Programme, as an intervention for struggling readers. This was in a single suburban school with two-form entry. The school had volunteered to be involved in Study 2, but did not have a sufficient teacher-to-pupil ratio to implement the programme as was intended. The school felt that they could only implement the programme with small selected groups of children and therefore a third study was designed for use as a small-group intervention for children who were falling behind their peers in reading progress.

There were two objectives: firstly, to explore the potential impact of a short-term intensive intervention of non-synthetic phonics-based activities for struggling readers, and secondly, to compare the use of different vocabularies (phonically decodable and non-phonically decodable) in this scenario. The study used a pre-post experimental design with randomisation of allocation to condition (A or P). Selection of participants was not randomised; teachers chose children who were deemed to be falling behind in their reading (excluding specific learning difficulties) and might benefit from extra input. However, allocation to the intervention condition was randomised.

The study began with the first round of assessments in April 2014. Unlike Studies 1 and 2, the intervention for Study 3 did not begin until the end of September and then ran for just two terms, but with greater frequency of teaching sessions. The class teachers had by this time been able to assess children's progress in order to select children who were not progressing at the expected rate to participate in the trial. All children were then re-assessed at the same time in April 2015. Programme delivery of Study 3 was identical to Study 2 except that the group sessions were smaller. Groupings for the three arms of the trial are shown in Figure 7.1 below.

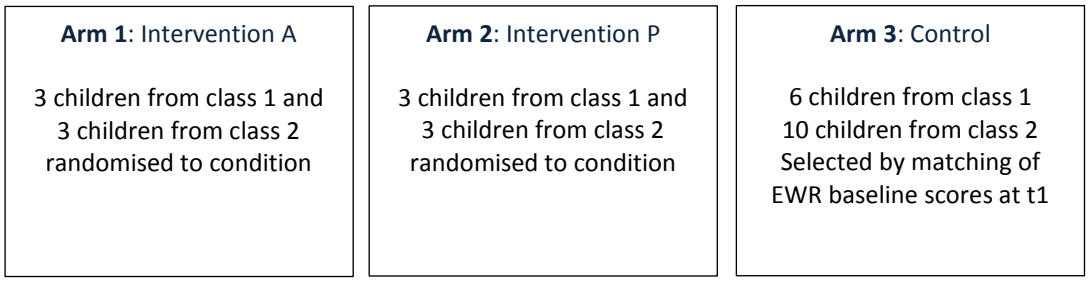


Figure 7.1 showing the three arms of the trial for Study 3.

7.2.2 Sample Size

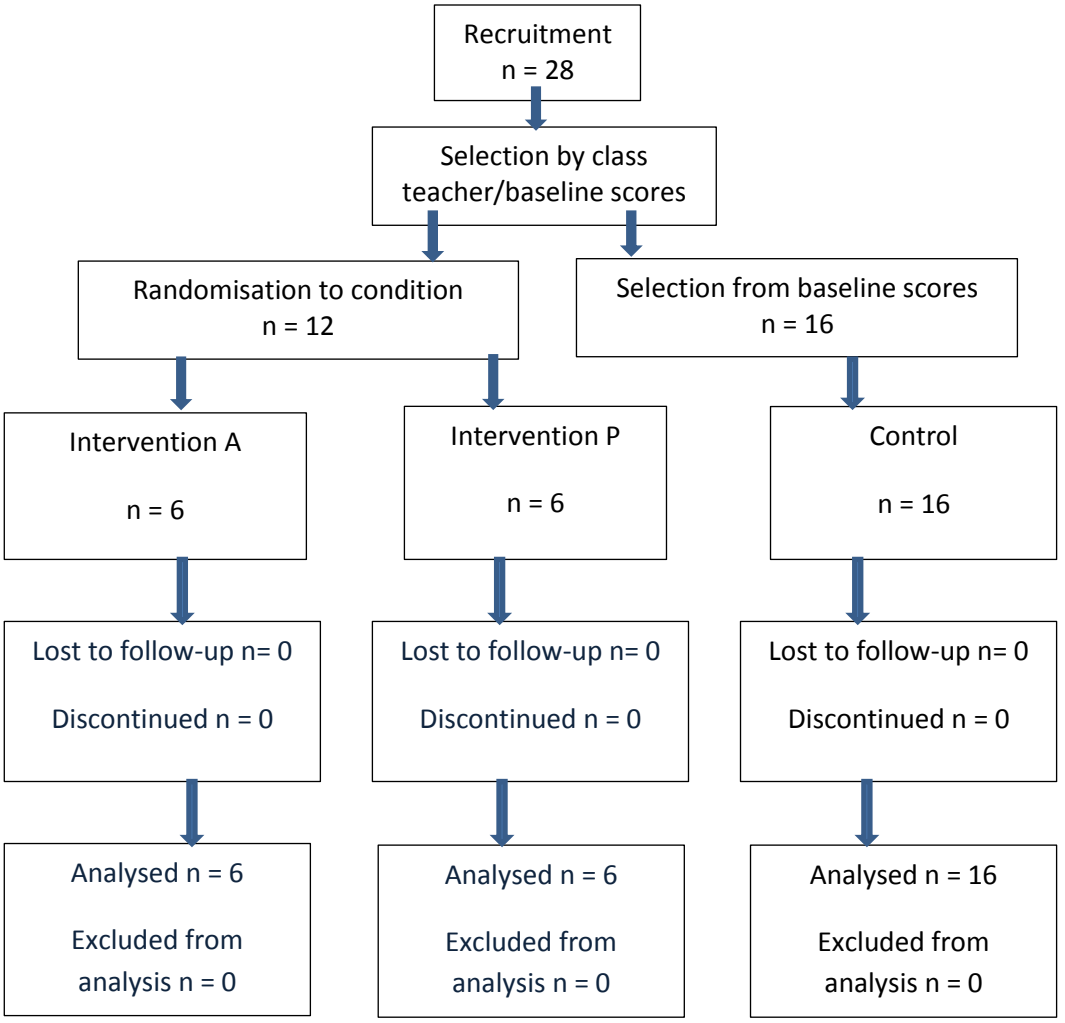
The very small sample size (28 at pre-test) was calculated to have 80% power to detect a minimum effect size of 1.10. According to Maxwell & Delaney (2008), due to the small sample size and the exploratory nature of the study, the level of statistical significance need only be p <.10 to indicate trends and to establish evidence upon which further larger scale and more rigorous studies can be based. In order to correct for potential bias (when using Cohen's *d*) resulting from the small sample size, Hedges' *g* statistic, which corrects for this bias, was included alongside Cohen's *d* where effect sizes are reported (Borenstein, Hedges, Higgins & Rothstein, 2009).

7.2.3 Recruitment

Class teachers selected six struggling readers from each of two classes. All the children had been assessed at the same pre-test time in April. By the end of September, the Year 1 class teachers had been able to gauge the children's progress and, in consultation with the Reception class teachers, were able to make a selection based on progress compared to the rest of the cohort. To reduce confounding factors, children with specific learning difficulties were excluded from the study. Baseline scores for the EWR test for the twelve participating children were used as a basis for selecting a Control group (BPVS would have led to higher numbers). Children selected were from the same cohort, from those who were within the same range of scores at pre-test as the twelve participants (standardised scores of between 81 and 111). There were sixteen children who fitted the criteria and could be used as the Control group for this study. Class teachers were not expected to participate directly and did not receive training. A diagram of the flow of children through the trial is shown below in Figure 7.2.

Figure 7.2 CONSORT flow diagram n = number of children

The diagram indicates the flow of children through the trial



7.2.4 Measures and Data Collection

Measures for Study 3 were the same as for Study 2: the British Picture Vocabulary Scale (Dunn et al, 2009), Letter Sound Knowledge, Early Word Reading, Passage Reading Comprehension, Sound Deletion, and Sound Isolation, from the York Assessment of Reading for Comprehension (Snowling et al, 2009) and Reading Speed (non-standardised). One set of books and resources were provided for each group (one for Intervention A and one for Intervention P). The list of children allocated to each arm of the intervention had to be given to the teachers, but the researcher was kept blind to the allocation. Cross-contamination was possible but unlikely due to the small numbers of children involved; secure separately-labelled boxes were provided to reduce the risk. In order to try to achieve high rates of

compliance to the programme protocol, teachers were given clear information regarding the aims of the study although this increased the risk of a type 1 error.

7.2.5 Timeline

The timeline for Study 3, showing recruitment, training, delivery and assessments, is indicated in Table 7.1 below.

Table 7.1
Timeline for Study 3

Date	Timeline for Study 3	Assessment details
March 2014	Recruitment of school	
April 2014		Assessments t 1* BPVS YARC: Letter Sound Knowledge Early Word Reading Sound Isolation Sound Deletion
May 2014	Training for programme delivery	
September 2014	Delivery of first term resources following randomisation of allocation	
November 2015	Audio recordings Feedback questionnaires	Observations Teacher feedback
January 2015	Delivery of second term resources	
February 2015	Delivery of second series of programme to second group	
March 2015	Audio recordings Feedback questionnaires	Observations Teacher feedback
April 2015	Collection of other data (tick sheets)	Assessments t 2** BPVS YARC: Letter Sound Knowledge Early Word Reading Sound Isolation Sound Deletion Passage Reading Comprehension Timed reading of extra book Teacher feedback

*t1 = time 1 (pre-intervention) ** t2 = time 2 (post-intervention)

7.2.6 Process and Implementation Fidelity Procedures

Training was given to the two teaching assistants together to allow opportunities for feedback and questions. They were first shown a presentation which detailed the fundamental principles underlying the development of the programme and the overarching aims that related to these principles. A copy of the Manual was provided to each teacher during training to allow opportunity for specific questions and discussion. In addition, the researcher modelled a number of sessions with one

of the groups for teachers to observe the correct procedures and types of language to be used.

A meeting with teachers was arranged at two time points: mid-way through the first term and mid-way through the second term. This enabled a discussion of progress and observations of a teaching session were made at each of these meetings. The sessions were recorded so that they could be analysed according to a predetermined schedule (Appendix K). Feedback to teachers, following observations, included positive reinforcement of desired behaviours and attention drawn to areas of non-compliance to programme protocol. In addition, further training was given where necessary in delivering the intervention at the correct pace and according to the original design protocol. Monitoring of implementation fidelity was through regular visits to the school to request updates on progress through the programme as well as providing an opportunity for teachers to ask questions or request support. These visits were more frequent than for Studies 1 and 2, and occurred twelve times.

7.3 Pupil Characteristics

The distribution of children with English as an additional language and the distribution of gender across the three arms of the trial are detailed here, as they were at post-test.

Table 7.2 below shows the number of children with English as an additional language (EAL) for each arm of the trial.

Table 7.2
Distribution of EAL in Study 3

	Control		Intervention P		Intervention A	
EAL	Number	Percentage	Number	Percentage	Number	Percentage
Yes	4	25	1	16	2	33
No	12	75	5	84	4	66
Total	16	100	6	100	6	100

There was no significant association between the balance of EAL in each arm of the trial as calculated using Pearson's chi-square: χ^2 (2) = 0.44 (less than the critical value of 5.991 and therefore not significant at .05). The percentage of EAL in the combined intervention (25%) was the same as in the Control condition.

Table 7.3 below shows the number of boys and girls for each arm of the trial.

Table 7.3
Distribution of gender in Study 3

	Control		Intervention P		Intervention A	
Gender	Number	Percentage	Number	Percentage	Number	Percentage
Male	10	62	5	83	3	50
Female	6	38	1	17	3	50
Total	16	100	6	100	6	100

There was no significant association between the balance of gender in each arm of the trial as calculated using Pearson's chi-square: χ^2 (2) = 1.50 (less than the critical value of 5.991 and therefore not significant at .05). The percentage of boys in the combined intervention was 66%, very similar to the percentage of boys in the Control condition. The effect of gender is analysed and reported for each of the separate measures, detailed later in the Chapter.

7.4 Impact and Process Evaluation

The outcomes reported here include analyses of both standardised and non-standard tests, using independent samples t-tests and analysis of variance (ANOVA), written feedback from teachers, and an evaluation of fidelity across the intervention arms using observations and written records.

7.4.1 Impact Evaluation

This section includes an analysis of the results from each of the standardised assessment measures used at pre-test: the British Picture Vocabulary Scale (BPVS) (Dunn et al, 2009), Letter Sound Knowledge (LSK), Early Word Reading (EWR), Sound Isolation (SI) and Sound Deletion (SD) from the York Assessment of Reading for Comprehension (YARC) (Snowling et al, 2009), and at post-test: BPVS, LSK, EWR, SI, SD, Passage Reading Comprehension (PRC) from the YARC and Reading Speed (RS) (non-standardised).

Planned contrasts were intended to evaluate measures in respect of Research Question 1: Intervention A (non-phonically decodable vocabulary) compared to Intervention P (phonically decodable vocabulary) and Research Question 2 (Intervention A with mixed teaching methods compared to the synthetic

phonics only of the Control condition and Intervention P with mixed teaching methods compared to the Control condition). Significant differences were to be measured using Independent samples t-tests where parametric tests were suitable and using Mann-Whitney where non-parametric tests were appropriate. Effect sizes (using Cohen's *d* and small-sample bias correction Hedges *g*) were calculated and reported as reading progress (in months). Results are presented in Table 7.4 below.

In addition, a gender analysis is shown for each test. The small sample sizes and imbalance of gender for Intervention P would make any analysis of gender for these two arms of the trial separately unreliable. Therefore, gender comparisons are only made between the Control group and the combined intervention groups.

7.4.1.1 Comparing Vocabulary (Interventions A and P)

The Kolmogorov-Smirnov test for normality of distribution (Field, 2013) showed non-normal distributions for Letter Sound Knowledge for this sample and therefore a non-parametric test was used in order to assess statistically significant group differences at pre and post-test. The Mann-Whitney test was used for this. Preliminary analysis showed that the data sets for all other measures were suitable for parametric tests.

Independent samples t-tests were first conducted on pre-test scores. Scores were not significantly different from one another at pre-test (BPVS, $t(10) = -1.92$. $p = .083$; EWR, $t(10) = -.365$, $p = .723$; SI, $t(10) = -1.15$, $p = .274$; SD, $t(10) = .483$, $p = .640$) which means the two groups were comparable. The Mann-Whitney test showed no significant differences for LSK at pre-test: $U = 16.50$, $z = .211$, $p = .818$, $r = 0.14$, $d = 0.30$ (Mean Rank P = 6.25, A = 6.75) or at post-test: $U = 15.0$, $z = 1.00$, $p = .699$, $r = 0.28$, $d = 0.60$ (Mean Rank P = 6.00, A = 7.00).

A preliminary analysis evaluating the homogeneity of regression assumption indicated that the relationship between the covariate at pre-test and the dependent variable at post-test did not differ as a function of the independent variable for any of the measures (BPVS, $F(1,11) = .002$, $p = .963$; EWR, $F(1,11) = 3.62$, $p = .564$; SI, $F(1,11) = .274$, $p = .615$; SD, $F(1,11) = 3.30$, $p = .107$; PRC, using BPVS as covariate $F(1,11) = .010$, $p = .923$; RS, using BPVS as covariate $F(1,11) = .008$, $p = .931$) and therefore ANCOVA could be run.

There was a non-significant effect of the covariate for any of the measures (BPVS, F (1,12) = .528, p .486, η_p^2 = .055; EWR, F (1,12) = .120, p = .737, η_p^2 = .013; SI, F (1,12) = 2.21, p = .171, η_p^2 = .197; SD, F (1,12) = .447, p = .520, η_p^2 = .047; PRC, F (1,12) = .038, p = .850, η_p^2 = .004; RS, F (1,12) = .472, p = .509, η_p^2 = .050). Except for PRC (using BPVS as covariate, F (1,12) = 9.33, p = .014, η_p^2 = .509), there was no significant effect of the condition after controlling for the covariate for the other measures (BPVS, F (1,12) = 2.31, p = .162, η_p^2 = .205; EWR, F (1,12) = 3.34, p = .101, η_p^2 = .271; SI, F (1,12) = .557, p = .475, η_p^2 = .058; SD, F (1,12) = 3.24, p = .105, η_p^2 = .265; RS, F (1,12) = .032, p = .862, η_p^2 = .004).

Passage Reading Comprehension (using BPVS pre-test scores as covariate) was the only measure which was significantly different between the two intervention arms after controlling for pre-test scores. Using raw scores d = 2.10 (g = 1.94) and using standard scores to control for age d = 1.64 (g = 1.52). After controlling for age and using Hedges correction, the effect size equated to 18 months difference. The relative between groups differences in pre-post effect sizes (Table 7.5) indicate a positive effect for Intervention A for BPVS, LSK, EWR, SI and SD compared to Intervention P for raw scores but after controlling for age, standardised scores show advantage for P over A for BPVS and SI.

7.4.1.2 Comparing Teaching Methods

Comparing Intervention A and Control

Independent samples t-tests were first conducted on pre-test scores. Scores were not significantly different from one another at pre-test for all measures except SD (BPVS, t (20) = .194, p = .849; EWR, t (20) = 1.20, p = .241, SI, t (20) = .334, p = .639; SD, t (20) = 1.57, p = .044) thus the two groups were comparable at pre-test except for SD. The Mann-Whitney test for LSK showed no significant difference between groups at pre-test: U = 43.5, z = .334, p = .747, r = 0.07, d = 0.14 (Mean Rank C = 12.22, A = 12.25) or at post-test: U = 30.0, z = 1.69, p = .203, r = 0.36, d = 0.77 (Mean Rank C = 10.38, A = 14.50).

Analysis evaluating the homogeneity of regression assumption indicated that the covariate at pre-test and the dependent variable at post-test did differ as a function of the independent variable for BPVS (F (1,22) = 7.77, p = .012) and SD (F (1,22) = 8.66, p = .009) therefore ANCOVA could not reliably be run. However, the

covariate and dependent variable did not differ significantly for EWR (F (1,22) = 2.28, p = .148), SI (F (1,22) = 3.69, p = .071), PRC (using BPVS as covariate F (1,22) = 2.17, p = .157) and RS (using BPVS as covariate F (1,22) = .460, p = .506) and therefore ANCOVA could be run. For EWR, ANOVA at pre-test showed no significant effect of condition F (1,20) = 21.45, p = .241, or at post-test F (1,20) = .276, p = .605. For SI, ANOVA at pre-test showed no significant effect of condition F (1,20) = .112, p = .742, or at post-test F (1,20) = 2.16, p = .157.

There was a significant effect of the covariate for EWR: F(1,21) = 15.62, p = .001, η_p^2 = .451; = η_p^2 =); SI: F(1,21) = 26.80, p < .001, η_p^2 = .585; PRC: F(1,21) = 14.46, p = .001, η_p^2 = .432; and RS: F(1,21) = 9.79, p = .006, η_p^2 = .340. There was no significant effect of the condition after controlling for the covariate for EWR: F(1,22) = 2.88, p = .106, η_p^2 = .132; SI: F(1,22) = 3.36, p = .082, η_p^2 = .150; or RS F(1,22) = .065, p = .802, η_p^2 = .003. However, there was a significant effect for PRC (using BPVS as the covariate): F(1,22) = 5.75, p = .027, η_p^2 =.232. Although there was no significant difference for SI following ANCOVA, this measure was not significant at pre-test but was at post-test (p = .037).

Passage Reading Comprehension (using BPVS pre-test scores as covariate) was the only measure which was significantly different between the two intervention arms after controlling for pre-test scores. Using raw scores d = 1.03 (g = 0.91) and using standard scores to control for age d = 0.86 (g = 0.82). After controlling for age and using Hedges correction, the effect size equated to 10 months difference. The relative between groups differences in pre-post effect sizes (see Table 7.5) indicate large and positive effects for BPVS, LSK, EWR, SI and SD for Intervention A compared to the Control.

Comparing Intervention P and Control

Independent samples t-tests were first conducted on pre-test scores. Scores were not significantly different from one another at pre-test for BPVS (t(20) = 1.49, p = .151), EWR (t (20) = 1.55, p = .136), SI (t (20) = .530, p = .602), and SD (t (20) = 1.17, p = .256), meaning that the two groups were comparable.

The Mann-Whitney test for LSK showed no significant difference between groups at pre-test: U = 46.50, z = .112, p = .914, r = 0.02, d = 0.04 (Mean Rank C

= 11.59, P = 11.25) or at post-test: U = 39.0, z = .803, p = .541, r = 0.17, d = 0.34 (Mean Rank C = 10.94, P = 13.00).

Analysis evaluating the homogeneity of regression assumption indicated that the covariate at pre-test and the dependent variable at post-test did not differ significantly as a function of the independent variable for SD (F (1,21) = 1.04, p = .321), PRC (using BPVS as covariate F (1,21) = 3.99, p = .061) and RS (using BPVS as covariate F (1,21) = .544, p = .470) and therefore ANCOVA could reliably be run. However it did differ significantly for BPVS (F (1,21) = 8.94, p = .008), EWR (F (1,21) = 5.77, p = .027) and SI (F (1,21) = 6.92, p = .017) and therefore ANCOVA could not reliably be run for these measures.

For BPVS, ANOVA at pre-test showed no significant effect of condition F (1,20) = 2.23, p = .151, or at post-test F (1,20) = .091, p = .766. For EWR, ANOVA at pre-test showed no significant effect of condition F (1,20) = 2.41, p = .136, or at post-test F (1,20) = 1.65, p = .213. For SI, ANOVA at pre-test showed no significant effect of condition F (1,20) = .281, p = .602, or at post-test F (1,20) = 1.88, p = .185.

There was a significant effect of the covariate for all the measures for which ANCOVA could be run (SD, F (1,22) = 49.28, p < .001, η_p^2 =.722; PRC, F (1,22) = 15.37, p = .001, η_p^2 = .447; RS, F (1,22) = 9.90 , p = .005, η_p^2 = .343). There was a significant effect of the condition after controlling for the covariate for SD, (F (1,22) = 9.32, p = .007, η_p^2 = .329), but not for PRC (F (1,22) = .275, p = .606, η_p^2 = .014) or RS (F (1,22) = .411, p = .529, η_p^2 = .021).

There was a statistically significant difference between these groups only for Sound Deletion as measured by ANCOVA. The relative between groups differences in pre-post effect sizes (see Table 7.5) show a positive effect for Intervention P compared to the Control condition for BPVS (although this diminishes after controlling for age using standard scores), EWR (only after controlling for age using standard scores) SI and SD (for both raw and standard scores).

Table 7.4

Group means, standard deviations, significance, effect sizes at pre and post-test

	Pre-test scores						Post-test scores					
	A Raw (SD) *Standard* *(SD)*	P Raw (SD) *Standard* *(SD)*	C Raw (SD) *Standard* *(SD)*	A v P Sig (*p*) (*d*)	A v C Sig (*p* (d))	P v C Sig (*p*) (*d*)	A Raw (SD) *Standard* *(SD)*	P Raw (SD) *Standard* *(SD)*	C Raw (SD) *Standard* *(SD)*	A v P Sig (*p*) (*d*)	A v C Sig (*p)* (*d*)	P v C Sig (*p*) (*d*)
BPVS	62.66	45.33	60.75	.083	.849	.151	80.83	71.00	73.62	.200	.394	.766
	(13.06)	(17.73)	(22.60)	(1.11)	(0.10)	(0.75)	(9.43)	(14.80)	(19.18)	(0.79)	(0.47)	(0.15)
	93.16	*80.83*	*91.68*	*.049*	*.807*	*.080*	*91.16*	*83.66*	*87.12*	*.147*	*.400*	*.478*
	(10.02)	*(9.06)*	*(13.20)*	*(1.29)*	*(0.12)*	*(0.95)*	*(7.83)*	*(8.68)*	*(10.38)*	*(0.90)*	*(0.49)*	*(0.36)*
LSK	26.33	23.16	23.37	.818	.476	.914	32.00	30.33	29.06	.699	.261	.541
	(4.80)	(10.00)	(9.21)	(0.30)	(0.40)	(0.02)	(0.00)	(4.08)	(6.12)	(0.60)	(0.67)	(0.24)
	115.00	*109.16*	*107.68*	*.571*	*.339*	*.861*	*120.16*	*114.33*	*108.75*	*.417*	*.097*	.473
	(13.31)	*(20.41)*	*(16.29)*	*(0.33)*	*(0.49)*	*(0.08)*	*(3.37)*	*(16.54)*	*(15.71)*	*(0.48)*	*(1.00)*	(0.34)
EWR	2.16	1.50	4.43	.723	.241	.136	21.16	13.33	18.93	.089	.605	.213
	(3.12)	(3.20)	(4.16)	(0.20)	(0.61)	(0.78)	(6.61)	(7.73)	(9.50)	(1.08)	(0.27)	(0.64)
	93.33	*85.83*	*97.75*	*.136*	*.298*	*.004*	*103.66*	*91.50*	*98.75*	*.157*	*.402*	*.259*
	(9.77)	*(5.74)*	*(8.22)*	*(0.93)*	*(0.48)*	*(1.68)*	*(11.27)*	*(15.22)*	*(12.22)*	*(0.90)*	*(0.41)*	*(0.52)*
SI	6.50	4.83	5.87	.274	.639	.602	11.16	11.00	8.62	.817	.037	.052
	(1.76)	(3.06)	(4.39)	(0.66)	(0.18)	(0.27)	(1.16)	(1.26)	(4.11)	(0.13)	(0.77)	(0.78)
	110.50	*97.66*	*102.25*	*.062*	*.341*	*.586*	*113.66*	*112.16*	*100.37*	*.834*	*.098*	*.134*
	(11.77)	*(9.26)*	*(19.23)*	*(1.21)*	*(0.51)*	*(0.30)*	*(12.69)*	*(11.46)*	*(16.98)*	*(0.12)*	*(0.88)*	*(0.81)*

SD	1.16	1.66	3.31	.640	.044	.256	7.66	5.83	4.68	.107	.012	.483
	(1.47)	(2.06)	(3.17)	(0.27)	(0.87)	(0.61)	(1.36)	(2.13)	(3.66)	(1.02)	(1.07)	(0.38)
	79.50	*82.83*	*92.93*	*.662*	*.133*	*.275*	*104.83*	*96.00*	*89.56*	*.140*	*.039*	*.385*
	(10.05)	*(15.06)*	*(19.88)*	*(0.26)*	*(0.85)*	*(0.57)*	*(8.86)*	*(12.61)*	*(15.87)*	*(0.81)*	*(1.18)*	*(0.44)*
PRC							10.66	5.66	6.75	.005	.061	.570
							(3.01)	(1.50)	(4.43)	(2.10)	(1.03)	(0.32)
							104.16	*88.00*	*92.81*	*.017*	*.089*	*.407*
							(12.87)	*(5.13)*	*(13.37)*	*(1.64)*	*(0.86)*	*(0.47)*
RS							45.22	36.89	40.98	.533	.748	.758
							(22.00)	(22.72)	(28.76)	(0.37)	(0.16)	(0.15)

Intervention A: n = 6; Intervention P: n = 6; Control: n = 16

Table 7.5

Pre-test to post-test effect sizes (Cohen's d)

	BPVS Raw	*BPVS Standard*	LSK Raw	*LSK Standard*	EWR Raw	*EWR Standard*	SI Raw	*SI Standard*	SD Raw	*SD Standard*
Intervention A	1.59	*0.22*	1.67	*0.53*	3.67	*0.97*	3.12	*0.25*	4.59	*2.67*
Intervention P	1.57	*0.31*	0.93	*0.06*	1.99	*0.49*	2.63	*1.45*	1.99	*0.94*
Difference	0.02	*0.09*	0.74	*0.47*	1.68	*0.48*	0.49	*1.20*	2.60	*1.73*
Intervention A	1.59	*0.22*	1.67	*0.53*	3.67	*0.97*	3.12	*0.25*	4.59	*2.67*
Control	0.61	*0.38*	0.72	*0.06*	1.97	*0.09*	0.64	*0.10*	0.40	*0.18*
Difference	0.98	*0.16*	0.95	*0.47*	1.70	*0.88*	2.48	*0.15*	4.19	*2.49*
Intervention P	1.57	*0.31*	0.93	*0.06*	1.99	*0.49*	2.63	*1.45*	1.99	*0.94*
Control	0.61	*0.38*	0.72	*0.06*	1.97	*0.09*	0.64	*0.10*	0.40	*0.18*
Difference	0.96	*0.07*	0.21	*0.00*	0.02	*0.40*	1.99	*1.35*	1.59	*0.76*

7.4.1.3 Gender Analyses

Contrasts were focused on the impact of the combined intervention arms compared with the Control arm on gender. Because of the imbalance in gender between Intervention A and Intervention P, a comparison of the impact of vocabulary on gender could not be made. Contrasts were only made between girls in the combined intervention arms and girls in the Control condition, and between boys in the combined intervention arms and boys in the Control condition. Where parametric tests could be run, analysis of covariance was conducted at post-test, using pre-test scores as covariates (using BPVS as covariates for PRC and Reading Speed). Scores between genders at pre-test and post-test (and pre-to-post effect sizes) for BPVS, LSK, EWR, SI and SD are shown in Table 7.6 below. Scores between genders at post-test for PRC and Reading Speed are shown in Table 7.7.

Comparing Teaching Methods (plus non-decodeable vocabulary for 3 out of 4 girls): Intervention (combined) and Control (Girls)

Tests for normality showed a non-normal distribution for LSK and EWR. The Mann-Whitney test showed no significant difference for LSK at pre-test U = 11.50, z = -.109, p = .914 (Mean Rank (MR): C = 5.58, I = 5.38) or at post-test U = 16.00, z = 1.21, p = .476 (MR: C = 4.83, I = 6.50). The Mann-Whitney test for EWR showed no significant difference at pre-test U = 4.50, z = -1.62, p = .114 (MR: C = 6.75, I = 3.62) or at post-test U = 9.00, z = -.645, p = .610 (MR: C = 6.00, I = 4.75). Preliminary analysis indicated that ANCOVA could reliably be run for all other measures (BPVS, F(1,6) = 2.03, p = .204; SI, F(1,6) = .405, p = .091; SD, F(1,6) = 5.86, p = .052; PRC, F(1,6) = .004, p = .953; RS, F(1,6) = .154, p = .708). There was no significant effect of condition after controlling for the covariate for any of the measures although SD was close to significance: (BPVS: F(1,7) = .235, p = .640, η_p^2 = .033, SI: F(1,7) = 3.59, p = .100, η_p^2 = .339, SD: F(1,7) = 5.39, p = .053, η_p^2 = .435, PRC: F(1,7) = .056, p = .823, η_p^2 = .008, or RS: F(1,7) = .218, p = .655, η_p^2 = .030). The difference in gains in pre-to-post effect sizes for LSK (d = 0.69) indicates a greater positive effect from the Control condition for girls on this measure in this study. However, for BPVS, EWR, SI and SD (d = 1.65 for BPVS, d = 1.75 for EWR, d = 2.84 for SI, d = 0.1.88 for SD), differences indicate a greater positive effect from the combined intervention conditions.

Table 7.6

Means, standard deviations and significance at pre and post-test, and pre-to-post-test effect sizes for gender

Test	Condition	Gender	Raw Time 1 (SD)	*Standard Time 1 (SD)*	Raw Time 2 (SD)	*Standard Time 2 (SD)*	Sig. (*p*) Time1	Sig. (*p*) Time2	Effect size (*d*) Pre-post test
BPVS	Control	M = 10	51.6(23.42)	*85.90(12.25)*	66.90(21.1)	*83.20(10.90)*	.013	.031	0.68
		F = 6	76.00(10.35)	*101.33(8.64)*	84.83(7.35)	*93.66(5.24)*			0.98
	Combined	M = 8	52.75(19.18)	*86.37(12.33)*	70.87(12.98)	*84.12(8.57)*	.743	.049	1.10
	Intervention	F = 4	56.50(15.43)	*88.25(9.91)*	86.00(3.65)	*94.00(5.09)*			2.63
LSK	Control	M = 10	21.00(10.95)	*105.50(19.48)*	27.70(7.49)	*107.10(18.99)*	.428	.562	0.71
		F = 6	27.33(2.94)	*111.33(9.37)*	31.33(1.21)	*111.50(8.78)*			1.77
	Combined	M = 8	25.87(4.70)	*113.37(10.51)*	30.75(3.53)	*116.12(14.46)*	.808	.808	1.17
	Intervention	F = 4	22.50(12.39)	*109.50(27.44)*	32.00(0.00)	*119.50(3.31)*			1.08
EWR	Control	M = 10	4.00(4.05)	*97.40(9.33)*	17.40(10.65)	*96.90(14.21)*	.428	.492	1.66
		F = 6	5.16(4.62)	*98.33(6.74)*	21.50(7.34)	*101.83(8.13)*			2.66
	Combined	M = 8	2.12(3.64)	*87.50(5.80)*	16.62(9.31)	*98.62(16.60)*	.933	.683	2.05
	Intervention	F = 4	1.25(1.50)	*93.75(12.52)*	18.50(5.32)	*95.50(9.71)*			4.41
SI	Control	M = 10	4.30(4.34)	*94.40(17.96)*	7.80(4.68)	*97.10(17.94)*	.061	.317	0.77
		F = 6	8.50(3.27)	*115.33(14.16)*	10.0(2.75)	*105.83(15.13)*			0.49
	combined	M = 8	5.25(2.96)	*100.37(10.45)*	11.12(1.12)	*113.87(12.00)*	.447	.870	2.62
	Intervention	F = 4	6.50(1.29)	*111.50(13.17)*	11.0(1.41)	*111.00(12.08)*			3.33
SD	Control	M = 10	3.30(3.52)	*92.90(22.39)*	4.40(4.37)	*88.60(19.36)*	.985	.654	0.27
		F = 6	3.33(2.80)	*93.00(16.81)*	5.16(2.31)	*91.16(8.70)*			0.71
	combined	M = 8	1.12(1.35)	*78.50(9.18)*	6.50(2.20)	*99.50(12.59)*	.435	.558	2.94
	Intervention	F = 4	2.00(2.44)	*86.50(17.44)*	7.25(1.50)	*102.25(9.87)*			2.59

Table 7.7
Means, standard deviations and significance of gender at post-test

Test	Condition	Gender	Raw Time 2 (SD)	*Standard Time 2 (SD)*	Sig. (*p*) Time 2
PRC	Control	M = 10	5.20(4.23)	*88.20(10.50)*	.069
		F = 6	9.33(3.72)	*100.50(14.97)*	
	combined	M = 8	7.25(3.19)	*94.62(12.22)*	.208
	Intervention	F = 4	10.0(3.65)	*99.00(14.69)*	
RS	Control	M = 10	30.49(21.31)		.056
		F = 6	58.46(32.82)		
	combined	M = 8	41.59(24.55)		.911
	Intervention	F = 4	39.99(18.11)		

Comparing Teaching Methods (plus non-decodable vocabulary for 3 out of 8 boys): Intervention (combined) and Control (Boys)

Tests for normality showed a non-normal distribution for LSK and EWR. The Mann-Whitney test showed no significant difference for LSK at pre-test U = 31.50, z = -.763, p = .460 (Mean Rank (MR): C = 8.65, I = 10.50) or at post-test U = 29.00, z = -1.23, p = .360 (MR: C = 8.40, I = 10.88). The Mann-Whitney test for EWR showed no significant difference at pre-test U = 28.50, z = -1.07, p = .315 (MR: C = 10.65, I = 8.06) or at post-test U = 39.00, z = 75.00, p = .965 (MR: C = 9.60, I = 9.38).

Preliminary analysis indicated that ANCOVA could reliably be run for SD, PRC and RS (SD, F(1,14) = 1.16, p = .298; PRC, F(1,14) = 1.17, p = .298; RS, F(1,14) = .210, p = .654) but not for BPVS or SI. For BPVS, ANOVA at pre-test showed no significant effect of condition F(1,16) = .013, p = .912, or at post-test F(1,16) = .215, p = .649. For SI, ANOVA at pre-test showed no significant effect of condition F(1,16) = .277, p = .606, or at post-test F(1,16) = 3.80, p = .069. There was a significant effect of condition after controlling for the covariate for SD (SD: F(1,15) = 15.43, p = .001, η_p^2 = .507) but not for PRC (F(1,16) = 2.18, p = .160, η_p^2 = .127) or RS (F(1,16) = 1.30, p = .272, η_p^2 = .080). The difference in gains in pre-to-post effect sizes for BPVS, LSK, EWR, SI and SD (d = 0.42 for BPVS, d = 0.46 for LSK, d = 0.39 for EWR, d = 1.85 for SI, d = 2.62 for SD) indicate a greater positive effect from the combined intervention conditions for boys on these measures in this study.

7.4.2 Process Evaluation

The outcomes reported here include written feedback from teachers and an evaluation of fidelity across the intervention arms using observations and written records.

7.4.2.1 Feedback

The teachers reported that they found both the manual and the training useful and valued the on-going support. The resources were reported as very easy to use. Regarding the games, activities and books, the teachers reported that the children enjoyed them a lot and looked forward to the sessions. An additional comment from one of the class teachers, during the first term, observed that they had already noticed improvement for some of the children participating in the trial. These teachers reported having noticed progress for most, but not all of the children. One child from each intervention arm appeared not to have had much benefit from the intervention although they clearly enjoyed the sessions. The teacher speculated that there may be specific learning issues for both of these children which had not yet been diagnosed.

At post-test, class teachers (who were not the programme deliverers) clearly stated that they felt that the intervention as a whole had been very beneficial; the children were motivated and looked forward to the sessions. The class teachers referred specifically to individuals who had made progress, and this is detailed in Appendix N. In summary, the class teachers attributed unspecified progress to the programme for three of the children, improvements in comprehension for another three children, improvements in confidence for four of the children, but for two children felt that there had been little effect, although the children had enjoyed it. Although only a small group, there were reported improvements in confidence, comprehension and enjoyment.

7.4.2.2 Fidelity

At post-test, when data was collected for analysis, all of the implementation data was available. No individuals had been lost to follow up for this trial. All the session records, and reading records, were completed and provided by the teachers. Because the intervention only ran for two terms, only eight of twelve books were completed by the group using Intervention A, and six books were completed by the

group using Intervention P. Both groups completed all the sessions for each book, but children in Intervention P had worked at a slower rate. The mean number of words read for each book, is shown in Table 7.8 below.

Table 7.8
Mean number of words per book (per intervention arm) Study 3

Book No	Mean No of words P	Number of sessions	Mean No of words A	Number of sessions
1	13.1	7	15	7
2	13.5	7	13.3	7
3	12	7	11.6	7
4	18.1	7	13.1	7
5	17	7	13.6	7
6	13.3	7	13.6	7
7			15	7
8			13.1	7
Total	87	42	108.3	56

The teachers reported that the time spent per book varied between three and four hours in total, in order to complete all the sessions before each book was read. Some of these sessions were completed within one or two weeks, but others, particularly near Christmas, took several weeks to get through because of extra commitments for the teachers and extra activities for the children.

Audio recordings of teaching sessions, to assess process fidelity, were made by the researcher at two time points: mid-way through the first term and midway through the second term. These were made at the same time as an observation of the sessions. Further training was given following the first of these observations. The data from these two sources (written observation and audio recordings) were both analysed and coded for desired and undesired behaviour (see Appendix K), with a research assistant for reliability. The behaviours were given a rating and the sum of the combined total scores was calculated as a percentage for each of desired and undesired behaviours as shown in Table 7.9 below.

Table 7.9
Observation results for Study 3 showing percentage ratings from combined audio and written observations at time 1 and 2

% of desired behaviours time 1	% of desired behaviours time 2	% of undesired behaviours time 1	% of undesired behaviours time 2
80.5	88.8	6	6

The high fidelity shown in both implementation and process is reflected in the high number of words read in the books which averaged at 14.5 for Intervention P and 13.5 for Intervention A, both of which were higher than either Study 1 or Study 2. This is an indication of the more intensive nature of this trial. The children were allocated approximately two hours per week and worked in small groups of no more than 6, which was approximately twice the length of time allocated for children in Study 2, as reported by their teachers.

7.5 Summary and Discussion of Study 3 outcomes

This section includes a discussion of the outcomes from the measures used in respect of the two main research questions, firstly regarding the effect of using non-phonically decodable vocabulary, and secondly regarding the use of mixed teaching methods. This is followed by a summary of the emergent group characteristics, the effects of gender, and the observed fidelity to the programme protocol.

7.5.1 Limitations of the Study Design

By having both sets of books and resources in the same classrooms, there was a small risk of cross-contamination, although they were kept in secure separate boxes and in colour-coded packages to reduce this risk. A significant potential weakness of this study was the small sample size which reduced the power of the design and this was partly addressed by using Hedges *g* to control for small sample sizes when calculating effect sizes. The nature of the trial made this unavoidable. Because the children had been selected by their class teachers, there was a risk of selection bias, although allocation to condition was randomised.

7.5.2 Research Question 1: *Are there measurable differences between vocabulary that is within a child's existing decoding ability (Intervention P) and vocabulary which is not so constrained (Intervention A)?*

There were no significant differences between the two intervention arms as measured by the BPVS, however, after controlling for age using standardised scores, Intervention P was the only arm of the trial which showed gains in scores between pre and post-test (d = 0.31, g = 0.29). For the SI measure the difference between raw scores and standard scores was considerable. For raw scores, Intervention A showed greater pre-post gains in effect size (d = 3.12) compared to Intervention P (d = 2.63). However, for standard scores, Intervention P showed greater gains in effect sizes from pre- to post-test (d = 1.45, g = 1.32) than Intervention A (d = 0.25, g = 0.23) a difference equivalent to 15 months.

There were no significant differences for LSK, EWR or SD, however, Intervention A showed greater gains in effect sizes than Intervention P from pre to post-test on all of these measures. For PRC, scores were significantly different, with an effect size, after controlling for age of d = 1.64 (g = 1.52) equivalent to 19 months difference. The RS test (not standardised) was not statistically different, however, mean scores for Intervention A were higher than those for Intervention P.

Other than for BPVS and SI, Intervention A (non-phonically decodable vocabulary) showed advantage over Intervention P (phonically-decodable vocabulary) for all other reported measures (LSK, EWR, SD, PRC and RS) for this group of struggling readers.

7.5.3 Research Question 2: *Are there measurable differences when comparing a synthetic phonics only approach with a mixed teaching approach?*

For the BPVS measure there appeared to be some effect from the teaching methods used in the intervention, as demonstrated by the gains in pre-post effect sizes for children in both Intervention A and Intervention P. These gains in both arms were similar (A: d = 1.59, P: d = 1.57) whereas gains in the Control were small (d = 0.61). There was a trend suggesting a small effect from the intervention as measured by LSK, although this was less noticeable after controlling for age using standardised scores. Both arms of the intervention made greater gains in EWR than the Control group, suggesting that the intervention had a positive effect, although

not statistically significant. For the SI measure, the combined intervention showed greater gains for both raw and standard scores compared with the Control group. For SD, the difference between Intervention P and the Control condition was significant after controlling for the covariate (p = .007) although this was not the case when comparing Intervention A with the control (p = .082). Nevertheless there were greater pre-post effect sizes in both arms of the Intervention (A: d = 4.59, P: d = 1.95) compared to the Control (d = 0.40). For the PRC measure there was a significant difference between Intervention A and the Control after controlling for the covariate (p = .027) but not for Intervention P (p = .606)

The impact of the mixed teaching methods used in the intervention varied across measures for this group of struggling readers, but there is evidence of children 'catching up' in EWR, SI and SD and making gains in LSK.

7.5.4 Group Characteristics

The balance of EAL between the two intervention arms was slightly different. However, the percentage within the combined intervention arms was identical to that of the Control condition and therefore unlikely to have had much impact. As with EAL, the gender balance was different between the two arms, but very similar between the combined intervention and the Control. Comparisons were therefore only made with the combined intervention for effects of gender.

Four of the measures, BPVS, EWR, SI and SD, showed greater gains for girls within the intervention condition. For the BPVS, the Control group showed a significant difference in gender at both pre and post-test and in the intervention condition there was a widening gender gap. For LSK, the Control condition showed a narrowing of the gender gap between pre and post-test. However, for the intervention, where boys had scored higher than girls at pre-test, at post-test girls were scoring higher, although not significantly.

For SD, boys made greater gains compared to girls in the intervention arms where there was a narrowing of the gap, and a widening gap in the Control condition. For SI in the control groups there was a narrowing of the gender gap, but at post-test girls were still scoring higher than boys. For the intervention, there was also a narrowing of the gender gap for SI. For PRC there was a greater gender difference in the Control condition compared to the combined intervention condition.

For the RS (non-standardised) there was a near significant gap (p = .056) in the Control condition, with girls scoring higher in the Control condition and boys scoring higher in the intervention condition. For this group of struggling readers, the outcomes suggest that for all measures used in this study, boys have benefitted from the mixed teaching methods (with non-decodable vocabulary for 3 out of 8 boys) used in the intervention when compared to boys in the Control group. Because of the imbalance in gender between the two intervention conditions (5:1 in Intervention P and 3:3 in Intervention A) it was not reasonable to compare the effects of gender between these two arms of the trial.

7.5.5 Fidelity

All the paper work was returned for Study 3. Although this study only ran for two terms, the teaching was more intensive. This is evidenced by the average number of words learned per book (14.5 for Intervention P and 13.5 for Intervention A). Both implementation and process fidelity were high.

7.6 Conclusions

The general aims for Study 3 were to extend children's reading vocabulary by developing word recognition and comprehension skills. The specific aim for this study was to evaluate the Reading Programme as a short-term intervention for struggling readers, while seeking to explore the two main research questions within this scenario.

The study design had a number of limitations, such as the small sample size, and a slight risk of cross-contamination of resources, although this is generally the nature of targeted interventions. There was an imbalance of EAL and gender across the three arms of the trial. However, there were some advantages from using the intervention with small groups, such as higher implementation and programme fidelity.

There were some patterns within this study consistent with the other two studies. For the majority of the measures, Intervention A showed advantage over Intervention P, even for struggling readers. However, in this study, Sound Isolation showed a particular advantage from Intervention P. Teaching methods from the intervention appear to have had an effect on measurements of Letter Sound Knowledge, Early Word Reading and phoneme awareness (Sound Isolation and

Sound Deletion), but not Passage Reading Comprehension or Reading Speed. For these latter two measures (PRC and RS), the non-decodable vocabulary used in Intervention A showed clear advantage over both Intervention P and the Control condition.

Results also suggest that boys made gains, relative to girls, from using the mixed methods teaching of the intervention in respect of Letter Sound Knowledge, Sound Deletion, Passage Reading Comprehension and Reading Speed. The Intervention A group showed scores which exceeded those in the comparison group on all measures. The Intervention P group exceeded those in the comparison group on three measures (LSK, SI and SD). Together with the feedback from the Class Teachers, there is evidence that the weebee Reading Programme can usefully be used for this kind of intervention with struggling readers.

Chapter Eight

Key Additional Findings

The aim of this chapter is to draw out key findings that have emerged from the research beyond the analyses which have been conducted so far. There are two main sections to this. Firstly, the data has been evaluated in terms of providing a clearer picture of the type of children who may or may not have benefitted from the intervention. Secondly, factors associated with conducting research in educational settings and in the Early Years sector in particular have been considered. This includes a few possible implications for future research in this area.

8.1 Factors Influencing Response to Intervention

Further analysis of the data collected has provided some insight regarding the type of children who may or may not have benefitted from the programme. This includes children with English as an additional language (EAL), children's responses to particular letter combinations, gender, and children's baseline scores at pre-test.

8.1.1 English as an Additional Language

The high number of children in Reception classes with EAL is a concern in many of today's schools (Strand, Malmberg & Hall, 2015); many teachers are having to deal with children who have very little knowledge of spoken English. The percentage of children with EAL in Study 1 was not significantly imbalanced across the three arms of the trial. Presented here is a comparison of mean scores for children having English as an Additional Language (EAL) with children having no EAL, within each arm of the trial.

The mean scores for pre-test and post-test are shown in Table 8.1. Tests for normality within the control condition showed a normal distribution for the British Picture Vocabulary Scale. Independent samples t-tests showed a significant difference between EAL children and non-EAL at both pre-test: t (81) = 3.87, p < .000 and at post-test: t (81) = 3.77, p <.000. The Mann-Whitney test was used for the other measures as they had non-normal distributions. There were no significant differences at pre-test (LSK: U = 308.0, z = -3.68, p = .713; EWR: U= 330.00, z = -.075, p = .941) or at post-test (LSK: U = 348.50, z = .232, p = .817, EWR: U = 441.0, z = 1.58, p = .113; PRC: U = 395.00, z = .926, p = .334; Nouns: U = 411.0 z = 1.14, p = .252). Nevertheless, Mean Rank scores (shown in Table 8.1) suggest

that children with EAL in the Control condition as a group had lower scores than non-EAL children at pre-test and higher scores at post-test for all measures except BPVS.

Tests for normality in Intervention P showed a normal distribution for the British Picture Vocabulary Scale. Independent samples t-tests showed a significant difference between EAL children and non-EAL at both pre-test: t (48) = 2.39, p = .020 and at post-test: t (48) = 2.76, p =.008. The Mann-Whitney test was used for the other measures as they had non-normal distributions. There were no significant differences at pre-test (LSK: U = 96.5, z = -1.06, p = .297; EWR: U = 105.00, z = -1.203, p = .439) or at post-test (LSK: U = 121.50, z = -.322, p = .760, EWR: U = 158.50, z = .793, p = .439; PRC: U = 111.00, z = -.633, p = .550; Nouns: U = 157.5 z = .763, p = .456). Mean Rank scores suggest that children with EAL in Intervention P as a group had lower scores at pre-test for BPVS, LSK and EWR, but at post-test had higher scores for EWR and the Nouns measures.

Tests for normality in Intervention A showed a normal distribution for BPVS and the Nouns measures. Independent samples t-tests showed a significant difference for the BPVS at pre-test: t (77) = 4.36, p < .000 and at post-test: t (77) = 3.95, p < .000. There was no significant difference for the Nouns measure: t (77) = .379, p = .706 although the receptive vocabulary scores (BPVS) were so different. Mann-Whitney tests showed no significant differences at pre-test (LSK: U = 134.5, z = -1.02, p = .319; EWR: U = 195.00, z = .303, p = .853) or at post-test (LSK: U = 171.00, z = -.292, p = .792, EWR: U = 193.00, z = .161, p = .884; PRC: U = 117.00, z = -1.38, p = .180). Mean Rank scores suggest that children with EAL in Intervention A as a group had lower scores at pre-test for BPVS and LSK, but higher scores for EWR. At post-test this had not changed.

One possible implication of these results may be that the concern regarding the number of children with EAL in Reception classes is to some extent unwarranted. Because the children are still so young when they start in Reception in England it is possible that children with EAL are able rapidly to catch up on many of the measures used. A similar analysis was not possible for Study 2 as there were no children with EAL in either of the intervention arms of the trial, or for Study 3 as there was only one child with EAL in Intervention P.

Table 8.1

Pre-test and Post-test means for EAL and Non-EAL in Study 1

		Pre-test			Post-test				
	N	BPVS	LSK [Mean Rank]	EWR [Mean Rank]	BPVS	LSK [Mean Rank]	EWR [Mean Rank]	PRC [Mean Rank]	Nouns [Mean Rank]
Control	9	39.33	4.66	1.44	56.22	29.00	15.88	15.33	15.33
EAL		(18.82)	(3.64)	(4.33)	(17.54)	(3.64)	(8.46)	(7.36)	(10.57)
			[39.22]	[41.67]		[43.72]	[54.00]	[48.89]	[50.67]
Control	74	62.54	4.97	0.35	76.35	28.39	11.10	12.64	11.10
Non-EAL		(16.72)	(3.28)	(1.17)	(14.82)	(5.11)	(8.55)	(10.57)	(9.39)
			[42.34]	[42.04]		[41.79]	[40.54]	[41.16]	[40.95]
Intervention P	6	58.33	3.83	0.00	70.50	30.16	16.33	17.83	18.00
EAL		(12.51)	(2.85)	(0.00)	(7.25)	(1.47)	(4.84)	(3.18)	(1.89)
			[19.58]	[21.00]		[23.75]	[29.92]	[22.00]	[29.75]
Intervention P	44	69.45	5.97	1.04	81.38	29.15	14.34	18.11	15.90
Non-EAL		(10.41)	(4.44)	(2.29)	(9.23)	(5.31)	(8.84)	(10.19)	(9.95)
			[26.31]	[26.11]		[25.74]	[24.90]	[25.98]	[24.92]
Intervention A	5	40.20	3.4	2.60	55.40	29.00	16.00	12.20	15.20
EAL		(10.54)	(3.97)	(5.81)	(9.71)	(3.93)	(11.24)	(9.62)	(14.37)
			[29.90]	[42.00]		[37.20]	[41.60]	[26.40]	
Intervention A	74	66.52	5.05	0.91	79.43	29.91	15.31	18.45	16.98
Non-EAL		(13.16)	(4.08)	(3.14)	(13.30)	(3.46)	(8.11)	(9.74)	(9.93)
			[40.68]	[39.86]		[40.19]	[39.89]	[40.92]	

8.1.2 Quantity and Type of Words Used

As discussed in Chapter Three, the decision to introduce twenty new words for each book was based on the success of a series of books for struggling readers of junior school age (7-11) designed in the 1970s (Harris, 1978). In Study 1, School 01 in particular demonstrated mean scores of 16/20 with some children learning all 20 words (see Table 5.11). In Study 2, two schools demonstrated mean scores of up to 16 and 17 out of 20 words (see Table 6.9), and class records showed that some children learned all 20 words. Looking at this table (6.9), it is noticeable that children in Intervention A were learning a higher mean number of words per book in School 92 (which provided these details), suggesting that children found it easier to learn a set of non-decodable words. The data collected during these studies supports the decision to introduce twenty new words for each book, even though the children in the research presented here were much younger.

The data provided by School 92 in Study 2, shows a drop in mean scores for words recognised for books 6, 7 and 8 in Intervention P which was not apparent in Intervention A. A brief analysis of the words used indicated that children were having difficulty with words ending in '_ll' and '_sh'. There was more repetition of these words in the books for Intervention P, in keeping with the controlled vocabulary and style of basal readers (see Table 8.3 below). The evidence (shown in Table 6.14) suggests that children found these words, situated in simple sentences, more difficult than a greater variety of words situated in more complex sentences. Two examples from Book 6 are shown below in Figure 8.1.

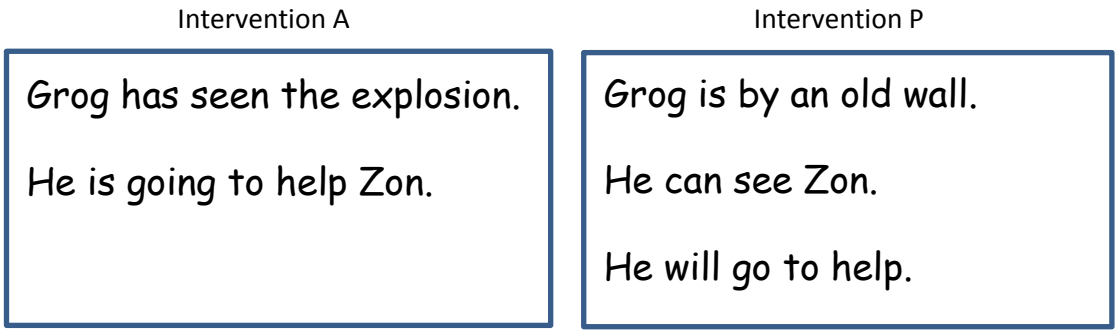


Figure 8.1 Example text from Book 6

Table 8.2
Word lists for Book 6

Intervention A					Intervention P				
soft	explosion	flash	upset	broken	fall	star	sky	flash	miss
space	twinkling	flying	crash	saucer	tall	moon	fly	crash	back
back	now	ship	fallen	hurt	wall	Zon	cry	ship	metal
seen	too	wants	not	shell	small	get	try	shell	upset

One possible explanation for this effect is that the words ending in _ll have so much similarity that they are more difficult to distinguish than words like 'explosion' and 'going'. It is also likely that children comprehend 'he is going to....' more easily than 'he will go to…' One effect of restricting the vocabulary to teach particular letter groups (seen in this example) is that the grammar (in this example the tense) has become less natural and thereby less familiar to young children. Two examples of the unnatural effect caused by restricting vocabulary in this way, taken from published reading schemes, are shown below. In the first example children would expect the cat to sit on the mat. In the second example the text makes no sense, does little to aid comprehension and makes incorrect use of puntuation.

> "Fat cat! Fat cat! Sam.
> A fat cat. Sam sat at a mat.
> A fat cat sat at a mat." (Greene and Wood, 2000, pp4-7, cited in Hassett, 2008).

> "the cow sat on a gate.
> the cow said, 'the gate is hot.'
> she said, 'I hate hot gates.'
> (Engelmann and Bruner, 1988: Story 95, cited in Hassett, 2008)

In Study 3, even though the participants were struggling readers, there were mean numbers of words recognised of up to 18 out of 20 (see Table 7.8). Intervention P showed a drop in mean number of words for Book 6 (compared to Books 4 and 5) which may be an indication of a similar phenomenon as for children in Study 2. By contrast there were a relatively constant mean number of words for Intervention A.

8.1.3 Letter-Sound Correspondence

Across all three studies, children who had been taught using a mixed approach in addition to synthetic phonics made greater gains (particularly in comprehension) than those children taught using only synthetic phonics. There were a few exceptions to this. In Study 1 the mean scores for Letter Sound Knowledge in the

Control condition were higher than those for Intervention P, suggesting that the focus on synthetic phonics and the explicit learning of letter sounds contributed to high scores on this measure.

Since the focus of synthetic phonics is the teaching of letter-sound correspondences there was a strong likelihood that the Control condition would have mean scores similar to or greater than Intervention P. However, mean scores were highest for Letter Sound Knowledge in Intervention A, suggesting that the use of non-decodable vocabulary may support the learning of letter-sound correspondences more than a controlled vocabulary.

8.1.4 Comparing Boys Across the Trial Arms

The data presented in Chapters 5-7 indicates that boys made greater gains from using both non-decodable text and mixed teaching methods in addition to synthetic phonics. For girls there were greater gains from using the non-decodable vocabulary but there was little observable effect from using mixed teaching methods in addition to synthetic phonics compared to synthetic phonics only.

In Study 1, the distribution of boys and girls was relatively even across two arms of the trial (Control: 45% boys, Intervention P: 44% boys) but there was a higher percentage of boys in Intervention A (57% boys). For the British Picture Vocabulary Scale and Early Word Reading boys performed best in Intervention P. For Letter Sound Knowledge, Passage Reading Comprehension and Nouns, boys performed best in Intervention A.

In Study 2, the random allocation resulted in there being a similar distribution of boys across two arms of the trial, with a higher percentage in Intervention A (Control: 56%; Intervention P 56%; Intervention A 65%). In this study boys performed best in Intervention P for the Letter Sound Knowledge measure. For all the other measures (British Picture Vocabulary Scale, Early Words Reading, Sound Isolation, Sound Deletion, Passage Reading Comprehension and Reading Speed) boys performed best in Intervention A. In this study, boys in Intervention A performed better than girls for Sound Deletion and Passage Reading Comprehension and equal to them for Sound Isolation. In Intervention P, boys out-performed girls for Early Word Reading and were equal to them for Passage Reading Comprehension and Reading Speed. Given that boys nationally are reported to perform 8% poorer than girls on measures of reading (National Literacy

Trust, 2012), it is interesting to note that boys in this study, with children two terms older than in Study 1, appear to 'buck' the trend.

In Study 3 the percentage distribution was unbalanced across the three arms of the trial (Control: 62% boys; Intervention P: 83% boys; Intervention A: 50% boys) and so comparisons could only be made between the Control condition and both arms of the intervention combined. In this study, boys in the combined intervention arms out-performed boys in the Control condition on all measures. For Letter Sound Knowledge and Sound Deletion, boys were equal to girls within the combined intervention. For the Sound Isolation and Reading Speed measures, boys out-performed girls in the intervention arms of the trial, counter to national trends.

8.1.5 Age Appropriateness

Feedback collected from teachers through questionnaires indicated a general approval of the overall design of the *weebee* Reading Programme. The design of the resources, the illustrations, story content and songs were considered by most to be age-appropriate. There was unanimous agreement amongst teachers that children in all studies enjoyed the teaching resources and the reading books; teachers and children liked the books, characters and songs. In Study 1, a number of parents had commented on the children's enthusiasm for the *weebee* characters.

> "'Sliding houses' worked well from the outset and a few parents said how much they had enjoyed reading the words again at home." (School 04)
> "Found it quite time consuming trying to fit in around all the other things we have to do. However it was worth it because the children have loved it, we have seen an improvement and they can't stop talking about the weebees!" (School 03)
> "The children have really enjoyed the games […] I will use it next year as an additional programme as the children have loved it" (School 01)

In the same study, in one class with a large number of very young children, the teacher thought the books were over long.

> "Books were simply too long. It was really rewarding to see the children reading so well but they struggled with reading stamina […] I had eleven children doing guided reading by the end of the year, I've never had that before." (School 04)

In Study 2 the resources were reported as motivating and enjoyed by all the children. On the Likert Scale (1-5) questionnaire (Appendix D) for questions 4 – 6 related to enjoyment, 61% gave a rating of 4 and 39% gave a rating of 5. In one of the schools in this study the teacher had devised actions to go with two of the

songs which the children clearly enjoyed. Two schools reported that having observed the benefits from the use of the programme with the children who were in Year 1 at post-test they had already begun using it with the new Reception cohort, acknowledging the appropriateness for use with the youngest age group.

In Study 3, teachers reported that the children looked forward to the sessions and enjoyed all the activities and books. Teachers associated the enjoyment of the sessions with an improvement in confidence and comprehension when reading (Appendix N). The empathetic response to the design of the weebee characters spanned the age groups; there was no suggestion that any of the children were too old or too young for the books.

8.2 Conducting Educational Research in Early Years Settings

There were a number of factors that affected participation in the research, fidelity to the implementation of the weebee Reading Programme, and the quality of delivery. Each of these is discussed below.

8.2.1 Participation

In the current climate of school inspections, assessment targets, record keeping and performance management, many schools were unwilling either to participate in or fully implement an additional intervention. Concerns were raised about an increase in paper work, time constraints and the need for compliance within statutory requirements. A survey of recruitment for randomised controlled trials in schools suggests that the complexities of participation increases unwillingness of schools to participate in research and that this is a widespread issue (Roschelle et al, 2014).

8.2.1.1 Recruitment Procedures

Recruitment in Study 1, although it proved difficult, nevertheless resulted in 12 schools agreeing to participate. Schools were approached directly via a phone call and a follow-up email to arrange a meeting. Head teachers and teachers made it clear that it was important for them to see the materials for the intervention before they would agree to making any commitment and they also needed to be clear regarding the amount of paper work that would be required, which was kept to a minimum. On the whole schools responded well to offers of free resources, and were particularly keen to have copies of the pre-test results for their own use as

baseline assessments. Impending changes to the curriculum may have contributed to the high levels of attrition and poor fidelity in this study, but numbers were nevertheless higher than for Study 2 as a result of using a direct approach.

Recruitment for Study 2 was approached differently, in the hope that there would be less attrition, by asking for volunteers rather than approaching schools directly and seeking to persuade them to participate. The response was very low, possibly as a result of changes to the curriculum, with only 4 schools volunteering out of 70 who were invited to participate. However, this approach did result in lower levels of attrition. Children who were lost to follow up had moved away; no children, staff or schools withdrew from the trials.

8.2.2 Fidelity (Compliance to Programme Protocol)

Compliance to the programme protocol was mixed across all studies. Poor compliance was associated with a lack of time allocation to the activities, levels of experience and reluctance to risk using an alternative approach. As a consequence of this many teachers failed to ensure that all children completed all activities and also failed to maintain records. The response to record keeping was variable with some teachers in all studies showing high levels of compliance and some teachers in Studies 1 and 2 showing poor compliance (there were no examples of poor compliance in Study 3).

Study 1

The cluster design, used in Study 1, made management of resources and within-class organisation easier (having only one set of resources) but resulted in outcomes which could have been attributable to the classroom environment and the style or experience of the teacher. Even with this simple design (all the children using the same materials) the workload was considered too heavy for many of the teachers. Some of these teachers withdrew part way through, did not even begin (see Figure 5.3) or handed all responsibility to Teaching Assistants without ensuring that protocol was being adhered to. The paper work was kept to a minimum by using a simple tick chart but most schools failed to keep either the individual children's reading records or the class session records. The audio recordings were not carried out according to the instructions given and this may have been a reflection on the actual delivery of the programme; it may have been an indication

that the teachers were not delivering all the activities. Only having the audio recordings at the end of the study meant that they could not be used to influence teacher behaviour but only provide a record of some of what had occurred. Their use had been intended only as a record, not originally as a tool to promote compliance.

Having gained the agreement of all leadership, teaching and support staff there was an expectation of compliance to protocol and even some interest in the research. All the resources had been pre-prepared, clearly packaged and had detailed but simple instructions designed for this young age group, and yet they were not made full use of in some schools. In addition, few of the schools had taught the children all the songs from the CD and although they were given sets of twelve books, these were not all used.

Audio recordings indicated that for many, the training given, particularly in terms of language to be used and the appropriate pace for activities, did not seem to impact on practice. This may have been due to underlying pre-conceived assumptions which had not been adequately countered by the training. It may be that this kind of training needs more in-depth explanation which includes research-based evidence to counteract resistance to alternative approaches.

Information relating to dosage and implementation was only collected at the end of the study and as such could only be used as a record not as a tool to promote adherence to the programme protocol. In this study there was evidence of a link between implementation and outcome, shown in the relative percentages of compliance within schools, as discussed in Chapter 5. Where there was both compliance to programme protocol and adherence to the intended implementation, the outcomes in terms of number of words learned was discernibly higher (see Table 5.8).

Study 2

The split cluster design used in Study 2 reduced the risk of attrition bias but did increase the complexity of the logistics in terms of delivery of the intervention. It became necessary to have two sets of books and teaching resources in each participating classroom. This added complexity may have contributed in part to the poor compliance in two of the schools. However, it did reduce the effect of having a

range of experience across those delivering the programme as this was shared across the two intervention arms. Observations and audio recordings in this study were intended to be used as a tool to promote compliance and this proved to be effective for some but not all teachers. Dosage and implementation data was only collected post-test for this study and not used as a tool to promote adherence. Had this information been collected termly or even half-termly, it could have been used for performance feedback and as an additional tool to promote implementation of the programme.

Two schools had recently adopted new synthetic-phonics-based reading schemes to which they were committed which may have led in part to poorer levels of compliance to the programme protocol. These same two schools also reported a reluctance fully to implement the programme because of concerns regarding imminent Ofsted inspections and peer pressure from other members of staff. This may be a product of school-based performance assessment and a reluctance to take what could be seen as a risk in implementing an alternative approach. The school that demonstrated highest levels of fidelity and compliance was using a mix of old and new reading schemes and the class teacher was open to using alternative approaches. The differences between schools, in terms of programme implementation, is unlikely to have greatly affected the outcome scores in terms of comparisons between the three arms of the trial, but may have reduced the measured impact of the intervention arms.

Audio recordings and observations demonstrated that School 92, which had the highest levels of implementation fidelity, also had highest levels of compliance to the programme protocol and was the only school to demonstrate none of the undesired behaviours listed. There appeared to be a clear link between experience (School 92 teacher had ten years of classroom experience) and compliance. The training given in Study 2 (although considerably more detailed and targeted than in Study 1) was nevertheless unable to prevent teachers in the other schools from continuing to demonstrate undesired behaviours in the programme delivery (see Table 6.15).

Study 3

Records of dosage were checked more frequently in Study 3 and it is possible that this contributed to the higher fidelity to programme protocol. Visits intended to

provide extra training and support were more frequent for this study than in Studies 1 and 2 as this was operating as a targeted intervention. Audio recordings and observations indicated high levels of desired behaviours (80 - 88.8%) and only 6% of undesired behaviours even though the teachers in Study 3 were less experienced Teaching Assistants. The high fidelity shown was reflected in the high numbers of words learned per book and the clear progress made on the measures used.

8.2.3 Quality of Delivery

The quality of delivery of the programme was associated in part with compliance to the programme protocol, but was also dependent on the effectiveness of the initial training and the experience and prior training of the teachers.

8.2.3.1 Training

Training in Study 1 was limited to demonstrating how the materials were to be used and emphasising the importance of not making the children read aloud or 'sound out' the words, but having the teacher model speaking the words at all times. Individual teachers or groups were met with and spent between one and two hours looking at all the resources, playing the games according to the rules, looking through the manual and discussing questions.

For Studies 2 and 3, training was only at group level and designed to encourage discussion and in-depth questioning as it was felt important that teachers should have a better understanding of the programme in order more easily to comply to protocol. These sessions consisted of a slide presentation followed by workshop activities to use the resources and included a question and answer session. Some of the background theory of the design of the materials was given, although this was kept to a minimum due to the mixed experience of the staff and time pressures in the schools. There was more opportunity to look at all the materials and the manual than with Study 1 since the design and production had already been completed. For Studies 2 and 3 there was on-going training after the initial session which followed each observation of a lesson (three sessions for Study 2 and twelve for Study 3). This gave an opportunity for performance feedback and reiteration of programme protocol. This was more frequent in Study 3 where, as seen above, levels of fidelity were highest.

For Study 3, due to the later start of the intervention, a second training session was given to the teachers in the September of the autumn term. In addition, the researcher modelled a number of sessions to demonstrate the language and pace to be used. Performance feedback was given shortly after the start of the intervention in the November. There were clear associations between higher levels of compliance and frequency and depth of training (the mean number of words – total number of words learned divided by number of books read – was higher in Study 3 than for Study 2).

8.2.3.2 The Value of Experience

In Study 1, the three schools that reported their data and had higher levels of fidelity and measured outcomes were also the schools where the intervention was delivered by experienced teachers. This has implications for this kind of research in schools in terms of differentiating between outcomes that appear to be related to an intervention and outcomes that are related to the quality of the teacher. It also raises questions regarding the use of Teaching Assistants for delivering interventions in schools unless they have had suitable training. Higher fidelity was also associated with Intervention P which may be attributable to its closer proximity to current approaches to teaching reading than Intervention A; in other words, teachers may have felt more confident in implementing Intervention P which was closer to existing classroom practice. In Study 3, higher fidelity was associated with increased frequency of training, support and modelling. Although the teachers for this study were both Teaching Assistants, the more intensive training and frequent visits, alongside a deep knowledge of the children from having worked with them in similar withdrawal groups may have resulted in high levels of compliance. In Studies 1 and 2, higher dosage was associated with more experienced teachers, especially when compared to Teaching Assistants. The audio recordings from School 01 (in Study 1) and 92 (in Study 2) were of experienced class teachers and demonstrated higher levels of compliance and vocabulary extension as well as the highest percentage of words learned. Amongst Teaching Assistants there was a high turnover (except in Study 3) and job shares also became an issue for some schools.

8.2.4 The Impact of Fidelity: A Case Study (one school)

In Study 2, School 92 demonstrated particularly high levels of fidelity and compliance to the programme protocol. It was the only school to complete all the

activities and all the books. There was evidence of high levels of desired behaviours in the programme delivery and low levels of undesired behaviour. The class was divided exactly between Intervention A and P so that there were equal numbers in each arm of the trial and the same experienced teacher taught all the children. Presented below are graphs of each measure at pre and post-test (only at post-test for Passage Reading Comprehension and Reading Speed) which show the spread of scores for each intervention arm in this class. Intervention P (Condition 2.00) is shown in blue and Intervention A (condition 3.00) is shown in green.

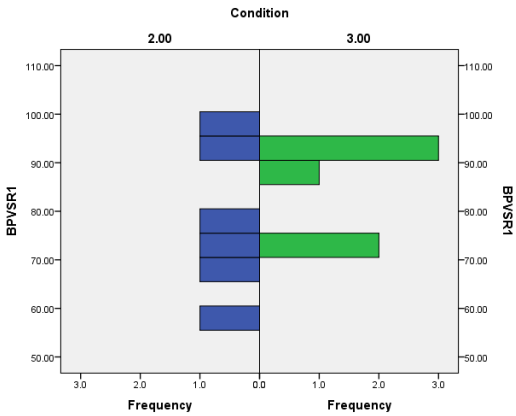


Figure 8.2a BPVS at pre-test.

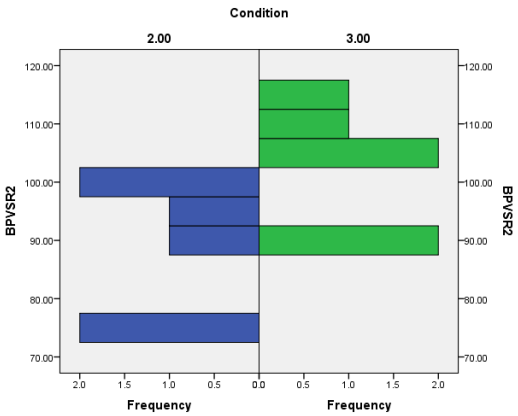


Figure 8.2b BPVS at post-test

The British Picture Vocabulary Scale: Audio recordings and observations of this class indicated high levels of vocabulary extension, particularly when using the Big Book with the whole group. The suggest that most children had made gains in both arms of the Intervention. However, there were more children in the higher range in Intervention A at post-test

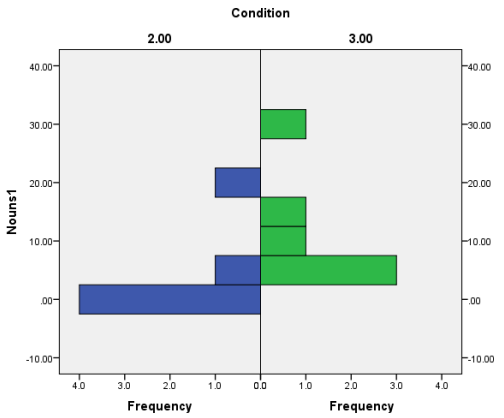


Figure 8.3a Nouns at pre-test

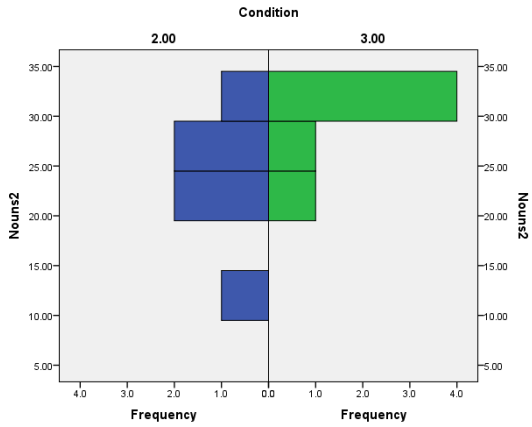


Figure 8.3b Nouns at post-test

Nouns: The grouping was closer at post-test, indicating that most children had made gains and learned most words in both arms of the intervention. Nevertheless, there were more children in the higher range in Intervention A at post-test.

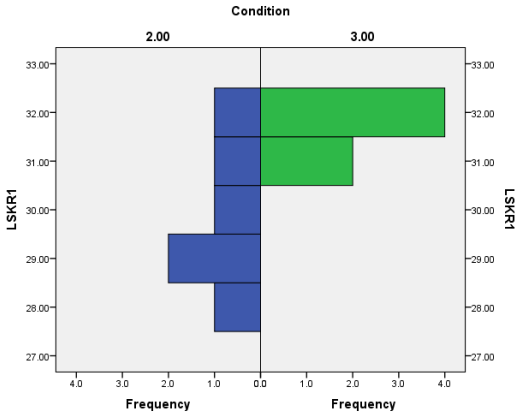


Figure 8.4a LSK at pre-test

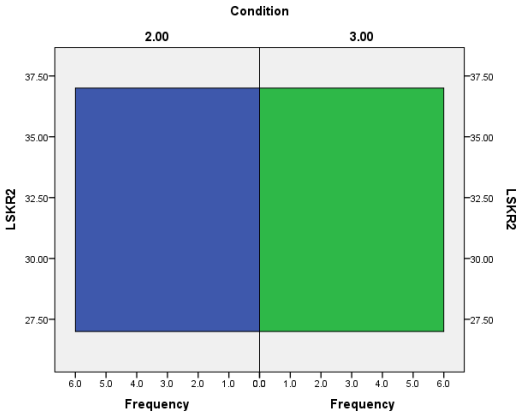


Figure 8.4b LSK at post-test

Letter Sound Knowledge: This graph shows the ceiling effect for this measure (they are all at the top). A more useful measure would have been a standardised test that included more of the complex grapheme-phoneme correspondences across a broader age range, although there does not appear to be such a test at the present time.

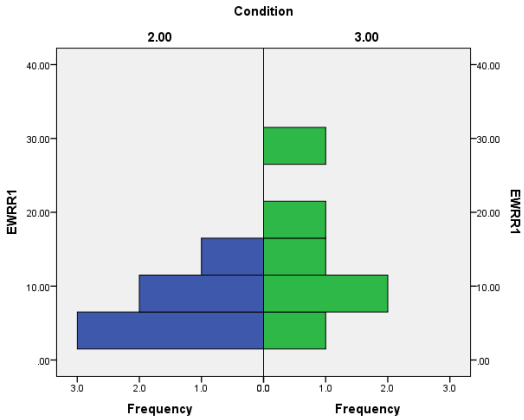


Figure 8.5a EWR at pre-test

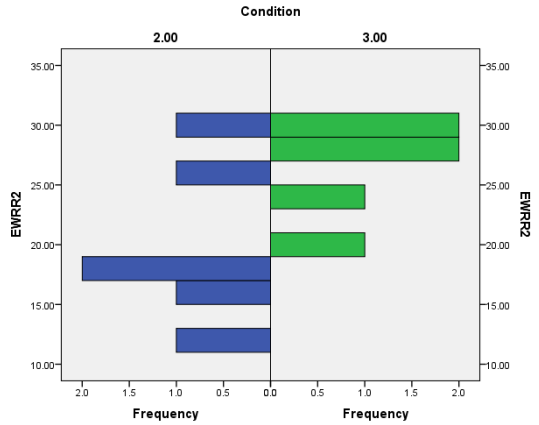


Figure 8.5b EWR at post-test

Early Word Reading: For this measure there was a greater spread at post-test, although again there were greater numbers in the higher range in Intervention A. This measure was also subject to ceiling effects. A more useful test for this older

age group would have been a graded reading test designed for a wider age range, such as the Schonell graded reading test (Schonell, 1971).

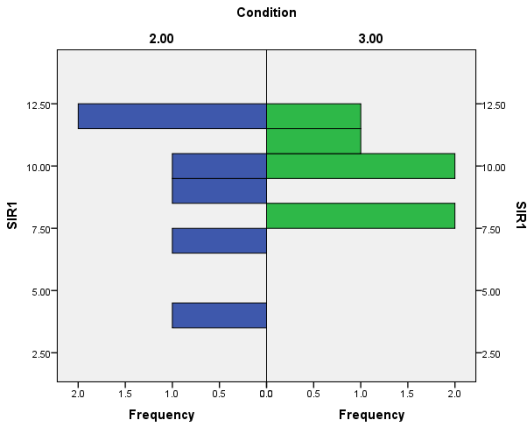


Figure 8.6a SI at pre-test

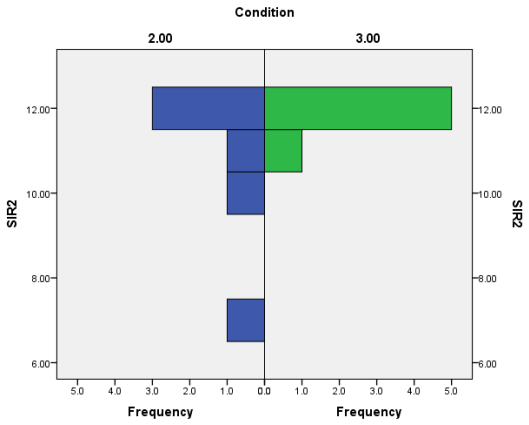


Figure 8.6b SI at post-test

Sound Isolation: For this measure, scores were mostly grouped at the higher end in both conditions at post-test, although there were greater numbers in Intervention A.

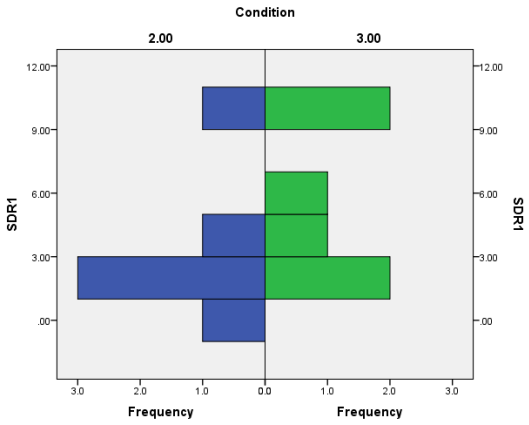


Figure 8.7a SD at pre-test

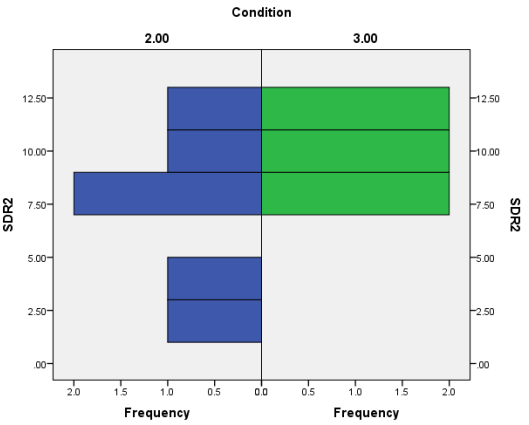


Figure 8.7b SD at post-test

Sound Deletion: For this measure, all the children in Intervention A were grouped at the high end at post-test. Ceiling effects were visible.

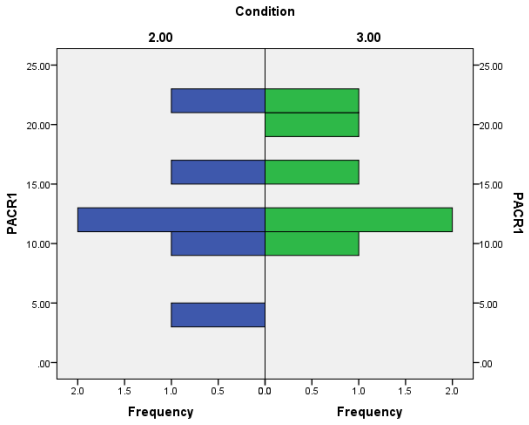


Figure 8.8a Phoneme Awareness at pre-test

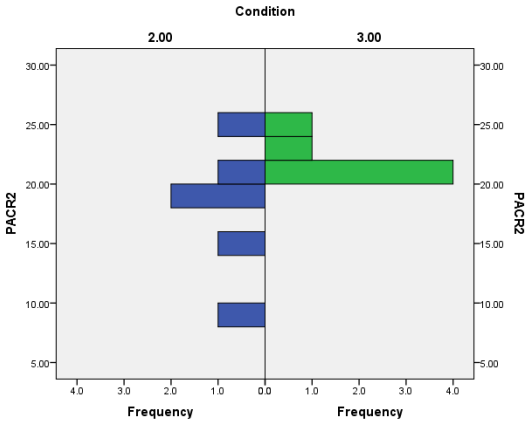


Figure 8.8b Phoneme Awareness at post-test

Phoneme Awareness Combined: For this measure, Sound Isolation and Sound Deletion were combined to give an indication of overall phoneme awareness. This was relatively balanced at pre-test but is clearly grouped towards the higher range in Intervention A at post-test.

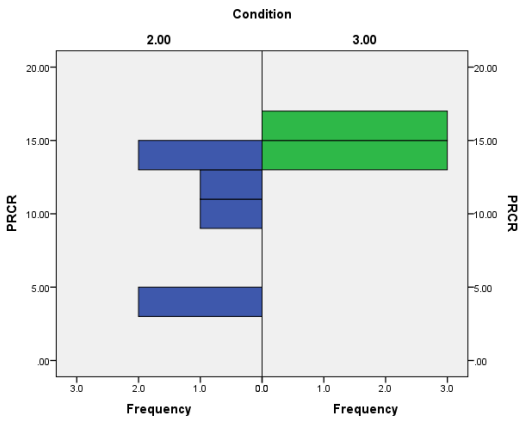


Figure 8.9a PRC at post-test

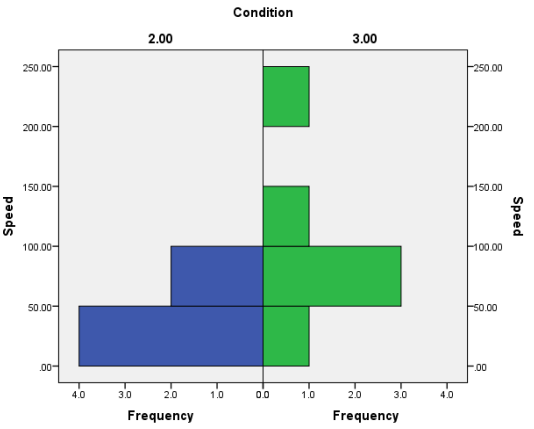


Figure 8.10 RS at post-test

Passage Reading Comprehension and Reading Speed: All the scores were grouped in the higher range for Intervention A for Passage Reading Comprehension. For Reading Speed more children were grouped in the higher range in Intervention A.

Although the small sample size renders most of these differences statistically non-significant, these graphs show clear differences between the two intervention arms. This group of children were all taught by the same highly experienced teacher whose objective will have been to ensure that all the children in her class made gains during the year and yet these differences have emerged from the use of

different vocabulary. These differences may have been more apparent if none of the measures had been subject to ceiling effects. Nevertheless, for all measures except Letter Sound Knowledge, children in Intervention A showed higher post-test groupings than Intervention P.

8.3 Summary

The key findings discussed in this chapter relate to conducting research in an Early Years educational setting and the factors influencing response to intervention. Pros and cons of either a direct approach to schools or asking for volunteers for recruitment were discussed. There was also a brief consideration of possible causes for reluctance to participate in research. This was followed by a more in-depth discussion of variability in compliance to the programme protocol for those schools who did agree to participate and the influence of training and experience on this compliance. The positive impact of compliance was illustrated in a case study of a single school which clearly showed a difference between the two intervention arms at post-test.

A more detailed look at the factors influencing response to intervention revealed a number of points of interest. For children with English as an additional language (EAL), on the whole there was little difference from their peers, but for those who were struggling readers, and had EAL in addition, these children appeared to be at a greater disadvantage in the Control condition and clearly benefitted in both arms of the intervention. There was good evidence to show that all children were capable of learning up to 20 new words for each book, however the type of word had an impact on this. The evidence from the word lists in Intervention P (illustrated in Book 6) suggested that words with close similarity were more difficult to learn than visually distinctive words.

For boys there was evidence of advantages from the intervention, even for struggling readers. This advantage was seen to be counter to national trends in boys' reading scores compared to girls' reading scores. There were children in the Control condition who had higher mean scores at post-test on a few of the measures (Letter Sound Knowledge, Early Word Reading and Sound Deletion). This was for a small group of the highest scorers at pre-test. It is possible that these children would have shown high scores regardless of condition. There may have been a similar effect for Children in Intervention P for Early Word Reading who had

the highest scores at pre-test. For Letter Sound Knowledge and Sound Isolation children who had lowest scores at pre-test performed better in Intervention P. However, for the majority of children across the ability ranges and ages, performance was best in these measures for children in Intervention A.

Chapter Nine

Discussion

The original contribution to knowledge presented here is an answer to the main research question for this Thesis, 'comparing the use of text which is within a child's current decoding ability with text which goes beyond it', which could only be answered in the light of the secondary question regarding teaching methods. This led firstly to an analysis of the literature relating to the phonics debate, and secondly to the inclusion of mixed teaching methods as part of the Reading Programme used in the research. The type of text used for teaching beginning reading is largely determined by the teaching approach used. For this research, the instructional texts were carefully designed to use either: vocabulary which should have been within the children's decoding ability, or vocabulary which went beyond this, including words that could not be sounded out and were multisyllabic. This Chapter starts with a short summary of the literature relating to the phonics debate and the 'real' books versus basal readers debate, as well as the design of the intervention. It includes a discussion of the two research questions, and implications of the results in respect of the gender gap, and for struggling readers. The Chapter concludes with a discussion of the implications of these results for policy, and suggestions for future research.

9.1 The Phonics Debate

In Chapter 1, teaching methods were considered, and focused mainly on the polarised debates between the skills-based and meaning-based, part-to-whole and whole-to-part, and phonics and whole-word approaches to teaching reading. A summary of this evidence echoes the views expressed in: the Bullock Report of 1975 (DES, 1975), the US review by the National Reading Panel in 2000 (NICHHD, 2000) and the Torgerson Review in 2006 (Torgerson et al, 2006), which hold that no single approach is the solution to beginning reading, but that teachers should be able to use all the tools at their disposal. The majority of evidence, presented in the research literature, demonstrates that children benefit from the use of a mix of skills-based and meaning-based approaches, and a mix of phonics and whole-word approaches. Further, there is little evidence, outside the Clackmannanshire study (Johnston & Watson, 2005), to suggest that synthetic phonics has greater value

than any other form of phonics teaching, although this is the approach currently required by statute in English schools.

9.2 Real Books versus Basal Readers

The existing evidence relating to the main research question, regarding instructional reading texts, was considered in Chapter Two. Very little research has as yet been conducted in this area, and what has is mainly focused on the comparison of predictable and non-predictable text, rather than specifically decodable or non-decodable text. Assumptions have been made by those who advocate a skills-based approach, that decodable text must be more appropriate. The research presented here does not support this assumption.

In addition, assumptions have been made by advocates of meaning-based approaches that the use of 'real books' must be more appropriate. By emphasising that children can learn from reading in context, they consider 'real books' to be the preferred instructional text. However, there is little clear evidence to support this assumption either. There is evidence indicating that some children learn better when using predictable, rather than non-predictable text, but much of the research literature is concerned only with word recognition rather than comprehension. There is no clear picture regarding instructional text within the current body of literature. Much of what is claimed has been based on assumption rather than empirical research.

9.3 The Intervention

The main research question led to the development of two parallel reading schemes, deployed in two intervention arms: Intervention A (using non-decodable vocabulary and mixed teaching methods) and Intervention P (using phonically decodable words and mixed teaching methods). In order to compare the effect of the use of different kinds of vocabulary, children in both intervention arms were required to learn large numbers of words (up to 240) that could be recognised easily. The activities for learning these words were all games that were played in groups. Some of these games were adaptations of traditional games and others, such as 'Pento', were developed especially for the reading scheme.

In the design of the activities, although the main objective was simply for the children to learn a particular set of sight vocabulary, consideration was given to

the requirements of the Early Years Foundation Stage curriculum. More importantly, a number of contemporary theories of learning were applied to the designs. These included the role of play and group activity in learning, joint attention and motivation (Mayer, 2011). A particular area of focus was on the extension of children's oral vocabulary to aid the development of world and background knowledge to support inference generation for comprehension. Models of eye tracking and theories related to the importance of initial and final letter positions informed the design of a number of activities. Models of reading, such as the connectionist model, statistical learning from repeated patterns, and learning through analogy also informed the design of the games and activities.

All three studies followed the same design, with three arms to each trial (two interventions and one Control). Both intervention arms had books and activities which were identical in all ways except for the choice of vocabulary. The same standardised assessments were used for all three trials, although Study 1 did not include the Sound Isolation or Sound Deletion measures for phoneme awareness as the children were too young. There was variation in both training and fidelity across the trials. Both were more rigorous for Studies 2 and 3. There were high levels of attrition in Study 1 as well as poor compliance. Although there were only low levels of attrition in Study 2, there was poor compliance in some schools, in spite of extra training. This shortcoming will have affected the results; however, it is the nature of this kind of research, undertaken in the natural classroom environment that makes it vulnerable to these effects. The framework constructed (Change Model, see Chapter Four) for the assessment of fidelity for Studies 2 and 3 nevertheless proved effective, and made it possible to measure target outcomes more easily. The framework also proved useful in the design of the observation schedules, by highlighting the target behaviours for the teachers. The data collected, to show the mean number of words children had read per book, revealed useful information as to the value of completing all the sessions in the programme, i.e. that higher fidelity led to improved results.

The main limitations of these studies were partly the result of having only a single researcher; as a consequence, training was limited for the most part to one session pre-trial and short supporting visits. It is possible that there may have been less attrition, and increased fidelity and compliance, if teachers had received more support. Simple measures of attendance may not have conveyed a true picture of

the extent to which individuals received the intervention; it is possible that some children may have received less instruction, relevant conversation or opportunities to respond (Kaiser & Hemmeter, 2013).

Sample size was also an issue, Studies 2 and 3 had smaller samples than had been expected, which resulted in less power for these trials. The design of Study 1 was weakened by the risk of class results being dependent on their teacher's experience and expertise, rather than the intervention, although randomisation will have ameliorated this affect. In addition, the design was vulnerable to imbalance from attrition. By contrast, Study 2, which used split clusters to avoid both of these issues, ran the risk of cross-contamination. The self-selecting nature of the Control school may have contributed to some sample bias; however, the schools were nevertheless a balanced mix of urban and rural schools.

Statistical significance is dependent on sample size. A very large sample with a very small effect will be statistically significant (leading to type I error); a very small sample can have a very large effect but not be statistically significant (leading to type II error). The *p* value can only give an indication as to whether an effect occurred by chance or not, it does not indicate the size of an effect. Therefore effect sizes (Cohen's *d*) were reported for all measures in all studies.

9.4 Research Question 1: *Are there measurable differences between vocabulary that is within a child's existing decoding ability (Intervention P) and vocabulary which is not so constrained (Intervention A)?*

Except for the British Picture Vocabulary Scale (BPVS) measure in Study 1 only, the mean scores for Intervention A were higher than Intervention P on all measures. For BPVS there were no statistically significant differences between the two interventions for any of the studies. In Study 1 the difference in pre-post effect sizes (d = 0.25) equated to 3 months difference; this difference remained after controlling for age using standard scores (d = 0.22). In Study 2 the difference in pre-post effect sizes (d = 0.67) equated to 8 months difference; this reduced to just 2 months difference after controlling for age using standard scores (d = 0.14). There was no difference in Study 3 (d = 0.02); this remained the same after controlling for age (d = 0.09).

For Letter Sound Knowledge there were no statistically significant differences between the two intervention arms for any of the studies. In Study 1 the difference in pre-post effect sizes (d = 1.54) equated to 19 months difference; this reduced to 4 months after controlling for age (d = 0.30). In Study 2 the difference in pre-post effect sizes (d = 0.15) equated to 2 months difference; this increased to 3 months after controlling for age (d = 0.22). In Study 3 the difference in pre-post-test effect sizes (d = 0.74) equated to 9 months difference; this reduced to 6 months after controlling for age (d = 0.47).

For Early Word Reading in Study 1 there was a statistically significant difference between the two intervention arms after controlling for pre-test scores. However, differences in pre-post effect sizes (d = 0.06) were minimal and remained so after controlling for age (d = 0.11). There was no significant difference in Study 2 after controlling for pre-test scores although post-test scores were significantly different. The difference in pre-post effect sizes (d = 0.86) equated to 10 months difference but this reduced to 0 months after controlling for age using standard scores (d = 0.01). In Study 3 there were no statistically significant differences between the two intervention arms. The difference in pre-post effect sizes (d = 1.68) equated to 20 months difference; this reduced to 6 months after controlling for age (d = 0.48).

For the Sound Isolation measure (not used in Study 1) there was no statistically significant difference between the two intervention arms after controlling for pre-test scores in Studies 2 or 3. In Study 2 the difference in pre-post effect sizes (d = 0.41) equated to 5 months difference, but this reduced to just 1 month difference after controlling for age (d = 0.09). In Study 3 the difference in pre-post effect size (d = 0.49) equated to 6 months difference and this increased to 15 months difference after controlling for age (d = 1.20), with children in Intervention P making greater gains than Intervention A. For this group of struggling readers the phonically decodable vocabulary appears to have been of greater benefit for this element of phoneme awareness.

For the Sound Deletion measure (not used in Study 1) there was no statistically significant difference between the two intervention arms after controlling for pre-test scores in Studies 2 or 3. In Study 2 the difference in pre-post effect sizes (d = 0.32) equated to 4 months difference and reduced to 3 months difference

after controlling for age (*d* = 0.25). In Study 3 the difference in pre-post effect sizes (*d* = 2.60) equated to 30 months differences and 21 months after controlling for age (*d* = 1.73). For this group of struggling readers the non-phonically decodable vocabulary appears to have been of greater benefit for this element of phoneme awareness. Generally scores were lower for this measure across Studies 1 and 2 in all arms of the trial suggesting that children of this age find this test more demanding.

In Study 1 there was no significant difference between the two interventions for Passage Reading Comprehension. In Study 2, with older children, there was a statistically significant difference between the two intervention arms after controlling for the covariate (BPVS at pre-test). The difference in effect size (*d* = 0.75) equated to 9 months difference, increasing to 12 months after controlling for age (*d* = 0.96). There was also a significant difference, after controlling for the covariate, for this measure in Study 3 with struggling readers. The difference in effect size (*d* = 2.10, g = 1.94) equated to 23 months difference; this reduced to 18 months difference after controlling for age (*d* = 1.64, g = 1.52).

For the two non-standardised tests (Nouns in Study 1 and Reading Speed in Studies 2 and 3), there were no significant differences. For the Reading Speed in Study 2, the difference in effect size (*d* = 0.51) equated to 6 months difference. In Study 3 the difference in effect size (*d* = 0.37) equated to 5 months difference.

There were significant or measurable differences between Intervention A and Intervention P on almost all the measures. Statistically significant differences were only observed in Study 1 for EWR and in Studies 2 and 3 for PRC. The evidence from these studies suggests that children who used the non-decodable vocabulary made greater gains than those who used phonically decodable vocabulary particularly for reading comprehension. The comparative gains, across the measures used, between Intervention A and P varied considerably between the studies and between measures but for the majority of these measures children in Intervention A made greater gains than children in Intervention P. This may seem counter intuitive, given that in the past assumptions have been made that children would make better progress using vocabulary that is within their current decoding ability. It does, however, add to the body of evidence which suggests that reading predictable text increases comprehension in early reading (Bus & van Ijzendoorn,

1999; Hatcher et al, 1994; Mesmer, 2001; Morris et al, 2000, Mosely, 2004). This is an area that warrants further investigation.

9.5 Research Question 2: *Are there measurable differences when comparing a synthetic phonics only approach with a mixed teaching approach?*

In order to make comparisons of teaching methods two contrasts were made, firstly the mixed teaching approach with a non-phonically decodable vocabulary (Intervention A) compared to synthetic phonics only, and secondly the mixed teaching approach with phonically-decodable vocabulary (Intervention P) compared to synthetic phonics only.

Comparing Intervention A with the Control condition

Mean scores at post-test were higher for children in Intervention A than in the Control condition on every measure in all three studies. For British Picture Vocabulary Scale in Study 1 there was no statistically significant difference after controlling for pre-test scores. The difference in pre-post effect sizes (d = 0.09) equated to just 1 month difference and there was very little difference after controlling for age (d = 0.11). In Study 2 there was a significant difference after controlling for pre-test scores. The difference in pre-post effect sizes (d = 0.92) equated to 11 months difference, but this reduced to 2 months difference after controlling for age (d = 0.17). In Study 3 ANCOVA could not be run for this measure. The difference in pre-post effect sizes (d = 0.98) equated to 12 months difference, but this reduced to 2 months difference after controlling for age (d = 0.16).

For Letter Sound Knowledge, ANCOVA could not be run in any of the studies. In Study 1 the pre-post difference in effect sizes (d = 0.99) equated to 12 months difference, this reduced to 4 months difference after controlling for age (d = 0.30). In Study 2, pre-post effect sizes (d = 0.45) equated to 6 months difference and increased to 7 months difference after controlling for age (d = 0.53). In Study 3, pre-post effect sizes (d = 0.95) equated to 11 months difference, but this reduced to 6 months after controlling for age (d = 0.47). The Letter Sound Knowledge test was subject to ceiling effects which is likely to have affected the measurable impact of the intervention in Studies 2 and 3; in Study 1 the children were younger and had not yet learned all their letters.

For Early Word Reading in Study 1 there was a statistically significant difference after controlling for scores at pre-test. The difference in pre-post-test effect sizes (d = 0.48) equated to 6 months difference, this reduced to 4 months after controlling for age (d = 0.31). In Study 2 ANCOVA could not be run. The difference in pre-post-test effect sizes (d = 0.30) equated to 4 months difference, reducing to 3 months after controlling for age (d = 0.20). In Study 3 there was no statistically significant difference after controlling for pre-test scores. The difference in pre-post effect sizes (d = 1.70) equated to 20 months difference, this reduced to 11 months after controlling for age (d = 0.88). The test of Early Word Reading was also subject to ceiling effects, which may have affected results for Study 2, as the children were older.

Both the Sound Isolation and Sound Deletion tests were also subject to ceiling effects. In Study 2 ANCOVA could not be run for Sound Isolation or Sound Deletion. In Study 3 there was no statistically significant difference after controlling for pre-test scores for Sound Isolation and ANCOVA could not be run for Sound Deletion. In Study 2, for Sound Isolation, the difference in pre-post-test effect sizes (d = 0.82) equated to 10 months difference, but this reduced to 1 month after controlling for age (d = 0.09). In Study 3, the difference in pre-post effect sizes (d = 2.48) equated to more than 24 months, but this reduced to just 2 months after controlling for age (d = 0.15). For Sound Deletion in Study 2 the difference in pre-post effect sizes (d = 1.00) equated to 12 months difference, reducing to 3 months after controlling for age (d = 0.20). In Study 3 the difference in pre-post effect sizes (d = 4.19) equating to 48 months, reduced to 27 months after controlling for age (d = 2.49). The children in Study 3 were struggling readers or children who had below average scores at pre-test and this will have had an impact on effect sizes.

The Passage Reading Comprehension test was not subject to ceiling effects and was used across all studies. In Study 1 there was a significant difference after controlling for the covariate at pre-test (BPVS). The difference in effect size at post-test (d = 0.51) equated to 6 months difference, and this reduced to 5 months after controlling for age (d = 0.42). In Study 2 there was no significant difference after controlling for the covariate at pre-test (BPVS). The difference in effect size at post-test (d = 1.18) equated to 14 months difference, this reduced to 12 months after controlling for age (d = 1.00). In Study 3, there was a significant difference after controlling for the covariate. The difference in effect size at post-test (d = 1.03)

equated to 12 months difference, reducing to 10 months after controlling for age (*d* = 0.86).

For the non-standard Nouns test in Study 1, there was a statistically significant difference after controlling for the covariate (BPVS at pre-test). The difference in effect size at post-test (*d* = 0.53) equated to 7 months difference. For the non-standard Reading Speed test in Studies 2 and 3 there were no significant difference after controlling for the covariate. In Study 2, the difference in effect size at post-test (*d* = 0.64) equated to 8 months difference and in Study 3 the difference (*d* = 0.16) equated to 2 months difference.

There were measurable differences between the children in the Control condition and children who were taught using mixed methods in combination with the non-decodable vocabulary (Intervention A). These differences represent gains for children in Intervention A compared to children in the Control condition, across all measures, ranging from just 1 month to 27 months (using standard scores controlling for age).

Comparing Intervention P with the Control Condition

Mean scores at post-test were higher in Intervention P than the Control condition for all measures in Study 1. In Study 2 this was the case for all measures except Early Word Reading. In Study 3, the Control had higher post-test mean scores for the British Picture Vocabulary Scale, Early Word Reading, Passage Reading Comprehension and Reading Speed. For the British Picture Vocabulary Scale there were no statistically significant differences between these two conditions after controlling for pre-test scores in any of the studies. The differences in pre-post effect sizes in Study 1 (*d* = 0.34) equated to 4 months difference and this remained the same after controlling for age (*d* = 0.33). In Study 2 this difference (*d* = 0.24) equated to 3 months difference, rising to 4 months after controlling for age (*d* = 0.31). In Study 3, the difference (*d* = 0.96) equated to 12 months difference, but this reduced to just 1 month after controlling for age (*d* = 0.07).

For Letter Sound Knowledge in Study 1 there was no significant difference between the groups after controlling for pre-test scores and ANCOVA could not be run in Studies 2 and 3. The differences in pre-post effect sizes in Study 1 (*d* = 0.55) equated to 6 months difference, but this reduced to just 1 month after controlling

for age (*d* = 0.08). In Study 2 this difference (*d* =0.30) equated to 4 months difference which increased to 9 months after controlling for age (*d* = 0.75). In Study 3, the difference (*d* = 0.21) equated to 3 months difference, but this reduced to nothing after controlling for age (*d* = 0.00).

For Early Word Reading in Study 1 there was no significant difference between the groups after controlling for pre-test scores and ANCOVA could not be run in Studies 2 and 3. The differences in pre-post effect sizes in Study 1 (*d* = 0.42) equated to 5 months difference, reducing to 3 months after controlling for age (*d* = 0.20). In Study 2, the difference (*d* = 0.56) which equated to 7 months difference, reducing to 3 months after controlling for age (*d* = 0.21), represented greater gains for the Control condition. In Study 3 there was no difference in raw scores (*d* = 0.02) but this increased to 5 months after controlling for age (*d* = 0.40), representing greater gains for Intervention P.

In Study 2 ANCOVA could not be run for Sound Deletion or Sound Isolation. In Study 3, ANCOVA could not be run for Sound Isolation but there was a significant difference after controlling for pre-test scores for Sound Deletion. In Study 2 the differences in pre-post effect sizes for Sound Isolation (*d* = 0.41) equated to 5 months difference, reducing to 2 months after controlling for age (*d* = 0.13). In Study 3, this difference (*d* = 1.99) equated to 24 months difference, reducing to 16 months after controlling for age (*d* = 1.35). For Sound Deletion in Study 2, this difference (*d* = 0.68) equated to 8 months difference, reducing to 6 months after controlling for age (*d* = 0.45). In Study 3 the difference (*d* = 1.59) equated to 19 months difference, reducing to 9 months after controlling for age (*d* = 0.76).

For Passage Reading Comprehension there was no significant difference between Intervention P and the Control condition after controlling for the covariate at pre-test (BPVS) in any of the studies. In Study 1 the difference in effect size at post-test (*d* = 0.51) equated to 6 months difference, reducing to 5 months after controlling for age (*d* = 0.36). In Study 2 this difference (*d* = 0.48) equated to 6 months difference, reducing to 4 months after controlling for age (*d* = 0.28). In Study 3, the difference (*d* = 0.32) equated to 4 months difference, increasing to 6 months after controlling for age (*d* = 0.47), representing greater gains for the Control condition.

For the non-standard Nouns test in Study 1 there was no statistically significant difference between these groups after controlling for the covariate at pre-test (BPVS). The difference in effect size at post-test (d = 0.48) equated to 6 months difference. For the non-standard Reading Speed test in Studies 2 and 3 there was no significant difference after controlling for the covariate at pre-test (BPVS). In Study 2 the difference in effect size at post-test (d = 0.17) equated to 2 months difference. In Study 3, this difference (d = 0.15) equated to 2 months difference.

There were measurable differences between the children in the Control condition and children who were taught using mixed methods in combination with the decodable vocabulary (Intervention P). These differences represent gains for children in Intervention P compared to children in the Control condition, across the majority of measures, with the exception of Early Word Reading in Study 2, and British Picture Vocabulary Scale, Letter Sound Knowledge and Passage Reading Comprehension in Study 3. The differences ranged from just 1 month to 16 months (using standard scores controlling for age). The evidence from these studies suggest that, in agreement with large-scale reviews, such as that undertaken by Torgerson et al (2006), a mixed approach to teaching has greater benefits than synthetic phonics alone.

9.6 Gender

According to an Ofsted report published in 2012, there was a 9% difference between girls and boys in reading on the Early Years Foundation Stage Profile (age 5) and an 8% difference still at the end of Key Stage 2 (Ofsted, 2012). Given the concerns raised by Ofsted, regarding the lower attainment of boys in reading, it was important to evaluate the Reading Programme used in the three trials detailed in this research in terms of the impact they had on gender differences.

In Study 3, the imbalance in gender rendered a gender analysis of each arm of the trial unreliable and the intervention arms were therefore combined in the gender analysis for that study. Across the other two studies, girls in Intervention A were the highest scoring group on all but one measure in Study 1 (for the British Picture Vocabulary Scale) and in Study 2 for Early Word Reading, Sound Isolation and Reading Speed. The implication of these results is that girls in Intervention A

benefitted from the use of the non-decodable vocabulary, in combination with a mixed methods teaching approach for these measures in particular.

For girls in Intervention P, the results were more mixed. In Study 1 they scored higher than the girls in the Control condition on all measures, but in Study 2 they scored lower for British Picture Vocabulary Scale, Early Word Reading, Passage Reading Comprehension and Reading Speed. In addition, in Study 2, they scored lower than the boys in Intervention P for the British Picture Vocabulary Scale, Letter Sound Knowledge and Passage Reading Comprehension. One possible implication here is that girls in Intervention P in Study 1 benefitted from the mixed teaching methods, but that this was not the case for the girls in Study 2. Alternatively, the higher scores for the boys in Study 2 may reflect the benefit of the mixed teaching methods for boys.

Boys in Intervention A scored higher than the boys in both Intervention P and the Control condition on most measures in Study 1 (with the exception of the British Picture Vocabulary Scale) and all measures in Study 2. This suggests that boys benefitted most from the non-decodable vocabulary in combination with a mixed methods teaching approach. Boys in Intervention P scored higher than boys in the Control condition on all measures in Studies 1 and 2. Boys in both intervention arms out-performed boys in the control condition. The most significant narrowing of the gender gap was in Intervention A for the Letter Sound Knowledge in Study 1, Passage Reading Comprehension in Study 2, and in Intervention P for Reading Speed in Study 2.

In Study 3, boys in the combined intervention arms outperformed boys in the control condition on all measures except Early Word Reading (although this changed after controlling for age). Girls in the combined intervention arms outperformed girls in the control condition on all but two measures (Early Word Reading and Reading Speed).

Overall, girls in these studies seem to have benefitted most from the use of non-decodable vocabulary, and boys seem to have benefitted from both the mixed teaching methods and the non-decodable vocabulary used. There was a positive impact from the use of non-decodable vocabulary and mixed teaching methods in addition to synthetic phonics on both word decoding and reading comprehension.

9.7 Struggling Readers

Study 3 involved a small group of struggling readers who used the same materials as those in Study 2, but were taught in smaller groups, and with greater frequency, over a shorter period of time. The aim was to explore how the Reading Programme might benefit struggling readers if used as a short-term intervention. The group was small (12), with more boys than girls, a ratio of 2:1.

Because of the imbalance, results for gender were not split between Interventions A and P. Within the Control condition, the gender gap (girls doing better than boys) was observed for all the measures used. By contrast, within the intervention conditions, boys showed advantage over girls on a number of measures (Early Word Reading after controlling for age, Sound Isolation and Reading Speed). This indicates that boys who are struggling readers benefitted from the mixed teaching methods used in both arms of the intervention.

In respect of Research Question 1, the results demonstrated that, other than for the British Picture Vocabulary Scale, children in Intervention A made greater gains than children in Intervention P. This suggests that even for struggling readers, a non-phonically-decodable vocabulary benefits reading progress in phoneme awareness, word decoding and comprehension. In respect of Research Question 2, the results demonstrated that, other than for the British Picture Vocabulary Scale, Early Word Reading, Passage Reading Comprehension and Reading Speed in Intervention P, children in the intervention arms made greater gains than children in the Control condition. This suggests that, for struggling readers, the mixed teaching methods used, in addition to synthetic phonics, benefits reading progress in phoneme awareness, word decoding and receptive vocabulary, especially in combination with non-decodable vocabulary.

Taken together, these results demonstrate that the children in Intervention A, made greatest gains compared to the children in the Control group for the measures used in this trial. In addition, the boys in the combined intervention made greater gains than boys in the Control group. This also seems counter intuitive, given that in the past assumptions have been made that struggling readers should be given more practice in synthetic phonics.

The results in this study for the tests of Sound Isolation (the ability to recognise and isolate the sounds of letters in initial and final letter positions) and Sound Deletion (the ability to separate onset and rime) support the findings of McGeown and Medford (2014) that children's reading skills reflect the kind of instruction they have received. Children in Intervention P (decodable vocabulary) made greater pre-post-test gains for the Sound Isolation test and children in Intervention A (non-decodable vocabulary including multi-syllabic words) made greater pre-post-test gains for the Sound Deletion test.

9.8 Summary of the Evidence

The evidence from these three studies clearly demonstrates that children benefit from the use of other teaching approaches in addition to synthetic phonics. Children of all abilities and both genders benefitted in terms of word decoding, phoneme awareness and reading comprehension, from a mix of skills-based and meaning-based approaches. The evidence presented here questions the assumption that children learn better with text which is within their current decoding ability. Instead, it suggests that children of all abilities and both genders benefit from using unconstrained, natural language that may go beyond their current decoding ability. This research adds to the body of literature regarding the use of instructional texts for beginning readers, by challenging long held assumptions, but without suggesting a return to the unstructured use of 'real' books.

9.9 Feedback

In addition to numerical data from the assessments administered, some feedback was also collected from teachers describing their experiences in delivering the weebee Reading Programme and the impact that they felt the programme had made on the children in general. On the whole there were positive comments regarding the teacher's Manual, the training, and the support. These comments ranged from satisfactory to very useful and the majority of teachers considered the resources easy to use and understand. There was unanimous agreement that the children had enjoyed all the activities, books, songs and the storybook characters, the weebees. Teachers commented that the children looked forward to the sessions, and were highly motivated by the activities and characters. Some parents had also commented positively on the children's response to the programme. Four of the participating schools stated that they intended to run the programme again

for subsequent cohorts. All the schools chose to keep all the books and resources. Teachers of the children in Study 3 stated that most of the children had made evident progress, especially in confidence and comprehension.

The only negative comments from schools were either to do with time constraints, which were felt by most schools, external pressures from imminent Ofsted inspections, and internal pressures from members of staff who felt that they should be following the National Curriculum guidelines to use synthetic phonics only.

9.10 Implications for Policy and Future Research

Results from all three studies indicate that children of all abilities, and boys in particular, benefit from a more eclectic approach to the teaching of reading than is currently being advocated in schools. This supports previous findings which have reported that there is no evidence that synthetic phonics alone is superior to any other method and goes further, by demonstrating that synthetic phonics alone is less effective than in combination with other teaching methods, particularly for boys. The results presented here also support the view that there is an optimal mix of skills-based and meaning-based approaches which varies for individuals (Connor et al, 2004; Dombey, 1999; Ellis & Moss, 2014; Flynn, 2007; Huang, 2014; Wilson & Comar, 2008). The implication is that there should be a re-evaluation of the current focus on synthetic phonics.

The results presented here from comparing the use of phonically decodable vocabulary in reading texts with non-phonically decodable vocabulary indicate that children, of all abilities and boys in particular, benefit from the use of texts which go beyond their current decoding ability. The implication here is that there should be a re-evaluation of the criteria for book publishers of early reading texts. The books used in Intervention A, although not constrained by the requirements of the core criteria for publishers, were nevertheless part of a structured reading scheme as opposed to 'real' books which have not been written as part of an instructional series. What has been demonstrated is that a structured reading scheme can be effective, which includes non-phonically decodable vocabulary, in predictable and meaningful text with illustrations, in which children can use context and grammar to assist their reading.

Areas for future research include an exploration of possible reasons why the use of non-decodable texts has been shown to improve some reading skills and comprehension in particular. Another question that warrants investigation is why boys should benefit less than girls from a synthetic-phonics-only approach. Not all children responded equally to the intervention, and therefore research that may identify which children benefit most would be valuable for teachers, in terms of future application of reading schemes such as the weebee Reading Programme. It may also be appropriate to consider a revised model of reading which incorporates a first stage which acknowledges recognition of whole words, but at the same time acknowledges recognition of letters and their sounds within words, rather than in isolation.

Appendices

Appendix A Outline of the ɯeebee Reading Programme from the Manual

Summary of Week 1 (of 12)

Objectives: • Introduction to 'ɯeebee's • Learn 'ɯeebee' song – The Valley • Introduction to games for book 1 • Introduction to phonics worksheets • Read book 1 (Session 5 only)
Resources: • 'ɯeebee' CD and 'big book' • Bingo game • Jig-words • Dominoes • Memory game • Fishing game • Snap • Phonics worksheets, scissors and felt pens • Book 1 • Tick sheets for Book 1 (one per child) and word list
Whole Group Activity: • 'ɯeebee' song • Look at 'big book'
Small Group Activities: • Bingo • Jig-words • Dominoes • Memory game • Fishing game • Snap • Phonics worksheets
Plenary: • Looking at the week's phonics worksheet and jig words

Appendix B

List of teaching resources and example lesson plan from the Teacher's Manual

Week 1

Session 1

Objectives:

1. To introduce the '*weebees*'
2. To learn the '*weebee*' song 'The Valley'
3. To introduce word bingo, dominoes and jig words

Resources:

1. '*weebee*' CD and 'big book'
2. Bingo game
3. Jig words
4. Dominoes
5. Draw-through phonics sheets (_an), scissors and felt pens

Activity 1 (all the children)

1. Listen to the '*weebee*' song on the CD and look at the first picture in the 'big book'.
2. Ask the children if they have ever seen a big pond or lake. (What was growing around it? Were there lots of trees? Were there any boats?)
3. Organise the children into groups of 3-4 for activities 2, 3 and 4.

Activity 2: Bingo (maximum 4 children)

1. Select 4 children to play bingo.
2. Give each child a board.
3. Put blank white cards where the children can reach them.
4. TA to take one word at a time from the bag.
5. TA to say aloud the selected word and show all the children in the group.
6. If the word appears on a child's board they should cover it with a blank card.
7. Children do not need to say the word, just find it on their board.
8. The winner is the first to cover their board.
9. Repeat the game swapping boards if time is left within the ten minutes.

Activity 3: Jig-words (maximum 4 children – 2 pairs)

1. Put the children in pairs where possible.
2. Ask the children to work together to find the matching two parts of a word.
3. If there is time left, separate the jig-words; remix them and begin again.

Activity 4: Dominoes (4 children)

1. Put the children in pairs.
2. Give each pair a pack of dominoes.
3. Each pair halves the pack and has a pile of half the cards each.
4. A first card is chosen and then the children take turns to place a card next to a matching word, one word at a time. (This will need either a large space on a table or it could be laid out on the carpet.)
5. The winner is the first to get rid of all their cards or the one who has the least number of cards when no more can be played.

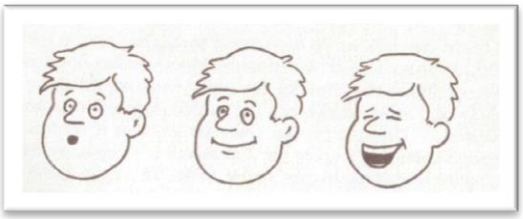
Activity 5: Phonics

Try to use one table for this activity so that the children can work together.

1. Give each child a cutting out sheet for the 'draw-through' activity (_**an**).
2. Give each child a pair of scissors and a coloured felt pen.
3. Each child cut out along the black lines. (Children may need assistance with parallel lines.)
4. With the coloured pen, write over the grey letters.
5. In pairs where possible, draw through strips for friend to see the words appearing.

Concluding Activity (all the children)

1. Choose one of the draw-through cards that the children have made.
2. Show the children how it works and that they can see new words appearing.
3. Whole group say the words together as they appear in the 'window'.



Appendix C Example Record Sheets (Session, Individual and Teacher checklists)

Session Record Sheet

Name	Bingo	Jig words/ Happy families	Dominoes	Memory	Fishing	Snap	Phonics

Individual Reading Record

Name:

Book 1	*Book 2*	*Book 3*	*Book 4*	*Book 5*	*Book 6*	*Book 7*	*Book 8*
a	by	an	as	feet	back	bed	all
and	flower	brown	been	blue	crash	bit	band
asleep	help	dark	big	bug	cry	box	bang
at	her	deep	bow	duck	fall	button	be
can	it	down	but	go	flash	fix	bees
green	on	from	for	had	fly	fluff	black
Grog	petal	fun	grass	Jig	get	Flup	clang
happy	pink	have	grip	land	metal	going	hang
he	Pip	here	has	luck	miss	good	hit
his	play	hoots	long	no	moon	grey	huff
in	put	hop	low	not	shell	hat	puff
is	see	jumps	Mop	of	ship	man	Saff
look	she	now	onto	old	sky	off	sang
nest	sink	owl	sing	rock	small	run	spy
pond	the	peep	song	sack	star	sit	stay
red	there	roots	swim	slip	tall	so	stuff
sleep	under	skip	swing	soft	try	tell	day
this	will	they	too	stick	upset	them	to
tree	with	Tod	well	top	wall	very	twig
weebee	you	up	yellow	wings	Zon	wood	way

Teacher Checklist for Week 1

Date:

Name:

Did I	**YES**	**NO**
1. Give every child a chance to play bingo?		
2. Give every child a chance to play Jig words?		
3. Give every child a chance to play dominoes?		
4. Give every child a chance to play memory game?		
5. Give every child a chance to play fishing game?		
6. Give every child a chance to play snap?		
7. Teach the 'ɯeebee' song?		
8. Discuss the pages in the 'big book'?		
9. Use the phonic sheets for _an?		
10. Hear every child read?		
11. Record on every child's word list any errors?		

Appendix D Example Feedback Questionnaire

Feedback May 2014

(Please circle the most representative number)

1. How useful have you found the manual?

1	2	3	4	5
Not at all	sometimes	satisfactory	useful	very useful

2. How useful was the initial training?

1	2	3	4	5
Not at all	sometimes	satisfactory	useful	very useful

3. How easy were the resources to use?

1	2	3	4	5
Not at all	sometimes	satisfactory	easy	very easy

4. Did the children enjoy the games?

1	2	3	4	5
Not at all	sometimes	satisfactory	enjoyed	enjoyed a lot

5. Did the children enjoy the phonic activities?

1	2	3	4	5
Not at all	sometimes	satisfactory	enjoyed	enjoyed a lot

6. Did the children enjoy the books?

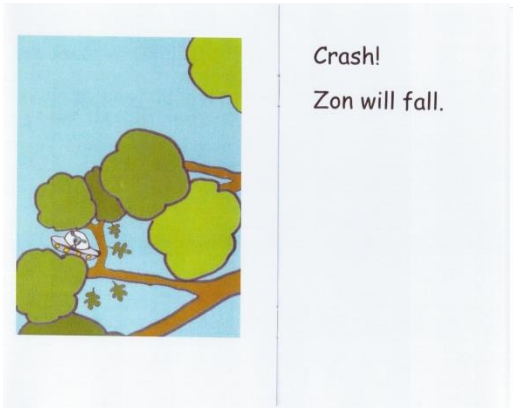
1	2	3	4	5
Not at all	sometimes	satisfactory	enjoyed	enjoyed a lot

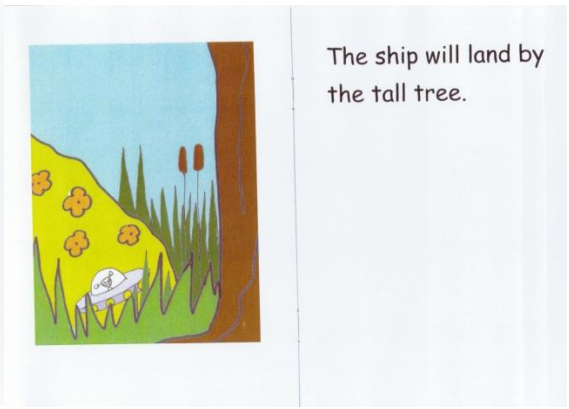
Additional comments:

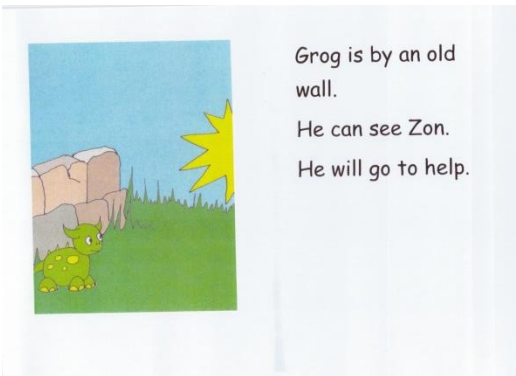
Appendix E Lesson-to-text match

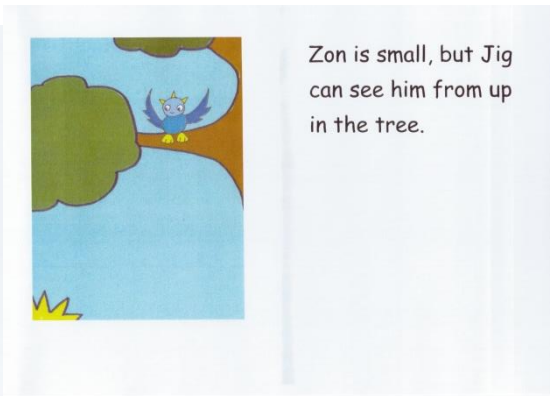
Example page from the Big Book

Example pages from Book 6, showing the words from the _all family being used in context

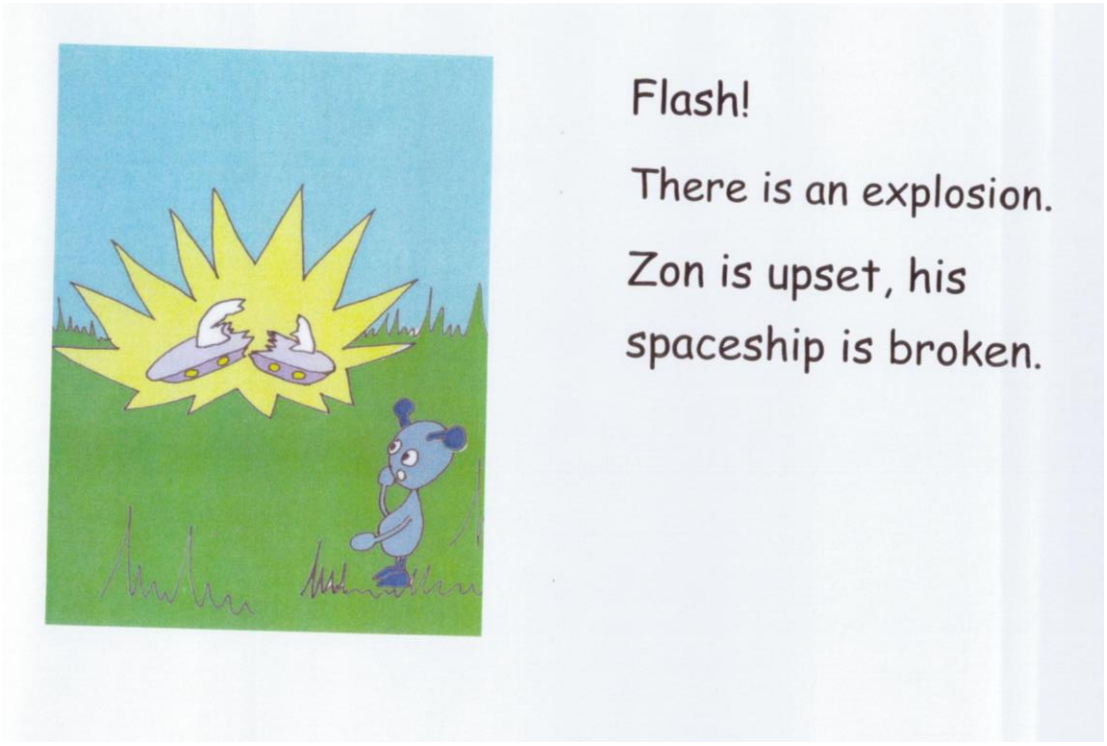








Appendix F Page from Book 6a (Intervention A) showing use of multisyllabic words



Flash!

There is an explosion.

Zon is upset, his spaceship is broken.

Appendix G

The introductory page and pages for weeks 2 and 9 from the Teacher's Manual (revised for Study 2)

Manual for using the 'weebee' Reading Programme

Introduction

The fictional characters appearing in this reading programme are tiny imaginary creatures that live in the roots of a tree and are called 'weebees'. By using these tiny fictional creatures we hope to avoid issues of gender and culture and hope that they will be accessible to all children including those who have English as an additional language. The illustrations are intended to assist learning and time should be allowed for children to look carefully at each picture and ask questions if they wish.

General Principles

- All the activities should be fun. There should be no pressure to read words aloud.
- Never ask the children to sound out words (*do not stop them if they choose to).*
- Draw attention to the first letter of words.
- You need to be the one to verbalise the words, the children should not be expected to (*do not stop them if they choose to*).
- Allow the children to help their friends/teammates; it saves you the job!
- The key elements are: repetition, recognition and attention to initial letters.
- Although the sessions are broken down into weeks this is not meant to be a definite time frame. Depending on the age of your children and the number in your class etc, the time for all the children to engage with all the activities and then read through the book could potentially take up to three weeks.
- Be flexible with your own time – the success of the programme is not dependent on adhering to any particular time scale. As the children move through the programme the pace will naturally pick up.

Key components: Vocabulary and use of onset/rime (initial letter sounds and patterns)

Summary of Week 2

Objectives:

- Practise 'weebee' song 'The Valley'
- Introduction to games for book 2
- Introduction to phonics worksheets
- Read book 2 (Session 5 only)

Resources:

- 'weebee' CD and 'big book'
- Bingo game
- Jig-words
- Dominoes
- Memory game
- Fishing game
- Snap
- Phonics worksheets, scissors and felt pens
- Book 2
- Tick sheets for Book 2 (one per child) and word list

Whole Group Activity:

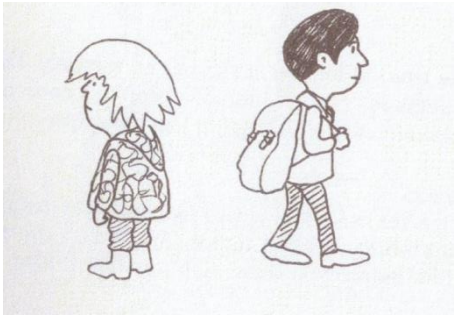
- **'weebee'** song
- Look at 'big book'

Small Group Activities:

- Bingo
- Jig-words
- Dominoes
- Memory game
- Fishing game
- Snap
- Phonics worksheets

Plenary:

- Looking at the week's phonics worksheet and jig words



Week 2

Session 1

Objectives:

1. To practise the 'weebee' song 'The Valley'
2. To use word bingo, dominoes and jig words

Resources:

1. 'weebee' CD and 'big book'
2. Bingo game
3. Jig words
4. Dominoes
5. Draw-through phonics sheets (_in), scissors and felt pens

Activity 1 (all the children)

1. Listen to the 'weebee' song on the CD and look at picture 5 in the 'big book'.
2. Ask the children if they can see the yellow field. Ask if anyone has ever seen a yellow field? Tell them some people come all the way from Japan just to see our yellow fields.
3. Organise the children into groups of 3-4 for activities 2, 3 and 4.

Activity 2: Bingo (maximum 4 children)

1. Select 4 children to play bingo.
2. Give each child a board.
3. Put blank white cards where the children can reach them.
4. TA to take one word at a time from the bag.
5. TA to say aloud the selected word and show all the children in the group.
6. If the word appears on a child's board they should cover it with a blank card.
7. Children do not need to say the word, just find it on their board.
8. The winner is the first to cover their board (it is likely that there will often be a tie).
9. Repeat the game swapping boards if time is left within the ten minutes.

Activity 3: Jig-words (maximum 4 children – 2 pairs)

1. Put the children in pairs where possible.
2. Ask the children to work together to find the matching two parts of a word.
3. If there is time left, separate the jig-words; remix them and begin again.

Activity 4: Dominoes (4 children)

1. Put the children in pairs.
2. Give each pair a pack of dominoes.
3. Each pair halves the pack and has a pile of half the cards each.
4. A first card is chosen and then the children take turns to place a card next to a matching word, one word at a time. (This will need either a large space on a table or it could be laid out on the carpet.)
5. The winner is the first to get rid of all their cards or the one who has the least number of cards when no more can be played.

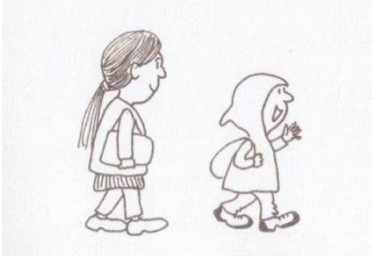
Activity 5: Phonics

Try to use one table for this activity so that the children can work together.

1. Give each child a cutting out sheet for the 'draw-through' activity (**_in**).
2. Give each child a pair of scissors and a coloured felt pen.
3. Each child cut out along the black lines. (Children may need assistance with parallel lines.)
4. With the coloured pen, write over the grey letters.
5. In pairs where possible, draw through strips for friend to see the words appearing.

Concluding Activity (all the children)

1. Choose one of the draw-through cards that the children have made.
2. Whole group say the words together as they appear in the 'window'
3. Choose two of the jig words to look at together.



Session 2

Objectives:

1. To practise the 'weebee' song 'The Valley'
2. To use word bingo, dominoes and jig-words with the rest of the children

Resources:

1. 'weebee' CD and 'big book'
2. Bingo game
3. Jig words
4. Dominoes
5. Draw-through phonics sheets (_in), scissors and felt pens

Activity 1 (all the children)

1. Sing the 'weebee' song.
2. Look at pictures 5 and 6 of the 'big book'.
3. Ask the children about what kind of flowers they have seen. Have they seen a daisy? Have they seen buttercups? Can they think of any pink flowers?
4. Organise the children into groups of 3-4 for activities 2, 3 and 4.

Activity 2: Bingo (maximum 4 children)

1. Select 4 children to play bingo.
2. Give each child a board.
3. Put blank white cards where the children can reach them.
4. TA to take one word at a time from the bag.
5. TA to say aloud the selected word and show all the children in the group.

6. If the word appears on a child's board they should cover it with a blank card.
7. Children do not need to say the word, just find it on their board.
8. The winner is the first to cover their board (it is likely that there will often be a tie).
9. Repeat the game swapping boards if time is left within the ten minutes.

Activity 3: Jig-words (maximum 4 children – 2 pairs)

1. Put the children in pairs where possible.
2. Ask the children to work together to find the matching two parts of a word.
3. If there is time left, separate the jig-words; remix them and begin again.

Activity 4: Dominoes (4 children)

1. Put the children in pairs.
2. Give each pair a pack of dominoes.
3. Each pair halves the pack and has a pile of half the cards each.
4. A first card is chosen and then the children take turns to place a card next to a matching word, one word at a time. (This will need either a large space on a table or it could be laid out on the carpet.)
5. The winner is the first to get rid of all their cards or the one who has the least number of cards when no more can be played.

Activity 5: Phonics (children not currently in one of the other groups)

1. Try to use one table for this activity so that the children can work together.
2. Give each child a cutting out sheet for the 'draw-through' activity (**_in**).
3. Give each child a pair of scissors and a coloured felt pen.
4. Each child cut out along the <u>black</u> lines. (Children may need assistance with parallel lines.)
5. With the coloured pen, write over the <u>grey</u> letters.
6. In pairs where possible, draw through strips for friend to see the words appearing.

Concluding Activity (all the children)

1. Choose one of the draw-through cards that the children have made.
2. Whole group say the words together as they appear in the 'window'
3. Choose two different jig words to look at together.

Session 3

Objectives:

1. To practise 'weebee' song 'The Valley'
2. To use memory game, snap and the 'fishing' game

Resources:

1. 'weebee' CD and 'big book'
2. Memory game
3. 'Fishing' game
4. Snap
5. Draw-through phonics sheets (_in), scissors and felt pens

Activity 1 (all the children)

1. Sing the '*weebee*' song
2. Look at pictures 5 – 7 of the 'big book'
3. Ask the children if they have ever seen a bulrush. Do they know what one is? Show them the bulrush in the picture and show them how high they can grow (about 1 metre). Tell them that they only grow near water and that there have been lots of stories about things found in the bulrushes.
4. Organise the children into groups of 3-4 for activities 2, 3 and 4.

Activity 2: Memory Game (maximum 4 children)

1. (Use half the number of cards for a first game and then use the other half for a second game.) Lay the cards face down in a regular pattern for example: 4 rows of 5 cards.
2. Each child takes a turn to turn over two cards. The TA says each of the words aloud before turning the cards back over. Do not ask children to repeat the word but do not prevent them either.
3. The aim is to find a matching pair of the same word. When a pair is found, the player may have a second go.
4. Order and neatness of layout needs to be maintained, gaps should be left.
5. The winner is the player who collects the most pairs.
6. Speed is not important.

Activity 3: Fishing Game (maximum 4 children)

1. The children work in pairs as two teams.
2. A set of cards with words written on is placed face down but within reach of all the children.
3. A set of words (on fish-shaped cards with metal attachments) is placed centrally in the 'pond' with all the words facing up and visible.
4. One child from the first pair turns over a card and reads the word, without showing the partner. The partner then 'catches' the matching fish word with a magnet fishing rod. The first child shows the word and it is checked against the fish.
5. If it is correct the pair keep the fish, if not it is 'thrown' back into the pond.
6. Within the pair the roles are reversed and then the next pair has their go.
7. The winning team has the most fish.
8. A variation can then be played with children finding matching initial letters

Activity 4: Snap (maximum 4 children – two pairs)

1. The cards are divided equally between the players.
2. The children take turns to turn over a card
3. As soon as a matching card is played the first person to call 'snap' takes the whole pile.
4. The game ends when one player has no cards left.
5. The winner is the player with the most cards.

Activity 5: Phonics

1. Try to use one table for this activity so that the children can work together.
2. Give each child a cutting out sheet for the 'draw-through' activity (**_in**).
3. Give each child a pair of scissors and a coloured felt pen.

4. Each child cut out along the black lines. (Children may need assistance with parallel lines.)
5. With the coloured pen, write over the grey letters.
6. In pairs where possible, draw through strips for friend to see the words appearing.

Concluding Activity (all the children)

1. Choose one of the draw-through cards that the children have made.
2. Whole group say the words together as they appear in the 'window'.
3. Choose two different jig words to look at together.

Session 4

Objectives:

1. To practise 'weebee' song 'The Valley'.
2. To use memory game, fishing game and snap with the rest of the children.

Resources:

1. 'weebee' CD and 'big book'
2. Memory game
3. 'Fishing' game
4. Snap
5. Draw-through phonics sheets (_in), scissors and felt pens

Activity 1 (all the children)

1. Sing the 'weebee' song.
2. Look at pages 5 – 8 of the 'big book'.
3. Ask the children what is the biggest flower they have ever seen.
4. Ask if they have ever seen a sunflower. Ask what the smallest flower they have seen is. Ask them to imagine being so small that a flower seems as big as a tree.
5. Organise the children into groups of 3-4 for activities 2, 3 and 4.

Activity 2: Memory Game (maximum 4 children)

1. (Use half the number of cards for a first game and then use the other half for a second game.) Lay the cards face down in a regular pattern for example: 4 rows of 5 cards.
2. Each child takes a turn to turn over two cards. The TA says each of the words aloud before turning the cards back over. Do not ask children to repeat the word but do not prevent them either.
3. The aim is to find a matching pair of the same word. When a pair is found, the player may have a second go.
4. Order and neatness of layout needs to be maintained, gaps should be left.
5. The winner is the player who collects the most pairs.
6. Speed is not important.

Activity 3: Fishing Game (maximum 4 children)

1. The children work in pairs as two teams.
2. A set of cards with words written on is placed face down but within reach of all the children.

3. A set of words (on fish-shaped cards with metal attachments) is placed centrally in the 'pond' with all the words facing up and visible.
4. One child from the first pair turns over a card and reads the word, without showing the partner. The partner then 'catches' the matching fish word with a magnet fishing rod. The first child shows the word and it is checked against the fish.
5. If it is correct the pair keep the fish, if not it is 'thrown' back into the pond.
6. Within the pair the roles are reversed and then the next pair has their go.
7. The winning team has the most fish.
8. A variation can then be played with children finding matching initial letters.

Activity 4: Snap (maximum 4 children – 2 pairs)

1. The cards are divided equally between the players.
2. The children take turns to turn over a card.
3. As soon as a matching card is played the first person to call 'snap' takes the whole pile.
4. The game ends when one player has no cards left.
5. The winner is the player with the most cards.

Activity 5: Phonics

1. Try to use one table for this activity so that the children can work together.
2. Give each child a cutting out sheet for the 'draw-through' activity (**_in**).
3. Give each child a pair of scissors and a coloured felt pen.
4. Each child cut out along the black lines. (Children may need assistance with parallel lines.)
5. With the coloured pen, write over the grey letters.
6. In pairs where possible, draw through strips for friend to see the words appearing.

Concluding Activity (all the children)

1. Choose one of the draw-through cards that the children have made.
2. Whole group say the words together as they appear in the 'window'.
3. Choose two different jig words to look at together.

Session 5

Objectives:

1. To hear each child read through Book 2
2. To record any errors

Resources:

1. Book 2
2. A check sheet of the words for each child
3. Word list for book 2

Activity:

1. Books are to be read by individual children to the TA.
2. If a child hesitates the TA should say the word aloud (without sounding out) then read the whole sentence.
3. The child can then continue on to the next sentence/page.
4. TA to make a record of unknown words on individual tick sheet.

5. At the end of the book TA to go back and find the words recorded and read them again to the child, pointing out any clues such as the initial letter, particular features or endings.
6. Recheck highlighted nouns on word list.

Summary of Week 9

Objectives:

- Learn 'weebee' song 'Pip's Song'
- Introduction to games for book 9
- Introduction to phonics worksheets
- Read book 9 (Session 5 only – please note change of instruction)

Resources:

- 'weebee' CD and 'big book'
- Snakes and ladders
- Pento games
- Grog's Journey
- Sentence matching
- Phonics worksheets, scissors and felt pens
- Book 9
- Tick sheets for Book 9 (one per child) and word list

Whole Group Activity:

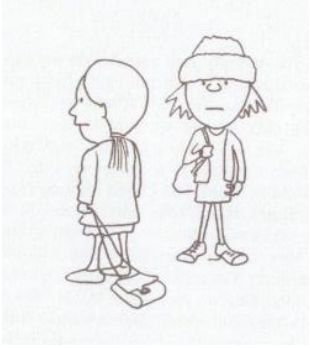
- 'weebee' song
- Look at 'big book'

Small Group Activities:

- Snakes and ladders
- Pento (two games)
- Grog's Journey
- Sentence matching
- Phonics worksheets

Plenary:

- Looking at the week's phonics worksheet and 'big book' word families



Week 9

Session 1

Objectives:

1. To learn the 'weebee' song 'Pip's Song'
2. To learn Pento, and Grog's Journey.
3. Revision work (sentence matching).
4. Phonics work

Resources:

1. 'weebee' CD and 'big book'
2. Pento game (2 sets)
3. Grog's Journey
4. Sentence Matching
5. Draw-through phonics sheets (_ee_), scissors and felt pens

Activity 1 (all the children)

1. Listen to the 'weebee' song on the CD and look at the picture on page 53 in the 'big book'.
2. Ask the children if they have had a holiday in the sun. Can they tell you what they need to have in hot weather? Do they like hot weather? Do they know how far away the sun is and how big it is?
3. Organise the children into groups of 3-4 for activities 2, 3 and 4.

Activity 2: Pento (4 children for each set)

1. Each player chooses a counter.
2. The game works like Monopoly, using dice to move around the board in a clockwise direction starting at *Go*.
3. The objective is to collect the set of 5 cards which match the chosen counter. These can be collected as each player lands on their own words.
4. If a player lands on a colour they should take the top one and follow the instructions replacing the card at the bottom of the pile facing down.
5. The winner is the first player to collect all 5 cards.

Activity 3: Grog's Journey (4 children)

1. Each player chooses a counter.
2. The objective is simple, the first to the top of the tree using dice to move along the path.
3. As the numbers are rolled, the TA should point to the word reached and verbalise each time a player lands on a word.

Activity 4: Sentence matching (8 children in pairs)

1. Working in pairs, children simply match sentences to pictures.
2. Some words may be interchangeable, this does not matter.
3. TA to encourage children to read the sentences after they have been correctly arranged.

Activity 5: Phonics worksheet (_ee_)

This requires two strips to be cut after the letters have been drawn over.

Concluding Activity (all the children)

1. Choose one of the draw-through cards that the children have made.
2. Whole group say the words together as they appear in the 'window'
3. Look at the word family on page 54 of the 'big book'. Make the highlighted sound together and then read the list out loud together.
4. Ask the children to think of a sentence to use each word and help as necessary.

Session 2

Objectives:

1. To practise the 'weebee' song 'Pip's Song'
2. To use Pento, and Grog's Journey
3. Revision work (sentence matching)
4. Phonics work

Resources:

1. 'weebee' CD and 'big book'
2. Pento game (2 sets)
3. Grog's Journey
4. Sentence Matching
5. Draw-through phonics sheets (_ee_), scissors and felt pens

Activity 1 (all the children)

1. Listen to the 'weebee' song on the CD and look at the picture on page 55 in the 'big book'.
2. Ask the children if they remember what is growing by the pond. Do they remember seeing bulrushes in one of the other books? Can they remember what happened?
3. Organise the children into groups of 3-4 for activities 2, 3 and 4.

Activity 2: Pento (4 children for each set)

1. Each player chooses a counter.
2. The game works like Monopoly, using dice to move around the board in a clockwise direction starting at *Go*.

3. The objective is to collect the set of 5 cards which match the chosen counter. These can be collected as each player lands on their own words.
4. If a player lands on a colour they should take the top one and follow the instructions replacing the card at the bottom of the pile facing down.
5. The winner is the first player to collect all 5 cards.

Activity 3: Grog's Journey (4 children)

1. Each player chooses a counter.
2. The objective is simple, the first to the top of the tree using dice to move along the path.
3. As the numbers are rolled, the TA should point to the word reached and verbalise each time a player lands on a word.

Activity 4: Sentence matching (8 children in pairs)

1. Working in pairs, children simply match sentences to pictures.
2. Some words may be interchangeable, this does not matter.
3. TA to encourage children to read the sentences after they have been correctly arranged.

Activity 5: Phonics worksheet (_ee_)

This requires two strips to be cut after the letters have been drawn over.

Concluding Activity (all the children)

1. Choose one of the draw-through cards that the children have made.
2. Whole group say the words together as they appear in the 'window'
3. Look at the word family on page 56 of the 'big book'. Make the highlighted sound together and then read the list out loud together.
4. Ask the children to think of a sentence to use each word and help as necessary.

Session 3

Objectives:

1. To practise the 'weebee' song 'Pip's Song'
2. To learn snakes and ladders and use Pento
3. Revision work (sentence matching)
4. Phonics work

Resources:

1. 'weebee' CD and 'big book'
2. Pento game (2 sets)
3. Snakes and ladders
4. Sentence Matching
5. Draw-through phonics sheets (_ee_), scissors and felt pens

Activity 1 (all the children)

1. Listen to the '*weebee*' song on the CD and look at the picture on page 57 in the 'big book'.
2. Ask the children what they can see in the grass. What kind of egg do they think it might be? What kinds of eggs have they seen? What kinds of eggs have they eaten? Can they guess what is inside?
3. Organise the children into groups of 3-4 for activities 2, 3 and 4.

Activity 2: Pento (4 children for each set)

1. Each player chooses a counter.
2. The game works like Monopoly, using dice to move around the board in a clockwise direction starting at *Go*.
3. The objective is to collect the set of 5 cards which match the chosen counter. These can be collected as each player lands on their own words.
4. If a player lands on a colour they should take the top one and follow the instructions replacing the card at the bottom of the pile facing down.
5. The winner is the first player to collect all 5 cards.

Activity 3: Snakes and ladders (4 children)

1. Each player chooses a counter.
2. The counters begin at the start. Using dice each player moves along the path.
3. The objective is simple, the first to or past the finish is the winner.
4. If a player lands at the foot of a ladder they move up to the top of the ladder. If a player lands at the head of a snake they move down the snake to the end of it's tail.
5. As the numbers are rolled, the TA should point to the word reached and verbalise each time a player lands on a word.

Activity 4: Sentence matching (8 children in pairs)

1. Working in pairs, children simply match sentences to pictures.
2. Some words may be interchangeable, this does not matter.
3. TA to encourage children to read sentences after correctly arranged.

Activity 5: Phonics worksheet (_ee_)

This requires two strips to be cut after the letters have been drawn over.

Concluding Activity (all the children)

1. Choose one of the draw-through cards that the children have made.
2. Whole group say the words together as they appear in the 'window'.
3. Look at the word family on page 58 of the 'big book'. Make the highlighted sound together and then read the list out loud together.
4. Ask the children to think of a sentence to use each word and help as necessary.

Session 4

Objectives:

1. To practise the 'weebee' song 'Pip's Song'
2. To use snakes and ladders and Pento
3. Revision work (sentence matching)
4. Phonics work

Resources:

1. 'weebee' CD and 'big book'
2. Pento game (2 sets)
3. Snakes and ladders
4. Sentence Matching
5. Draw-through phonics sheets (_ee_), scissors and felt pens

Activity 1 (all the children)

1. Listen to the 'weebee' song on the CD and look at the picture on page 59 in the 'big book'.
2. Ask the children if they can see who is swimming. Ask who can swim. Who has swimming lessons? Do they like swimming? Where have they been swimming?
3. Organise the children into groups of 3-4 for activities 2, 3 and 4.

Activity 2: Pento (4 children for each set)

1. Each player chooses a counter.
2. The game works like Monopoly, using dice to move around the board in a clockwise direction starting at *Go*.
3. The objective is to collect the set of 5 cards which match the chosen counter. These can be collected as each player lands on their own words.
4. If a player lands on a colour they should take the top one and follow the instructions replacing the card at the bottom of the pile facing down.
5. The winner is the first player to collect all 5 cards.

Activity 3: Snakes and ladders (4 children)

1. Each player chooses a counter.
2. The counters begin at the start. Using dice each player moves along the path.
3. The objective is simple, the first to or past the finish is the winner.
4. If a player lands at the foot of a ladder they move up to the top of the ladder. If a player lands at the head of a snake they move down the snake to the end of it's tail.
5. As the numbers are rolled, the TA should point to the word reached and verbalise each time a player lands on a word.

Activity 4: Sentence matching (8 children in pairs)

1. Working in pairs, children simply match sentences to pictures.

2. Some words may be interchangeable, this does not matter.
3. TA to encourage children to read sentences after correctly arranged.

Activity 5: Phonics worksheet (_ee_)

This requires two strips to be cut after the letters have been drawn over.

Concluding Activity (all the children)

1. Choose one of the draw-through cards that the children have made.
2. Whole group say the words together as they appear in the 'window'.
3. Look at the word family on page 60 of the 'big book'. Make the highlighted sound together and then read the list out loud together.
4. Ask the children to think of a sentence to use each word and help as necessary.

Session 5

Objectives:

1. To hear each child read through Book 9
2. To record any errors
3. To encourage fluency and self-correction

Resources:

1. Book 9
2. A check sheet of the words for each child
3. Word list for book 9

Activity:

1. Books are to be read by individual children to the TA.
2. The method of instruction is to differ for this next set of books. Ask the child to read all the words of each sentence in their heads first then tell you 'what it says'.
3. The child can then continue on to the next sentence/page.
4. TA to make a record of unknown words on individual tick sheet.
5. At the end of the book TA to go back and find the words recorded and read them again to the child, pointing out any clues such as the initial letter, particular features or endings.
6. Recheck highlighted nouns from word list.

Appendix H Consent Forms and Information Sheets for parents and schools

THE UNIVERSITY of York

Institute for **Effective Education**
Empowering educators with evidence

School and Teacher Consent Form (Version 1, 09.05.2013)
Pilot of Resources

Name of school..

Name of Head teacher...

I confirm that I have read the information sheet (Version 1, 30.04.2013), had a chance to ask questions and I am willing to participate in the pilot of resources for the research project. I understand that I can withdraw from the study at any time.

Signature of Head teacher..

Signature of Class Teacher...

Thank you

Please feel free to contact me, Ruth, on 07970 151275 or rmm519@york.ac.uk

In the case of any complaint about the conduct of this study, please contact: Dr. Louise Tracey on 01904 328160 or write to Professor Chris Kyriacou, Department of Education, University of York, Heslington, York YO10 5DD

Institute for Effective Education Berrick Saul Building University of York YO10 5DD
Tel 01904 328166 Fax 01904 328156 Email iee@york.ac.uk www.york.ac.uk/iee

THE UNIVERSITY of York

Institute for **Effective Education**
Empowering educators with evidence

School Information Sheet (Version 1, 09.05.2013)
Pilot of Resources

Researcher: Ruth Mace
University: York

Dear

This pilot study is a student PhD project. Ruth Mace is the lead researcher working on this project, supervised by Dr. Louise Tracey at the Institute for Effective Education, and Dr. Claudine Bowyer-Crane, Department of Education, at the University of York.

Children in the reception class will be using the 'wee bee' reading programme activities for this pilot study. The Teaching Assistant in reception class will be trained to deliver small group sessions of the programme which is designed to support reading development. The pilot programme will begin in June of this year (2013).
The sessions will be part of the school day, 30 minutes per day for 4 weeks.

Ruth Mace, PhD student
rmm519@york.ac.uk
07970 151275

Institute for Effective Education Berrick Saul Building University of York YO10 5DD
Tel 01904 328166 Fax 01904 328156 Email iee@york.ac.uk www.york.ac.uk/iee

Pilot Evaluation of 'wee bee' reading programme activities
Parent 'opt out' Form (Version 1, 30.04.2013)

I confirm that I have read and understood the information sheet version 1 dated 30.04.2013.

I understand that if I am happy for my child to participate in the 'weebee' reading programme activities, then I do not need to do anything.

I also understand that I can object to my child participating by signing this form and returning it to the school.

If I sign this form:
I do not wish for my child to participate in the 'wee bee' reading programme activities.
I understand that my child will not join in with the 'wee bee' activities or use the related reading materials.

Signature of parent.....Date.....

Your name in print please.....

Your child's name.....

Your child's school.....

Institute for Effective Education Berrick Saul Building University of York YO10 5DD
Tel: 01904 328166 Fax: 01904 328156 Email: iee@york.ac.uk www.york.ac.uk/iee

Pilot Evaluation of 'wee bee' reading programme activities
Parent Information Sheet (Version 1, 30.04.2013)

Dear Parent
This study is a student PhD project. Ruth Mace is the lead researcher working on this project, supervised by Dr. Louise Tracey at the Institute for Effective Education, and Dr. Claudine Bowyer-Crane, Department of Education, at the University of York.

Invitation
Your child attends a school that has agreed to be part of this study. Children in the reception class will be using the 'wee bee' reading programme activities for this study. The Teaching Assistant in your child's class is being trained to deliver small group sessions of the programme which is designed to support your child's reading development. The pilot programme will begin in June of this year (2013).
The sessions will be part of the school day, 30 minutes per day for 4 weeks.

Do I have to take part?
It is entirely up to you whether or not you decide to allow your child to be part of this pilot research project and participate in the activities. If you are happy for your child to be part of the study, you are free to withdraw from the research at any time and you do not need to give a reason. Your withdrawal from the study will not affect the support that your child will receive from the school. If you do decide to opt out, would you please sign the attached 'opt out' form to return to the school.

What are the benefits of taking part?
If you agree for your child to take part, your child will be given extra reading support using fun-filled group activities.

What are the possible disadvantages or risks of taking part?
There are no obvious disadvantages or risks to your child.

Will my child's details be kept confidential?
Yes. All the information about your child and their school will remain strictly confidential. Results will only be reported for the group of children taking part as a whole and these will be kept anonymous.

Thank you

In the case of any complaint about the conduct of this study, please contact: Dr. Louise Tracey on 01904 328160 or write to Professor Chris Kyriacou, Department of Education, University of York, Heslington, York YO10 5DD

Institute for Effective Education Berrick Saul Building University of York YO10 5DD
Tel: 01904 328166 Fax: 01904 328156 Email: iee@york.ac.uk www.york.ac.uk/iee

THE UNIVERSITY of York

Institute for **Effective Education**
Empowering educators with evidence

Proposed teaching methods:

Cooperative and group learning
Structure: Four days using resources, one day reading the book.
TAs to work with groups of 4 for 15 mins x 2 per day
(a total of 30 mins per day for 4 days, plus 1 day hearing children read)
Other children to work in pairs unaided

All resources will be provided by the researcher. These will be demonstrated during the TA training session and will include a manual for use by TAs.

What are the benefits to you?

Free training for the Teaching Assistant (Half a day)
Free resources and books
Short reports on assessment scores for participating children
Participating in research that could significantly add to the field of reading/literacy

Rest assured that your identity, the school's identity, and the children's identity as well as the data collected will remain strictly confidential and anonymous. The school, teachers and parents will of course have the right to withdraw from the research at any stage.

We feel this is an important piece of research which will inform the teaching community as well as policy makers. If you have any questions, please do not hesitate to contact me.

Ruth Mace, PhD student
rmm519@york.ac.uk
07970 151275
www.york.ac.uk/iee/studying/studentprofiles/rmace.htm

THE UNIVERSITY of York

Institute for **Effective Education**
Empowering educators with evidence

School Information Sheet(Version 1, 30.04.2013)
'wee bee' Reading Programme

Researcher: Ruth Mace
University: York

Dear [redacted]

I am writing to invite your school to participate in a PhD research project.The researcher, Ruth Mace, is being supervised by Dr. Louise Tracey at the Institute for Effective Education,and Dr. Claudine Bowyer-Crane, at the Department of Education, at the University of York. Ethical approval is being sought through the Department of Education ethics committee and will be confirmed shortly.

In 2005, a House of Commons Education and Skills Committee report 'Teaching Children to Read' recommended that the Department of Education and Skills commission a large scale comparative study to establish:

> 'the effect of teaching texts which go beyond a child's existing knowledge of phonics compared to that of limiting instructional texts to those within a child's current decoding abilities' (Eighth Report of Session , 2005, paragraph 52)

To date, there have been a number of reviews which have made comparisons between schools using different reading programmes both in the UK and US. However, no direct comparisons have as yet been made.

The aim of the study is to create two parallel reading programmes. Both will have the same illustrations and the same number of new words to learn for each book. The teaching methods will be the same for both programmes. Rather than relying solely on synthetic phonics, a range of approaches will be used. Both programmes will have graded and sequenced books. The only variable will be the vocabulary. This will be for children starting in reception in September 2013.

Short literacy assessments (approx.20-30 minutes/child) with individual children will be carried out by the lead researcher (with DBS) in schools before and after the proposed intervention (a quiet area will be needed).

The proposed timeline is:

Summer Term 2013: TA training in participating schools (and early September)
Autumn Term 2013: Assessments (1)
Begin intervention
Four weeks per half term
Summer Term 2014: Finish intervention
Assessments (2)

Institute for Effective Education Berrick Saul Building University of York YO10 5DD
Tel: 01904 328166 Fax: 01904 328156 Email: iee@york.ac.uk www.york.ac.uk/iee

THE UNIVERSITY of York

Institute for **Effective Education**
Empowering educators with evidence

School Information Sheet(Version 1, 30.04.2013)
Literacy assessments

Researcher: Ruth Mace
University: York

Dear

I am writing to invite your school to participate in a PhD research project.The researcher, Ruth Mace, is being supervised by Dr. Louise Tracey at the Institute for Effective Education,and Dr. Claudine Bowyer-Crane, at the Department of Education, at the University of York. Ethical approval is being sought through the Department of Education ethics committee and will be confirmed shortly.
The aim of the study is to look at the value of a new reading programme. In order to establish if the programme is effective we need some schools to act as comparison schools. This will be for children starting in reception in September 2013.
Literacy assessments with individual children will be carried out by the lead researcher (with DBS) in schools before and after the proposed intervention (a quiet area will be needed).

The proposed timeline is:

Autumn Term 2013: **Assessments (1)**
Begin intervention
Summer Term 2014: Finish intervention
Assessments (2)

Proposed Assessment tools:
British Picture Vocabulary Scales (BPVS)
York Assessment of Reading Comprehension (YARC)

What are the benefits to you?
Short reports on assessment scores for participating children

Rest assured that your identity, the school's identity, and the children's identity as well as the data collected will remain strictly confidential and anonymous. The school, teachers and parents will of course have the right to withdraw from the research at any stage.

We feel this is an important piece of research which will inform the teaching community as well as policy makers. If you have any questions, please do not hesitate to contact me.

Ruth Mace, PhD student
rmm519@york.ac.uk
07970 151275

Institute for Effective Education Berrick Saul Building University of York YO10 5DD
Tel: 01904 328166 Fax: 01904 328156 Email: iee@york.ac.uk www.york.ac.uk/iee

THE UNIVERSITY of York

Institute for **Effective Education**
Empowering educators with evidence

School and Teacher Consent Form(Version 1, 30.04.2013)

Name of school.....

Name of Head teacher.....

Total number of pupils eligible for free school meals in reception class.....

Any other information about the school which you think may be of interest to us.....

.....

.....

I confirm that I have read the information sheet (Version 1, 30.04.2013), had a chance to ask questions and I am willing to participate in the 'wee bee' reading programme research project. I understand that this will involve half a day of training for a TA as well as a small amount of paper work (weekly check list). I understand that there will be 1-2 days of assessments at the beginning of the intervention and post-intervention. I understand that the intervention will require approximately 30 minutes per day, four days a week, a fifth day for hearing the children read and that this will run four weeks per half term. I understand that I can withdraw from the study at any time.

Signature of Head teacher.....

Signature of Class Teacher.....

Signature of Teaching Assistant.....

Thank you

Please feel free to contact me, Ruth, on 07970 151275 or rmm519@york.ac.uk

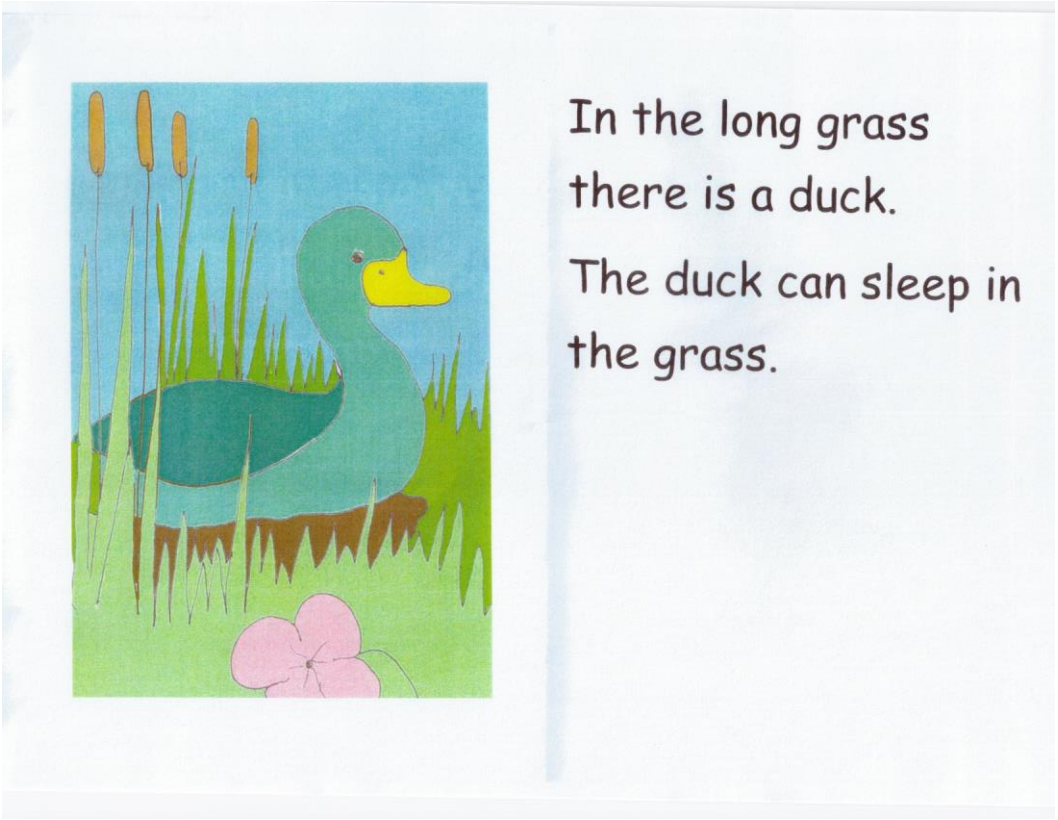
In the case of any complaint about the conduct of this study, please contact: Dr. Louise Tracey on 01904 328160 or write to Professor Chris Kyriacou, Department of Education, University of York, Heslington, York YO10 5DD

Institute for Effective Education Berrick Saul Building University of York YO10 5DD
Tel: 01904 328166 Fax: 01904 328156 Email: iee@york.ac.uk www.york.ac.uk/iee

Appendix I List of 32 Nouns for assessment:

ant	nest	blue	button
bed	rock	chick	caterpillar
bee	root	grass	dragon
duck	sack	petal	flower
egg	shell	pink	monster
eye	ship	stick	orange
hat	spot	wall	spider
owl	tree	wood	yellow

Appendix J Example page from the book for assessment of Reading Speed



In the long grass
there is a duck.

The duck can sleep in
the grass.

Appendix K Observation Schedule

School Name

Researcher

Date of analysis

Time/Length of observation

Target outcome	Objective	Observed = 1 Not observed = 0	Level: Mechanical = 1 Routine = 2 Refined = 3
	Allows children to make additional suggestions Clear evidence of extending children's vocabulary		
Desired behaviour	Teacher pronounces words clearly Teacher models correct pronunciation Teacher asks children to look to see if they have the same word. Teacher points to initial letter for identification Teacher points to other salient features for identification Teacher allows other children to assist Teacher encourages other children to assist Children help each other without prompting Children discuss the salient features without prompting		
Undesired behaviour	Teacher encourages children to sound out Teacher stops children helping each other Teacher points to all the letters in turn Teacher expects children to read every word, rather than focusing on the game Teacher shuts down conversation rather than using the opportunity to extend the children's vocabulary		

Appendix L Example questionnaire for teacher re: experience (School code 92)

Amount of time having worked with children	*10 years*
Describe teaching experience	*All primary and early years. From Reception up to years 2 and 3.*
Describe any CPD	*A lot of phonics CPD and Book Talk*
How much overlap is there in terms of phonics and sight vocabulary	*A lot of overlap. We tend to do phonics each day and also teach the children a sight vocabulary. We believe that children learn indifferent ways and so we use a mixed approach and don't expect children to rely on synthetic phonics.*
Approximate total time spent on the weebees	*We used to do 3 sessions per week of 25 minutes each. Now we include some of the activities during the phonics sessions, but it works out about the same amount of time.*

Appendix M

Group means for the control condition after individuals with EAL have been removed (Study 3).

BPVS pre-test	98.43	SI pre-test	112.39
BPVS post-test	93.95	SI post-test	109.54
LSK pre-test	116	SD pre-test	101.13
LSK post-test	117	SD post-test	100.6
EWR pre-test	108.6	Passage Reading	101.6
EWR post-test	108.82	Reading Speed	53.47

Appendix N

Feedback for individuals Study 3

Class Teacher: 1 Child Code	**Feedback for individuals:**
9513(p)	He has made a lot of progress which she [the teacher] attributes to his involvement in the weebee project. (He suffers from absence epilepsy)
9524(p)	He has not made any noticeable progress but has enjoyed the project. (SEN)
9504(p)	His comprehension has really improved, which she attributes to the weebee project. (SEN)
9516(a)	She has made huge improvements especially in comprehension.
9522(a)	He has made some progress, particularly in confidence.
9515(a)	Has made obvious progress.
	Feedback for whole group:
	The weebee project was worthwhile for the whole group. She would use it again but not for so much time and plan her own programme with the resources. The children were motivated by the characters and looked forward to the sessions as they clearly enjoyed them. Overall it was definitely worthwhile being involved in the project.
Class Teacher: 2 Child Code	**Feedback for individuals:**
9422(p)	She did not get much out of it as she did not seem to be able to match any of the words and couldn't access the activities.
9414(p)	Real progress in his reading, especially in his confidence. He does not read much at home and so the weebee input was very useful.
9413(p)	He has shown a massive improvement, but she [the teacher] found it hard to separate the weebee input from other interventions also in place for him.
9416(a)	Huge improvement in confidence.
9407(a)	Her confidence in reading has grown.
9427(a)	He is making much more sense of what he is reading.
	Feedback for whole group:
	The project took up a lot of time on just a small group of individuals. The TA would normally be spread more evenly across the class. However it did benefit the children and the children evidently enjoyed the project and looked forward to being taken out in their respective groups. The children found the weebees engaging. She would use the programme again but plan the use of the resources differently according to both the needs of the children and the requirements of the day-to-day classroom management.

Appendix O

Sub-analysis of the impact of fidelity in Study 2 (using standardised scores)

School comparisons	Test	Mean Time 2 (SD)	Sig. (*p*) Time 2 only	Effect size (Cohen's *d*) Time 2 only
92 compared to	BPVS	103.66(10.12)	.634	0.17
91 + 93		101.95(9.56)		(3 months difference)
	LSK	120.91(3.23)	.464	0.27
		119.95(3.83)		(4 months difference)
	EWR	108.08(13.31)	.992	0.00
		108.13(10.54)		
	PRC	110.50(12.71)	.244	0.38
		106.17(9.64)		(5 months difference)
92 compared to Control	BPVS	103.66(10.12)	.004	1.10
		91.55(11.75)		(more than 12 months)
	LSK	120.92(3.23)	.033	0.62
		114.55(13.98)		(8 months difference)
	EWR	108.08(13.31)	.819	0.08
		106.88(15.55)		(less than 2 months)
	PRC	110.50(12.71)	.036	0.78
		99.40(15.41)		(10 months difference)
91 + 93 compared to	BPVS	101.95(9.56)	.001	0.97
Control		91.55(11.75)		(12 months difference)
	LSK	119.95(3.83)	.079	0.52
		114.55(13.98)		(6 months difference)
	EWR	108.13(10.54)	.747	0.09
		106.88(15.55)		(less than 2 months)
	PRC	106.17(9.64)	.066	0.52
		99.40(15.41)		(6 months difference)

The table shown above compares School 92 (highest fidelity in Study 2) with Schools 91 + 93 (equally low fidelity in Study 2) and the Control school from that study. This sub-analysis demonstrates the impact of higher fidelity levels on the four measures common to all three studies. There has been no effect on the Early Word Reading (EWR) measure. However, there appears to have been some impact on Letter Sound Knowledge (LSK), although this measure (and EWR) was subject to ceiling effects. The greatest difference between School 92 and Schools 91 + 93 is in Passage Reading Comprehension (PRC) and this difference equates to 5 months reading progress. This measure is not subject to ceiling effects. These results are evidence that higher fidelity to an intervention can lead to higher outcome measures. For the PRC measure (not subject to ceiling effects), the differences between the intervention arms was greater in School 92 (P: M = 102.33 and A: M =

118.66) compared to Schools 91 + 93 (P: *M* = 103.41 and A: *M* = 109.18). The effect size for Intervention A between School 92 and 91+93 (*d* = 1.01, *g* = 0.94) equated to more than 12 months and between School 92 and the control the difference was even greater (*d* = 1.53, *g* = 1.29: 19 months progress).

Abbreviations

BPVS	British Picture Vocabulary Scale
EAL	English as an Additional Language
EWR	Early Word Reading
GPC	Grapheme-Phoneme Correspondence
ITT	Intention to Treat Analysis
LSK	Letter Sound Knowledge
PRC	Passage Reading Comprehension
RCT	Randomised Controlled Trial
RS	Reading Speed
SD	Sound Deletion
SI	Sound Isolation
SSP	Systematic Synthetic Phonics
YARC	York Assessment of Reading for Comprehension

Glossary

Analogy	Similarity of letter groups between words
Analytic phonics	Identifying (analysing) common phonemes in words
Basal readers	Textbooks used to teach reading
Bayes nets	Graphical representation of a probability distribution
Blending	Hearing phonemes and merging them together
Dialogic reading	Asking and expanding questions whilst reading together
Etymology	The origins of a word
Grapheme-phoneme	Letter-to-sound relationships
Guided reading	The teacher drawing out the meaning whilst reading
Inference Generation	When something is inferred by the reader
Intention-to-Treat Analysis	Analysis based on the treatment assigned
Joint Attention	Shared focus of two individuals on an object
Morphology	Analysis of the structure of words, e.g. root words
Onset	The initial consonant of a word that precedes the rime
Orthography	The spelling system of a language
Phoneme awareness	The ability to hear, identify and manipulate phonemes
Phoneme	Perceptually distinct units of sound in a language
Phonics	Correlating sounds with symbols in alphabetic writing
Phonology	Contrasting relationships of speech sounds
Real books	Written by children's authors with no intention to teach
Rimes	The rest of a syllable that follows the onset
Salient Features	A feature that is prominent or 'stands out' in a word
Segmenting	To split up a word into its individual phonemes
Self-teaching hypothesis	Phonological recoding acts as a self-teaching mechanism
Sight word	Words which are recognised immediately on sight
Synthetic phonics	Teaching letter sounds which are then blended into words
Systematic phonics	Teaching phonics systematically in a pre-defined order
Whole word	Using whole words first before individual letters are taught

References:

Abrami, P.C., Lou, Y., Chambers, B., Poulsen, C., & Spence, J.C. (2000). Why should we group students within-class for learning? *Educational Research and Evaluation: An International Journal on Theory and Practice*, 6(2), 158-179.

Akhtar, N., & Gernsbacher, M.A. (2007). Joint Attention and Vocabulary Development: A Critical Look. *Language and Linguistic Compass*, 1(3), 195-207.

Allen, L. (1998). An Integrated Strategies Approach: Making Word Identification Instruction Work for Beginning Readers. *The Reading Teacher*, 52(3), 254-268.

Allington, R.L., & Woodside-Jiron, H. (1999). The Politics of Literacy Teaching: How "Research" Shaped Educational Policy. *Educational Researcher*, 28(8), 4-13.

Arizpe, E., & Styles, M. (2003). *Children reading pictures: Interpreting visual texts*. London: RoutledgeFalmer.

Ashby, J., & Rayner, K. (2012). Reading in alphabetic writing systems: evidence from cognitive neuroscience. In S. Della Sala & M. Anderson (Eds.), *Neuroscience in Education: The good, the bad and the ugly*. Oxford: Oxford University Press.

Atherton, J. S. (2013). *Learning and Teaching: Piaget's developmental theory* Retrieved February 25, 2015 from http://www.learningandteaching.info/learning/piaget.htm

Barab, S. (2014) Design-Based Research. In R.K. Sawyer (Ed.) *The Learning Sciences* (pp 151-170). New York: Cambridge University Press

Baumann, J.F., Ware, D., & Carr Edwards, E. (2007). "Bumping into Spicy, Tasty Words That Catch Your Tongue": A Formative Experiment on Vocabulary Instruction. *The Reading Teacher*, 61(2), 108-122.

Baylis, P., & Snowling, M.J. (2011). Evaluation of a phonological reading programme for children with Down syndrome. *Child Language Teaching and Therapy*, 28(1), 39-56.

Beech, J.R. (2005). Ehri's model of phases of learning to read: a brief critique. *Journal of Research in Reading*, 28(1), 50-58.

Beech, J.R. (2010). Young Readers' Strategic Approaches to Reading Unfamiliar Words in Text. *Reading & Writing Quarterly: Overcoming Learning Difficulties,* 26(3), 264-284.

Beech, J.R., & Mayall, K.A. (2005). The word-shape hypothesis re-examined: evidence for an external feature advantage in visual word recognition. *Journal of Research in Reading*, 28(3), 302-319.

Beverly, B.L., Giles, M., & Buck, K.L. (2009). First-grade reading gains following enrichment: phonics plus decodable texts compared to authentic literature read aloud. *The Reading Teacher*, 53(4), 292-307.

Blais, C., Fiset, D., Arguin, M., Jolicoeur, P., Bub, D., & Gosselin, F. (2009). Reading between eye saccades. *PLOS ONE*, 7(7), 1-7.

Booil, J., Asparouhov, T., & Muthén, B.O. (2008). Intention-to-treat analysis in cluster randomized trials with noncompliance. *Statistics in Medicine*, 27, 5565-5577.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester: Wiley.

Bowyer-Crane, C. and Snowling,M.J. (2005). Assessing children's inference generation: What do tests of reading comprehension measure? *British Journal of Education Psychology*, 75, 189-201.

Bowyer-Crane, C., Snowling, M.J., Duff, F.J., Fieldsend, E., Carroll, J.M., Miles, J., Götz, K., & Hulme, C. (2008). Improving early language and literacy skills: differential effects of an oral language versus a phonological with reading intervention. *Journal of Child Psychology and Psychiatry,* 49(4), 422-432.

Brown, K.J. (2000). What kind of text – For whom and when? Textual scaffolding for beginning readers. *The Reading Teacher*, 53(4), 292-307.

Bucci, M.P., Nassibi, N., Gerard, C-L., Bui-Quoc, E., & Seassau, M. (2012). Immaturity of the Oculomotor Saccade and Vergence Interaction in Dyslexic Children: Evidence from a Reading and Visual Search Study. *PLoS ONE*, 7(3), 1-8.

Burnett, J. (2007). A pleasurable path to literacy: can Steiner contribute to the literacy debate? *Journal of Early Childhood Literacy*, 7(3), 321-331.

Bus, A.G., van Ijzendoorn, M.H., & Pellegrini, A.D. (1995). Joint book reading makes for success in learning to read. A meta-analysis on intergenerational analysis of literacy. *Review of Educational Research*, 65(1), 1-21.

Bus, A.G., & van Ijzendoorn, M.H. (1999). Phonological Awareness and Early Reading: A Meta-Analysis of Experimental Training Studies*. Journal of Educational Psychology*, 91(3), 403-414.

Campbell, R. (2007). Literacy learning – valuing books, children and teachers. *Education 3-13: International Journal of Primary, Elementary and Early Years Education*, 31(1), 53-57.

Camilli, G., Vargas, S., & Yurecko, M. (2003). Teaching Children to Read: The Fragile Link Between Science and Federal Education Policy. *Educational Policy Analysis Archives*, 11(15), 1-51.

Camilli, G., Kim, S.H., & Vargas, S. (2008). A Response to Steubing et al., "Effects of Systematic Phonics Instruction are Practically Significant": The Origin of the National Reading Panel. *Education Policy Analysis Archives*, 16(16), 1-19.

Carroll, L. (1871). *Through the Looking-Glass, and What Alice Found There*. London: Macmillan.

Chall, J. (1967). *Learning to Read: The Great Debate*. New York: McGraw-Hill.

Chen, V., & Savage, R.S. (2014). Evidence for a simplicity principle: teaching common complex grapheme-to-phonemes improves reading and motivation in at-risk readers. *Journal of Research in Reading*, 37(2), 196-214.

Chu, M.C., & Chen, S.H. (2014). Comparison of the effects of two phonics training programs on L2 word reading. *Psychological Reports: Relationships & Communications*, 114(1), 272-291.

Clark, M.M. (1976). *Young Fluent Readers: What can they teach us*? London: Heineman Educational.

Clark, M.M. (2014a). *Synthetic Phonics and Literacy Learning*. Birmingham: Glendale Education.

Clark, M.M. (2014b). *Learning to be Literate*. Birmingham: Glendale Education.

Clarke, L.J. (2009). Inference Generation and Reading Disability, Unpublished Doctoral Thesis, University of York, 2009.

Clarke, P., Snowling, M.J., Truelove, E., & Hulme, C. (2010). Ameliorating Children's Reading-Comprehension Difficulties: A Randomised Controlled Trial. *Psychological Science*, 20(10), 1-11.

Coles, G. (2004). Real Books in the Caboose. *Knowledge Quest*, 33(2), 22-25.

Collier-Meek, M.A., Fallon, L.M., Sanetti, L.M.H., & Maggin, D., (2013). Assessing and Promoting Treatment Fidelity. *Teaching Exceptional Children*, 45(5), 55-59.

Compton, D.L., Miller, A.C., Ellenan, A.M., & Steacy, L.M. (2014). Have We Forsaken Reading Theory in the Name of "Quick Fix" Interventions for Children with Reading Disability? *Scientific Studies of Reading*, 18(1), 55-73.

Comaskey, E.M., Savage, R.S., & Abrami, P. (2009). A randomised efficacy study of web-based synthetic and analytic programmes among disadvantaged urban kindergarten children. *Journal of Research in Reading*, 32(1), 92-108.

Connor, C.M., Morrison, F.J., & Katch, L.E. (2004). Beyond the Reading Wars: Exploring the Effect of Child-Instruction Interactions on Growth in Early Reading. *Scientific Studies of Reading*, 8(4), 305-336.

Cook, M., Littlefair, A., & Brooks, G. (2007). Responses to Wyse and Styles' article, "Synthetic phonics and the teaching of reading: the debate surrounding England's 'Rose Report'" (Literacy, 41(1), April, 2007). *Literacy*, 41(3), 169.

Corriveau, K., Pasquini, E., & Goswami, U. (2007). Basic Auditory Processing Skills and Specific Language Impairment: A New Look at an Old Hypothesis. *Journal of Speech, Language and Hearing Research*, 50, 647-666.

Corriveau, K., Goswami, U., & Thomson, J.M. (2010). Auditory Processing and Early Literacy Skills in Preschool and Kindergarten Population. *Journal of Learning Disabilities,* 43(4), 369-382.

Cunningham, A.E., Perry, K.E., Stanovich, K.E., & Share, D.L. (2002). Orthographic learning during reading: examining the role of self-teaching. *Journal of Experimental Child Psychology*, 82, 185-99.

Cunningham, J.E. (2011). Intention-to-treat analysis: the parachute revisited. *ANZ Journal of Surgery*, 81(12), 920-922.

Darrow, C.L. (2013). The Effectiveness and Precision of Intervention Fidelity Measures in Preschool Intervention Research. *Early Education and Development*, 24(8), 1137-1160.

Davis, A. (2012). A Monstrous Regimen of Synthetic Phonics: Fantasies of Research-Based Teaching 'Methods' Verses Real Teaching. *Journal of Philosophy of Education,* 46(4), 561-573.

Davis, C., & Bryant, P. (2006). Causal connections in the acquisition of an orthographic rule: a test of Uta Frith's developmental hypothesis. *Journal of Child Psychology and Psychiatry*, 47(8), 849-856.

Department of Education and Science [DES] (1975). *A Language for life*. London: HMSO.

Department for Education and Employment [DfEE] (1998). *The National Literacy Strategy*. London: HMSO.

Department for Education and Skills [DfES] (2006). *Primary Framework for literacy and mathematics*. London: HMSO.

Department for Education and Skills [DfES] (2007). *Letters and Sounds: Principles and practice of high quality phonics six-phase teaching programme*. London: HMSO.

Department for Education and Skills [DfES] (2009). *Gender and education: the evidence on pupils in England*. London: HMSO.

Department for Education [DfE] (2010). *Phonics teaching materials: core criteria and the self-assessment process*. London: HMSO.

Department for Education [DfE] (2012a). *Early Years Foundation Stage, Department for Education*. London: HMSO. https:www.education.gov.uk/publications/standard/AllPublications/Pages/DFE-OOO23-2012

Department for Education [DfE] (2012b). *Criteria for assuring high-quality phonic work*. London: HMSO.

Department for Education [DfE] (2013a). *National Curriculum for English Key Stages 1 and 2*. London: HMSO.

Department for Education[DfE] (2013b). *Phonics Screening Check and National Curriculum Assessments at Key Stage 1 in England 2012/13*. London: HMSO.

Department for Education [DfE] (2014a*). National Curriculum for English Key Stages 1 and 2*. London: HMSO.

Department for Education (2014b). *National Curriculum Assessments at Key Stage 2*. London: HMSO.

Devonshire, V., Morris, P., & Fluck, M. (2013). Spelling and reading development: the effect of teaching children multiple levels of representation in their orthography. *Learning and Instruction*, 25, 85-94.

Diliberto, J.A., Beattie, J.R., Flowers, C.P., & Algozzine, R.F. (2008). Effects of Teaching Syllable Skills Instruction on Reading Achievement in Struggling Middle School Readers. *Literacy Research and Instruction*, 48(1), 14-27.

Dingfelder, H.E., & Mandell, D.S., (2011). Bridging the research-to-practice gap in autism interventions: an application of diffusion of innovation theory. *Journal of Autism and Developmental Disorders*, 41(5), 597-609.

Dombey, H. (1999). Picking a path through the phonics minefield. *Education 3-14: International Journal of Primary, Elementary and Early Years Education*, 27(1), 12-21.

Dombey, H. (2005). Revolutionary reading. *Journal of Curriculum Studies*, 37(2), 209-221.

Downing, J., & Latham, W. (1967). *Evaluating the Initial Teaching Alphabet: A Study of the Influence of English Orthography in Learning to Read and Write*. London: Cassell.

Duff, F.J., Fieldsend, E., Bowyer-Crane, C., Hulme, C., Smith, G., Gibbs, S., & Snowling, M.J. (2008). Reading with vocabulary intervention: evaluation of an instruction for children with poor response to reading intervention. *Journal of Research in Reading*, 31(3), 319-336.

Duff, F.J., & Clarke, P.J. (2011). Practitioner Review: Reading disorders: what are the effective interventions and how should they be implemented and evaluated? *The Journal of Child Psychology and Psychiatry*, 52(1), 3-12.

Duff, F.J., Hayiou-Thomas, M.E., & Hulme, C. (2012). Evaluating the effectiveness of a phonologically based reading intervention for struggling readers with varying language profiles. *Reading and Writing*, 25(3), 621-640.

Dunn, L.M., Dunn, D.M., Sewell, J., Styles, B., Brzyska, B., Shamsan, Y., & Burge, B. (2009). *The British Picture Vocabulary Scale* (3rd ed.). London: GL Assessment.

Ehri, L.C., Nunes, S.R., Stahl, S.A., & Willows, D.M. (2001a). Systematic Phonics Instruction Helps Students Learn to Read: Evidence from the National Reading Panel's Meta-Analysis. *Review of Educational Research*, 71(3), 393-447.

Ehri, L.C., Nunes, S.R., Willows, D.M., Schuster, B.V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001b). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly*, 36(3), 250-287.

Ehri, L.C. (2005). Learning to read words: Theory, Findings and Issues. *Scientific Studies of Reading*, 9(2), 167-188.

Elliott, J.G., & Grigorenko, E.L. (2014). *The Dyslexia Debate*. Cambridge: Cambridge University Press.

Ellis, S. (2007). Policy and research: Lessons from the Clackmannanshire Synthetic Phonics Initiative. *Journal of Early Childhood Literacy*, 7(3), 281-297.

Ellis, S., & Moss, G. (2014). Ethics, education policy and research: the phonics question reconsidered. *British Educational Research Journal*, 40(2), 241-260.

Engbert, R., Nuthmann, A., Richter, E.M., & Kliegl, R. (2005). SWIFT: A dynamic model of saccade generation during reading. *Psychological Review*, 112(4), 777-813.

Ferre, J.M. (1987). Pediatric Central Auditory Processing Disorder: Considerations for Diagnosis, Interpretation and Remediation. *Journal of Academy of Rehabilitative Audiology,* 20, 73-81.

Field, A. (2013) *Discovering Statistics using IBM SPSS Statistics* (4th ed.). London: Sage.

Fiser, J., & Aslin, R.N. (2002). Statistical learning of new visual feature combinations by infants. *Proceedings of the National Academy of Sciences*, 99(24), 15822-15826.

Flynn, N. (2007). Good practice for pupils learning English as an additional language: Lessons from effective literacy teachers in inner-city primary schools. *Journal of Early Childhood Literacy*, 7(2), 177-198.

Foorman, B.R. (1994). The Relevance of a Connectionist Model of Reading for 'The Great Debate'. *Educational Psychology Review*, 6(1).

Foorman, B.R., Francis, D.J., Novy, D.M., & Liberman, D. (1991). How Letter-Sound Instruction Mediates Progress in First-Grade Reading and Spelling. *Journal of Educational Psychology*, 83(4), 456-469.

Foorman, B.R., Francis, D.J., Fletcher, J.M., & Schatschneider, C. (1998). The Role of Instruction in Learning to Read: Preventing Reading Failure in At-Risk Children. *Journal of Educational Psychology*, 90(1), 37-55.

Fraser, J., Goswami, U., & Conti-Ramsden, G. (2010). Dyslexia and Specific Language impairment: The Role of Phonology and Auditory Processing. *Scientific Studies of Reading*, 14(1), 8-29.

Fricke, S., Bowyer-Crane, C., Haley, A.J., Hulme, C., & Snowling, M.J. (2012). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry,* 54(3), 280-290.

Frith, U. (1985). Beneath the surface of developmental dyslexia. *Surface Dyslexia*, 32, 301-330.

Frith, U. (2001). What framework should we use for understanding developmental disorders? *Developmental Neuropsychology*, 20(2), 555-563.

Fuchs, L.S., Fuchs, D., & Hosp, M.K. (2001). Oral Reading Fluency as an Indicator of Reading Comptetence: A Theoretical, Empirical, and Historical Analysis. *Scientific Studies of Reading*, 5(3), 239-256.

Garner, J.K., & Bochna, C.R. (2004). Transfer of a listening Comprehension Strategy to Independent Reading in First-Grade Students. *Early Childhood Education Journal*, 32(2), 69-74.

Gibson, E.J., & Levin, H. (1975). *The Psychology of Reading*. Cambridge, Massachusetts: The M.I.T. Press.

Goodman, K.S. (1967). Reading: A Psycholinguistic Guessing Game. *Journal of the Reading Specialist*, 6, 126-135.

Goodman, K.S. (1976). Reading: A psycholinguistic guessing game. In H. Singer & R.B. Ruddell (Eds.), *Theoretical models and process of reading* (pp497-508). Newark, DE: International Reading Association.

Goodman, K.S. (1989). Whole-Language Research: Foundations and Development. *The Elementary School Journal*, 90(2), 207-221.

Gomez, R.L., & Gerken, L. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, 4(5), 178-186.

Gontijo, P.F.D., Gontijo, I., & Shillock, R. (2003). Grapheme-phoneme probabilities in British English. *Behaviour Research Methods, Instruments & Computers*, 35(1), 136-157.

Gopnik, A., Glynmour, C., Sobel, D.M., Banks, D., Schulz, L.E., & Kushir, T. (2004). A theory of causal learning in children: causal maps and Bays nets. *Psychological Review*, 111(1), 3-32.

Gorard, S., Siddiqui, N., See, B.H. (2015). An evaluation of the 'Switch-on-Reading' literacy catch-up programme. *British Educational Research Journal*, 41(4), 596-612.

Goswami, U. (1999). Causal connections in beginning reading: the importance of rhyme. *Journal of Research in Reading*, 22(3), 217-240.

Goswami, U. (2005). Synthetic Phonics and Learning to Read: A Cross-language Perspective. *Educational Psychology in Practice*, 21(4), 273-282.

Goswami, U., Wang, S.H.L., Cruz, A., Fosker, T., Mead, N., & Huss, M. (2010). Language-universal Sensory Deficits in Developmental Dyslexia: English, Spanish and Chinese. *Journal of Cognitive Neuroscience*, 23(2), 325-337.

Goswami, U., Fosker, T., Huss, M., Mead, N., & Szucs, D. (2011). Rise time and formant transition duration in the discrimination of speech sounds: the Ba-Wa distinction in developmental dyslexia. *Developmental Science*, 14(1), 34-43.

Gough, P.B., & Tunmer, W.E. (1986). Decoding, reading and reading disability. *Remedial and Special Education*, 7, 6-10.

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing Inferences During Narrative Text Comprehension. *Psychological Review*, 101(3), 371-395.

Haring, M.J. & Fry, M.A. (1979). Effect of pictures on children's comprehension of written text. *Educational Communication and Technology*, 27(1), 185-190.

Harn, B., Parisi, D., & Stoolmiller, M. (2013). Balancing Fidelity with Flexibility and Fit: What Do We Really Know about Fidelity of Implementation in School? *Exceptional Children*, 79(2), 181-193.

Harris, C. (1978). *fuzzbuzz a remedial reading scheme*. Oxford: Oxford University Press.

Hassett, D.D. (2008). Teacher Flexibility and Judgement: A multi-dynamic literacy theory. *Journal of Early Childhood Literacy*, 8(3), 295-327.

Hatcher, P.J., Hulme, C., & Ellis, A.W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development*, 65, 41-57.

Hatcher, P.J., Hulme, C., Miles, J.N.V., Carroll, J.M., Hatcher, J., Gibbs, S., Smith, G., Bowyer-Crane, C., & Snowling, M.J. (2006). Efficacy of small group intervention for beginning readers with reading-delay: a randomised controlled trial. *Journal of Child Psychology and Psychiatry*, 47(8), 820-827.

Havelka, J., & Frankish, C. (2010). Is RoAsT Tougher than StEAK/ The effect of case mixing on perception of multi-letter graphemes. *Psihologija*, 43(1), 103-116.

Hay, I., & Fielding-Barnsley, R. (2007). Facilitating children's emergent literacy using shared reading: A comparison of two models. *Australian Journal of Language and Literacy*, 30(3), 191-202.

Hay, I., & Fielding-Barnsley, R. (2009). Competencies that underpin children's transition into early literacy. *Australian Journal of Language and Literacy*, 32(2), 148-162.

Heibert, E.H., Martin, L.A., & Menon, S. (2005). Are There Alternatives in Reading Textbooks? An Examination of Three Beginning Reading Programs. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 21(7), 7-32.

Helenious, P., Uutela, K., & Hari, R. (1999). Auditory stream segregation in dyslexic adults. *Brain*, 122, 907-913.

Hempenstall, K. (1997). The Whole Language-Phonics Controversy: an historical perspective*. Educational Psychology: An International Journal of Experimental Educational Psychology*, 17(4), 399-418.

Her Majesty's Inspectorate of Education [HMIE] (2006). *Pilot inspection of the Education Functions of Clackmannanshire Council in October 2006*. HMIE 2006, SEED, Edinburgh.

Higgins, S., Katsipataki, M., Kokotsaki, D., Coe, R., Major, L.E., & Coleman, R. (2013). *The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit: Technical Appendix*. London: Education Endowment Foundation.

Hill, D.R., King, S.A., Lemons, C.J., & Partanen, J.N. (2012). Fidelity of Implementation and Instructional Alignment in Response to Intervention Research. *Learning Disabilities Research and Practice*, 27(3), 116-124.

Hindson, B., Byrne, B., Fielding-Barnsley, R., Newman, C., Hine, D.W., & Shankweiler, D. (2005). Assessment and Early Instruction of Preschool Children at Risk for Reading Disability. *Journal of Educational Psychology*, 97(4), 687-704.

Hodgson Burnett, F. (1905). *A Little Princess*. London: Frederick Warne.

Hoffman, J.V., Sailors, M., & Patterson, E.U. (2002). Decodable Texts for Beginning Reading Instruction: The Year 2000 Basals. *Journal of Literacy Research*. 34(3), 269-298.

Høien-Tengesdal, I. (2010). Is the Simple View of Reading too Simple? *Scandinavian Journal of Educational Research*, 54(5), 451-469.

Holliman, A.J., Wood, C., & Sheehy, K. (2010). The contribution of sensitivity to speech rhythm and non-speech rhythm to early reading development. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 30(3), 247-267.

Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *British Medical Journal*, 319, 670-674.

House of Commons Education and Skills Committee (2005). *Teaching Children to Read, Eighth Report of Session 2004-05*. House of Commons, London: The Stationery Office.

House of Commons (2011). *Report of the Inquiry into Overcoming the Barriers to Literacy.* London: The Stationery Office.

Huang, L.Y. (2014). Learning to Read with the Whole Language Approach: The Teacher's View. *English Language Teaching*, 7(5), 71-77.

Huey, E.D. (1908). *The Psychology and Pedagogy of Reading*. New York: The Macmillan Company.

Hutchison, D., & Styles, B. (2010). *A guide to running randomised controlled trials for educational researchers*. Slough: NFER.

Ialongo, N., Poduska, J., Werthamer, L. & Kellam, S. (2001). The Distal Impact of Two First-Grade Preventive Interventions on Conduct Problems and Disorder in Early Adolescence. *Journal of Emotional and Behavioural Disorders*, 9(3), 146-160.

Imrie, A. (2008). Rhythm as a resource to generate phonetic and phonological coherence in lists. *York Papers in Linguistics Series 2*, 9, 23-47.

Inhoff, A.W., Radach, R., Eiter, B.M., & Skelly, M. (2003). Exterior Letters Are Not Privileged in the Early Stage of Visual Word Recognition During Reading: Comment on Jordan, Thomas, Patching and Scott-Brown (2003). *Journal of Experimental Psychology, Learning, Memory and Cognition*, 29(5), 894-899.

Jenkins, J.R., Peyton, J.A., Sanders, E.A., & Vadasy, P.F. (2004). Effects of Reading Decodable Texts in Supplemental First-Grade Tutoring. S*cientific Studies of Reading*, 8(1), 53-85.

Johnston, A.M., & Barnes, M.A. (2008). Reading Comprehension: Developmental Processes, Individual Differences and Interventions. *Canadian Psychology*, 49(2), 125-132.

Johnston, F.R. (2000). Word Learning in Predictable Text. *Journal of Educational Psychology*, 92(2), 248-255.

Johnston, R.S., & Morrison, M. (2007). Towards a Resolution of Inconsistencies in the Phonological Deficit Theory of Reading Disorders: Phonological Reading Difficulties Are More Severe in High-IQ Poor Readers. *Journal of Learning Disabilities*, 40(1), 66-79.

Johnston, R.S., & Watson, J.E. (2005). A Seven Year Study of the Effects of Synthetic Phonics Teaching on Reading and Spelling Attainment*. Insight 17.* Scottish Executive Education Department, Edinburgh: SEED.

Joseph, L.M., & Schisler, R.A. (2007). Getting the "Most Bang for Your Buck". *Journal of Applied School Psychology*, 24(1), 69-90.

Juel, C., & Roper-Schneider, D. (1985). The Influence of Basal Readers on First Grade Reading. *Reading Research Quarterly*, 20(2), 134-152.

Kaiser, A.P., & Hemmeter, M.L. (2013). Treatment Fidelity in Early Childhood Special Education Research: An Introduction to the Special Issue. *Journal of Early Intervention*, 35(2), 79-84.

Keller-Margulis, M.A. (2012). Fidelity of implementation framework: A critical need for response to intervention models. *Psychology in the Schools*, 49(4), 342-352.

Kendeou, P., Savage, R., & van den Broek, P. (2009). Revisiting the simple view of reading. *British Journal of Educational Psychology*, 79, 353-370.

Kucirkova, N., Messer, D., & Sheehy, K. (2014). Reading personalised books with preschool children enhances their word acquisition. *First Language*, 34(3), 227-243.

Kuhl, P. (2004). Early Language Acquisition: Cracking the speech code. *Nature Reviews Neuroscience*, 5, 831-843.

Kuppen, S., Huss, M., Fosker, T., Fegan, N., & Goswami, U. (2011). Basic Auditory Processing Skills and Phonological Awareness in Low-IQ Readers and Typically Developing Controls. *Scientific Studies of Reading*, 15(3), 211-243.

Lacey, P., Layton, L., Miller, C., Goldbart, J., & Lawson, H. (2007). What is literacy for students with severe learning difficulties? Exploring conventional and inclusive literacy. *Journal of Research in Special Educational Needs*, 7(3), 149-160.

Lapin, S., & Shieber, S.M. (2007). Machine Learning theory and practice as a source of insight into universal grammar. *Journal of Linguistics*, 43(2), 393-427.

Lavidor, M. (2011). Whole-word shape effect in dyslexia. *Journal of Research in Reading*, 34(4), 443-454.

Leong, V.,Hamalainen, J., Soltesz, F., & Goswami, U. (2011). Rise time perception and detection of syllable stress in adults with developmental dyslexia. *Journal of Memory and Language*, 64, 59-73.

Lever-Chain, J. (2008). Turning boys off? Listening to what five-year-olds say about reading. *Literacy*, 42(2), 83-91.

Levin, H., & Williams, J. (Eds.) (1970). *Basic Studies in Reading*. New York: Basic Books.

Literacy Task Force (1997). *A Reading Revolution: How We Can Teach Every Child to Read Well. The Preliminary Report of the Literacy Taskforce*. London: University of London, Institute of Education.

Logan, S., & Johnston, R. (2010). Investigating gender differences in reading. *Educational Review*, 62(2), 175-187.

Lucas, P.J. (1997). Jabberwocky back to Old English: Nonsense, Anglo-Saxon and Oxford. *Language History and Linguistic Modelling*, 503-520.

Mace, R.M. (2014) *weebee Reading Programme*. Worcester: New Alphabet.

Mackay, T. (2007). *Achieving the Vision: The Final Research Report of the West Dunbartonshire Literacy Initiative*. West Dunbartonshire Council: Dunbarton.

Mandell, D.S., Stahmer, A.C., Shin, S., Xie, M., Reisinger, E., & Marcus, S. (2013). The role of treatment fidelity on outcomes during a randomised field trial of an autism intervention. *Autism*, 17(3), 281-295.

Marshall, B. (2011). English in the National Curriculum: a simple redraft or a major rewrite? *The Curriculum Journal*, 22(2), 187-199.

Mathes, P.G., & Torgesen, J.K. (1999). A Call for Equity in Reading: A Response to Allington and Woodside-Jiron*. Educational Researcher*, 29(6), 4-14.

Mayer, R.E. (2011). Applying the science of learning to multimedia instruction. In P. Mestre & B.H. Ross (Eds.), *The Psychology of Learning and Motivation, Cognition in Education*, 55, (pp77-106), London: Academic Press.

Maxwell, S.E., & Delaney, H.D. (2008). *Designing experiments and analysing data* (2nd ed.). New York, NY: Routledge Academic.

McGeown, S.P., Johnston, R.S., & Medford, E. (2012). Reading instruction affects the cognitive skills supporting early reading development. *Learning and Individual Differences*, 22, 360-364.

McGeown, S.P., & Medford, E. (2014). Using method of instruction to predict the skills supporting initial reading development: insight from a synthetic phonics approach. *Reading and Writing*, 27(3), 591-608.

McNally, J., & Murray, W. (1962). *Key Words to Literacy*. London: The Schoolmaster Publishing Co.

Mesmer, H.A.E. (2001). Decodable text: A review of what we know. *Reading Research and Instruction*, 40(2), 121-141.

Mesmer, H.A.E. (2009). Textual Scaffolds for Developing Fluency in Beginning Readers: Accuracy and Reading Rate in Qualitatively Levelled and Decodable Text. *Literacy Research and Instruction*, 49(1), 20-39.

Mohr, K.A. (2006). Children's Choices for Recreational Reading: A three-Part Investigation of Selection Preferences, Rationales and Processes. *Journal of Literacy Research,* 38(1), 81-104.

Morgan, P.L., Fuchs, D., Compton, D.L., Cordray, D.S, & Fuchs, L.S. (2008). Does Early Reading Failure Decrease Children's Reading Motivation? *Journal of Learning Disabilities*, 41(5), 387-404.

Morris, D., Tyner, B., & Perney, J. (2000). Early Steps: Replicating the Effects of a First-Grade Reading Intervention Program. *Journal of Educational Psychology*, 92(4), 681-693.

Morris, R. (1963). *Success and Failure in Learning to Read*. London: Oldbourne.

Mosely, D. (2004). The diagnostic assessment of word recognition and phonic skills in five-year-olds. *Journal of Research in Reading*, 27(2), 132-140.

Moustafa, M. (1995). Children's productive phonological recoding. *Reading Research Quarterly*, 30, 464-476.

Moustafa, M., & Maldonado-Colon, E. (1999). Whole-to-part Phonics Instruction: Building on What Children Know to Help Them Know More. *The Reading Teacher*, 52(5), 448-458.

Mowbray, C.T., Holter, M.C., Teagure, G.B., & Bybee, D. (2003). Fidelity Criteria: Development, Measurement and Validation. *American Journal of Evaluation*, 24(135), 315-340.

Muter, V., Hulme, C., Snowling, M., & Taylor, S. (1997). Segmentation, not rhyming, predicts early progress in learning to read. *Journal of Experimental Child Psychology*, 65, 370-396.

Muter, V., Hulme, C., Snowling, M.J., & Stevenson, J. (2004). Phonemes, Rimes, Vocabulary and Grammatical Skills as Foundations of Early Reading Development: Evidence From a Longitudinal Study*. Developmental Psychology*, 40(5), 665-681.

Nahachewsky, J. (2013). Understanding the importance of ethos in composing the "everyday" new literacies classroom. *Language and Literacy*, 15(1), 74-92.

Nation, K., & Hulme, C. (1997). Phonemic segmentation, not onset-rime segmentation, predicts early reading and spelling skills. *Reading Research Quarterly*, 32, 154-167.

Nation, K., & Snowling, M.J. (1997). Assessing reading difficulties: the validity and utility of current measures. *British Journal of Educational Psychology*, 72, 549-560.

Nation, K., & Snowling, M.J. (1998). Individual Differences in Contextual Facilitation: Evidence from Dyslexia and Poor Reading Comprehension. *Child Development*, 69(4), 996-1011.

National Institute of Child Health and Human Development [NICHHD]. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction* (NIH Publication No. 00-4769). Washington, DC: U.S. Government Printing Office.

National Literacy Trust (2012). *Boy's Reading Commission*. National Literacy Trust: London.

Nelson, M.C., Cordray, D.S., Hulleman, C.S., Darrow, C.L., & Sommer, E.C. (2012). A Procedure for Assessing Intervention Fidelity in Experiments Testing Education and Behavioural Interventions. *The Journal of Behavioural Health Services & Research,* 39(4), 374-396.

Newton, M., & Thomson, M. (1982). *Aston Index* (Revised ed.), Cambridge: Learning Development Agency.

O'Donnell, C.L. (2008). Defining, Conceptualizing and Measuring Fidelity of Implementation and its Relationship to Outcomes in K-12 Curriculum Intervention Research. *Review of Educational Research*, 78(1), 33-84.

Ofsted (2012). *Moving English Forward*. Ofsted: Manchester.

Papanicolaou, A.C., Simos, P.G., Breier, J.I., Wheless, J.W., Mancias, P., Baumgartner, J.E., Maggio, W.W., Gormley, W., Constaninou, J.E.C., & Butler, I.I. (2001). Brain Plasticity for Sensory Linguistic Functions: A Functional Imaging Study Using Magneto encephalography With Children and Young Adult. *Journal of Child Neurology*, 16, 241-252.

Pearson, P.D. (2004). The Reading Wars. *Educational Policy*, 18(1), 216-252.

Pickett, L. (1998). Literacy Learning During Block Play. *Journal of Research in Childhood Education*, 12(2), 225-230.

Pitchford, N.J., Ledgeway, T., & Masterton, J. (2008). Effect of orthographic procession letter position encoding. *Journal of Research in Reading*, 31(1), 97-116.

Potter, B. (1902). *The Tale of Peter Rabbit*. London, Frederick Warne.

Poulsen, M. (2008). Do dyslexics have auditory input processing difficulties? *Applied Psycholinguistics*, 32, 245-261.

Protopapas, A., Simos, P.G., Sideridis, G.D., & Mouzaki, A. (2012). The components of the simple view of reading: A confirmatory factor analysis. *Reading Psychology*, 33, 217-240.

Protopapas, A., Mouzaki, A., Sideridis, G.D., Kotsolakou, A., & Simos, P.G. (2013). The Role of Vocabulary in the Context of the Simple View of Reading*. Reading & Writing Quarterly: Overcoming Learning Difficulties,* 29(2), 168-202.

Rayner, K., Johnson, R.L., & Perea, M. (2007). Transposed-letter effects in reading: evidence from eye movements and parafoveal preview. *Journal of Experimental Psychology: Human Perception and Performance*, 33(1), 209-229.

Rayner, K., Slattery, T.J., Drieghe, D., & Liversedge, S.P. (2011). Eye movements and word skipping during reading: Effects of word length and predictability. *Journal of Experimental Psychology: Human Perception and Performance*, 37(2), 514-528.

Readence, J.E., & Baldwin, R.S. (1978). The Relationship of Cognitive Style and Phonics Instruction*. The Journal of Educational Research*, 72(1), 44-51.

Reed, D.K., Cummings, K.D., Schaper, A., & Biancarosa, G. (2014). Assessment Fidelity in Reading Intervention Research: A Synthesis of the Literature. *Review of Educational Research*, 20(10), 1-47.

Reichle, E.D., Reineberg, A.E., & Schooler, J.W. (2010). Eye Movements During Mindless Reading. *Psychological Science*, 21(9), 1300-1310.

Ricketts, J., Nation, K., & Bishop, D.V.M. (2007). Vocabulary is Important for Some, but Not All Reading Skills. *Scientific Studies of Reading*, 11(3), 235-257.

Ricketts, J., Bishop, D.V.M., Pimperton, H., & Nation, K. (2011). The Role of Self-Teaching in Learning Orthographic and Semantic Aspects of New Words. *Scientific Studies of Reading*, 15(1), 47-70.

Ritter, S.M., Damian, R.I., Simonton, D.K., VonBaaren, R.B., Strick, M., Derks, J., & Dijksterhuis, A. (2012). Diversifying experiences enhance cognitive flexibility. *Journal of Experimental Social Psychology*, 48(4), 961-964.

Rochelle, J., Feng, M., Gallagher, H., Murphy, R., Harris, C., Kamdar, D., & Trinidad, G. (2014). *Recruiting Participants for Large-Scale Random Assignemtn Experiemnts in School Settings*. Menlo Park, CA: SRI International.

Romberg, A.R., & Saffran, J.R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 906-914.

Rose, J. (2006). *Independent review of the teaching of early reading*. London: Department for Education and Skills.

Rumelhart, D.E. (1989). The Architecture of Mind: A Connectionist Approach. In M. Posner (Ed.), *Foundations of Cognitive Science* (pp. 133-159). Cambridge, MA: MIT Press.

Rutter, M., Caspi, A., Fergusson, D., Horwood, L.J., Goodman, R., Maughan, B., Moffitt, H., Meltzer, H., & Carroll, J. (2004). Sex differences in developmental reading disability. *Journal of the American Medical Association*, 291(16), 2007-12.

Sadoski, M., & Paivio, A. (2001). *Imagery and text: A dual coding theory of reading and writing*. Mahwah, NJ: Lawrence Erlbaum.

Savage, R., Burgos, G., Wood, E., & Piquette, N. (2015). The Simple View of Reading as a framework for national literacy initiatives: a hierarchical model of pupil-level and classroom-level factors. *British Educational Research Journal*, 41(5), 820-844.

Schonell, F.J. (1971). *Reading and spelling tests*. Edinburgh: Oliver & Boyd.

Shannon, P., & Crawford, P. (1997). Manufacturing Descent: Basal Readers and the Creation of Reading Failures. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 13(3), 227-245.

Shannon, P. (2001). Basal Readers: Three Perspectives. *Theory into Practice*, 28(4), 235-239.

Share, D.L. (1995). Phonological recoding and self-teaching: sine qua non of reading acquisition. *Cognition*, 55(2), 151-218.

Silby, A., & Watts, M. (2015). Making the tacit explicit: children's strategies for classroom writing. *British Educational Research Journal*, 41(5), 801-819.

Silverman, R.D., Speece, D.L., Harring, J.R., & Ritchey, K.D. (2013). Fluency Has a Role in the Simple View of Reading. *Scientific Studies of Reading*, 17(2), 108-133.

Siraj-Blatchford, I. (2009). Conceptualising progression in the pedagogy of play and sustained shared thinking in early childhood education: a Vygotskian perspective. *Educational and Child Psychology*, 26(2), 77-89.

Slavin, R.E. (2003). A Reader's Guide to Scientifically Based Research*. Educational Leadership*, 60(5), 12-16.

Slavin, R.E, Lake, C., Chambers, B., Cheung, A., & Davis, S. (2009). Effective Reading Programs for the Elementary Grades: A Best-Evidence Synthesis. *Review of Educational Research*, 79(4), 1391-1466.

Smith, E.B., Goodman, K.S., & Meredith, R. (1976). *Language and Thinking in School* (2nd ed.). New York: Holt, Rinehart and Winston.

Smith, F. (1973). *Understanding reading: A psycholinguistic analysis of reading and learning to read* (2nd ed.). New York: Holt, Rinehart & Winston.

Smith, S.W., Daunic, A.P., & Taylor, G.G. (2007). Treatment Fidelity in Applied Educational Research: Expanding the Adoption and Application of Measures to

Ensure Evidence-Based Practice. *Education and Treatment of Children*, 30(4), 121-134.

Snowling, M.J., Gallagher, A., & Frith, U. (2003). Family Risk of Dyslexia Is Continuous: Individual Differences in the Precursors of Reading Skill. *Child Development*, 74(2), 358-373.

Snowling, M., Stothard, S.E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E., Nation, K., & Hulme, C. (2009). *York Assessment of Reading for Comprehension.* London: GL Assessment.

Snowling, M.J., & Hulme, C. (2011). Evidence-based interventions for reading and language difficulties: Creating a virtuous circle*. British Journal of Educational Psychology*, 81, 1-23.

Snowling, M.J., & Hulme, C. (2012). Children's reading impairments: From theory to practice. *Japanese Psychological Research*, 55(2), 186-202.

Solar, J., & Openshaw, R. (2007). 'To be or not to be?': The politics of teaching phonics in England and New Zealand. *Journal of Early Childhood Literacy*, 7(3), 333-352.

Solity, J., & Vousden, J. (2009). Real books vs reading schemes: a new perspective from instructional psychology. *Educational Psychology: An International Journal of Experimental Educational Psychology*, 29(4), 469-511.

Southgate, V., Arnold, H., & Johnson, S. (1981). *Extending Beginning Reading*. London: Heinemann.

Stephen, C. (2010). Pedagogy: the silent partner in early years learning. *Early Years,* 30(1), 1-14.

Strand, S., Malmberg, L., & Hall, J. (2015). *English as an Additional Language (EAL) and educational achievement in England: An analysis of the National Pupil Database*. London: Educational Endowment Foundation.

Stokes, S.F., Kern, S., & DosSantos, C. (2012). Extended Statistical Learning as an account for slow vocabulary growth. *Journal of Child Language*, 39(1), 105-129.

Stuart, M., Stainthorp, R., & Snowling, M.J. (2008). Literacy as a complex activity: deconstructing the simple view of reading. *Literacy*, 42(2), 59-66.

Stuebing, K.K., Barth, A.E., Cirino, P.T., Francis, D.J., & Fletcher, J.M. (2008). A Response to Recent Reanalysis of the National Reading Panel Report: Effects of Systematic Phonics Instruction Are Practically Significant. *Journal of Educational Psychology*, 100(1), 123-134.

Sylva, K. (1984). A Hard-headed look at the fruits of play*. Early Child Development and Care*, 15(2), 171-183.

Thomson, J.M., Fryer, B., Maltby, J., & Goswami, U. (2006). Auditory motor rhythm awareness in adults with dyslexia. *Journal of Research in Reading*, 29(3), 334-348.

Torgerson, C.J., Brookes, G., & Hall, J. (2006). *A Systematic Review of the Research Literature in the use of phonics in the teaching of reading and spelling*. London: Department for Education and Skills (DfES).

Torgerson, D.J., & Torgerson, C.J. (2008). *Designing Randomised Trials*. New York: Palgrave Macmillan.

Tormanen, M.R.K., & Takala, M. (2009). Auditory processing in developmental dyslexia: An exploratory study of an auditory and visual matching training program with Swedish children with developmental dyslexia. *Scandinavian Journal of Psychology*, 50, 277-285.

Tracey, L., Chambers, B., Slavin, R.E., Hanley, P., & Cheung, A. (2014). *Success for All in England: Results From the Third Year of a National Evaluation.* York: University of York, Institute for Effective Education.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., & Richmond-Welty, E.D. (1995). The special role of rimes in the description, use, and acquisition of English orthography*. Journal of Experimental Psychology: General*, 124, 107-136.

Treiman, R., & Kessler, B. (2006). Spelling as statistical learning: using consonantal context to spell vowels*. Journal of Educational Psychology*, 98(3), 642-652.

Tunmer, W.E., & Hoover, W.A. (1993). Phonological recoding skill and beginning reading. *Reading and Writing: An Interdisciplinary Journal*, 5, 161-179.

Tunmer, W.E., & Chapman, J.W. (2012). The Simple View of Reading Redux: Vocabulary Knowledge and the Independent Components Hypothesis. *Journal of Learning Disabilities*, 45(5), 453-466.

Vandermosten, M., Boets, B., Luts, H., Poelmans, H., Golestani, N., Wouters, J., & Chesquiere, P. (2010). Adults with dyslexia are impaired in categorizing speech and nonspeech sounds on the basis of temporal cues. *PNAS,* 107(23), pp10389-10394.

Vaughan Van Hecke, A., Mundy, P., Block, J.J., Delgado, C.E.F., Parlade, M.V., Pomares, Y.B., & Hobson, J.A. (2012*).* Infant responding to joint attention, executive processes and self-regulation in preschool children. *Infant Behaviour and Development*, 35(2), 303-311.

Vellutino, F.R. (1991). Introduction to Three Studies on Reading Acquisition: Convergent Findings on Theoretical Foundations of Code-Oriented Versus Whole-Language Approaches to Reading Instruction. *Journal of Educational Psychology*, 83(4), 437-443.

Vousden, J.I., Ellefson, M.R., Solity, J. & Chater, N. (2011). Simplifying reading: applying the simplicity principle to reading. *Cognitive Science*, 35(1), 34-78.

Vygotsky, L. (1978). *The Role of Play in Development, Mind in Society*. Cambridge, MA: Harvard University Press.

Wade, N., & Tatler, B. (2005*). The Moving Tablet of the Eye: The origins of modern eye movement research*. Oxford: Oxford University Press.

Wang, H.C., Nickels, L., Nation, K., & Castles, A. (2013). Predictors of Orthographic Learning of Regular and Irregular Words. *Scientific Studies of Reading*, 17(5), 369-384.

Walker, M., Bartlett, S., Betts, H., Sainsbury, M., & Mehta, P. (2013). *Evaluation of the Phonics Screening Check: First Interim Report*. National Foundation for Educational Research, London: NFER.

Walker, M., Bartlett, S., Betts, H., Sainsbury, M., & Worth, J. (2014). *Phonics screening check evaluation*. National Foundation for Educational Research, London: NFER.

Walker, M., Sainsbury, M., Worth, J., Bamforth, H. & Betts, H. (2015). *Phonics screening check evaluation: Final report*. National Foundation for Educational Research, London: NFER.

Wallach, G.P. (2011). Peeling the Onion of Auditory Processing Disorder: A Language/Curricular-Based Perspective*. Language, Speech and Hearing Services in Schools*, 42, 273-285.

Walsh, M. (2003). ‘Reading’ pictures: what do they reveal? Young children’s reading of visual texts. *Literacy*, 37(3), 123-130.

Warrington, M.J., & George, P. (2014). Reading for pleasure in paradise: paired reading in Antigua and Barbuda. *Literacy,* 48(2), 66-71.

Watts, Z., & Gardner, P. (2013). Is systematic synthetic phonics enough? Examining the benefit of intensive teaching of high frequency words in a year one class. *Education 3-13: International Journal of Primary, Elementary and Early Years Education*, 41(1), 100-109.

Wedell, K. (2014). Points from the SENCo Forum: A dilemma in ‘personalised’ teaching. *British Journal of Special Education*, 41(1), 105-6.

Weiser, B.L. (2013). Ameliorating Reading Disabilities Early: Examining an Effective Encoding and Decoding Prevention Instructions Model. *Learning Disability Quarterly*, 36(3), 161-177.

Wenz-Gross, M., & Upshur, C. (2012). Implementing a Primary Prevention Social Skills Intervention in Urban Preschools: Factors Associated with Quality and Fidelity. *Early Education and Development*, 23(4), 427-450.

White, I.R., Carpenter, J., & Horton, N.J. (2012). Including all individuals is not enough: Lessons for intention-to-treat analysis. *Clinical Trials*, 9, 396-407.

White, S.J. (2008). Eye movement control during reading; effects of word frequency and orthographic familiarity. *Journal of Experimental Psychology: Human perception and Performance,* 34(1), 205-223.

Wilkins, A., Cleave, R., Grayson, N., & Wilson, L. (2009). Typography for children may be inappropriately designed*. Journal of Research in Reading*, 32(4), 402-412.

Willam, J. (2009). Revisiting Susan Isaacs – a modern educator for the twenty-first century*. International Journal of Early Years Education*, 17(2), 151-165.

Williams, J.C. (2014). Recent official policy and concepts of reading comprehension and inference: the case of England's primary curriculum. *Literacy*, 48(2), 95-102.

Wilson, J., & Colmar, S. (2008). Re-evaluating the Significance of Phonemic Awareness and Phonics in Literacy Teaching: The Shared Role of School Counsellors and Teachers. *Australian Journal of Guidance and Counselling*, 18(2), 89-105.

Wolery, M. (2011). Intervention Research: The importance of fidelity measures. *Topics in Early Childhood Special Education*, 31, 155-157.

Wright, C., Conlon, E.G., Wright, M., & Dyck, M.H. (2011). An Open, Pilot Study of the Understanding Words Reading Intervention Program. *SAGE Open*. Retrieved February 6, 2012, from http://dx.doi.org/10.1177/2158244011420452.

Wyse, D., & Styles, M. (2007). Synthetic phonics and the teaching of reading: the debate surrounding England's 'Rose Report'. *Literacy*, 41(1), 35-41.

Ziegler, J.C., & Goswami, U. (2005). Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory. *Psychological Bulletin*, 131(1), 3-29.