A Comparison of Vocabulary Instruction Using Word Meaning Versus a Combined Sound-Meaning Approach on Outcomes of Vocabulary, Phonemic Awareness and Nonword Reading in 5-6 Year Olds

Rose Brooks

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

The University of Sheffield

Faculty of Medicine, Dentistry and Health

Department of Human Communication Sciences

April 2022

TABLE OF CONTENTS

Abstract

Vocabulary yields a powerful influence on children's attainment, behaviour, mental health and life prospects. Given the large number of primary-age pupils with limited vocabulary, there is a need for effective models of whole-class instruction. Teaching has traditionally focussed on word meaning to improve reading comprehension. However, evidence also points to links between oral vocabulary and phonemic skills, which could potentially be harnessed to improve not only vocabulary outcomes but also wider phonemic awareness and phonic reading. Interest is growing in a teaching approach that engages children in learning about the sound structure of the words, alongside the traditional emphasis on meaning. This thesis investigates the relative impacts of this combined method compared to meaning-based pedagogy.

A two-phase research design was implemented with the full range of Year One children (age 5-6) in mainstream UK classrooms. To further investigate relationships between vocabulary, phonemic and word-level literacy, a cross-sectional study was conducted with 152 children. Results confirmed existing findings that vocabulary significantly predicts performance in phonemic awareness and phonic reading in this age group.

The cross-sectional results informed an experimental study with 273 Year One pupils to causally test the impact of combined sound-meaning versus meaning-only instruction, compared to an age-matched waiting control group receiving treatment as usual. A 26 week programme was delivered involving explicit daily vocabulary teaching linked to storybooks, based on evidence-based principles of effective vocabulary instruction, differing only on the cue type provided. Testing across three timepoints determined that combined instruction produced significantly higher results on taught vocabulary and phonic reading than the meaning-based intervention or control group receiving usual instruction, as well as significantly higher phonemic awareness outcomes than the control group.

The current thesis enhances the research base by providing experimental evidence of the effectiveness of the combined approach for whole-class mainstream vocabulary instruction in the early years of schooling.

Acknowledgements

I extend my heartfelt gratitude to the dear friends, family, work colleagues and the HCS department who provided practical, financial and moral support for my PhD and Master's research. Thank you to everyone who went out of their way to contribute to this research project and to support the important aim of improving children's education.

- Thank you Jenny and Meesha - my wonderful, intelligent and supportive PhD supervisors. You were perfect for me, and I feel so grateful that you were my mentors
- I also wish to thank the rest of the Human Communication Sciences Department for their inspiration as well as for awarding me the departmental scholarship that made the research viable
- My appreciation is given to my insightful managers at Babcock Education for the financial support that enabled a large-scale testing programme, as well as my travel and accommodation in Sheffield over the years
- My sincere gratitude to Widgit Software and the authors of Wordaware who allowed free use of their resources to support children's vocabulary development
- I am particularly indebted to my colleagues past and present who did not flinch (much) at the mammoth programme of testing, marking and other activities that made this research possible
- Special recognition goes to my legion of testers - Maggie, Janice and Kate deserve special recognition for sticking with the project to the very end
- Enormous thanks to all the schools, families and pupils who took part in the project and especially to the participating teachers who were my partners in the research process
- I am so grateful to my patient statistics tutors. Believe it or not, I have grown to love stats!
- Thank you to my friends who listened endlessly, marked assessments and provided fun diversions along the way
- Heartfelt gratitude to my family, who always took an interest in my research, provided encouragement and helped in every practical way imaginable
- However, saving the best for last, my deepest thanks are for my husband David whose enduring practical and emotional support made this research possible.

List of Tables

List of Figures

List of Abbreviations

ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
ATL	Association of Teachers and Lecturers
CLPE	Centre for Literacy in Primary Education
CPD	Continuing Professional Development
CR	Correct Rejections
DBS	Disclosure and Barring Service
DCFS	Department for Children, Families and Schools
DFE	Department for Education
DFES	Department for Education and Schools
EAL	English as an Additional Language
EEF	Education Endowment Fund
EHCP	Education, Health and Care Plan
FSM	Free School Meals
HRS	Homogeneity of Regression Slopes
IDACI	Income Deprivation Affecting Children Index
IMD	Index of Multiple Deprivation
NASUWT	National Union of Schoolmasters Union of Women Teachers
NFER	National Foundation for Educational Research
OFSTED	Office for Standards in Education
PAN	Pupil Admission Number
PIN	Pupil Identification Number
RCT	Randomised Control Trial
RQ	Research Question
SALT	Speech and Language Therapist
SD	Standard Deviation
SEMH	Social, Emotional and Mental Health
SEN	Special Educational Needs
SES	Socioeconomic Status
SLCN	Speech, Language and Communication Needs
SPSS	Statistical Package for the Social Sciences
STAR	Select, Teach, Apply, Review
TA	Teaching Assistant

Assessment abbreviations and their descriptions are available in Appendix B.

Presentations and Publications Arising from the Thesis

Presentations

Association of Teachers and Lecturers: Regional Event (March 2015) *High quality teaching approaches for vocabulary.*

Babcock Annual Primary English Conference: Reading Revisited (March 2017) *Vocabulary development and support*.

Pupil Premium Network, Exeter Consortium (October 2018) *Supporting vocabulary.*

HCS Postgraduate Research Conference (July 2017) *Poster: Piloting an assessment battery for a Vocabulary intervention study.*

Vocabulary Conference with Jean Gross CBE: The Wonder of Words (November 2019) *Vocabulary development and support.*

Developmental Research Cluster (July 2020) *A comparison of combined sound-meaning vocabulary instruction versus semantic instruction in 5-6 year olds.*

LinkLive conference (May 2021) *Vocabulary teaching using sound and meaning cues*.

HCS Postgraduate Research Conference (July 2021) *The impact of vocabulary instruction on vocabulary, phonemic awareness and nonword reading in 5-6 year olds.*

TechEd Showcase SEN AT Bitesize webinar (March 2022) *Supporting language and literacy attainment through vocabulary teaching in Key Stage One*.

NAPLIC Conference (May 2022) *The impact of a combined sound-meaning approach for whole-class vocabulary teaching.*

Education Show North (June 2022) *Building a vocabulary recovery plan for schools.*

Livewell SouthWest Children's Speech and Language Therapy Service Research Group (July 2022) *Sound and meaning cues to support whole-class vocabulary instruction.*

Publications

Babcock LDP Newsletter (October 2014) *The vital role of vocabulary*.

Speechlink Blog (May 2021) *The impact of vocabulary teaching using sound and meaning cues*.

Widgit Newsletter (June 2021) *The impact of symbols on vocabulary attainment.*

Widgit Resources (May 2022) *Sound and meaning vocabulary teaching cards.*
https://www.widgit.com/resources/literacy-language/vocabulary/index.htm

Declaration

I Rosemarie Brooks confirm that the thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (**www.sheffield.ac.uk/ssid/unfair-means**). This work has not been previously presented for an award at this, or any other, university.

Chapter 1: Thesis Introduction and Overview

1.1 Rationale for the Study

Words are the building blocks of spoken and written communication. A steadily increasing literature highlights the important role that oral vocabulary plays in a range of life outcomes, including academic achievement, behaviour, mental well-being and employment. Vocabulary is a strong predictor of school attainment (Conti-Ramsden, et al., 2009; Kurdek & Sinclair, 2001; Roulstone et al., 2011; Scarborough, 2001; Spencer et al., 2017; Stothard et al., 1998), largely due to its impact on literacy development (Schoon et al., 2010). Individuals with low levels of vocabulary are at greater risk of delinquent behaviour and criminality. Two-thirds of those with behavioural difficulties across a wide age range are known to have significant gaps in oral language, including pupils at risk of exclusion, young offenders and the prison population (Bryan et al., 2007; Bryan et al., 2015; Clegg et al., 2009; Coles et al., 2017; Ripley & Yuill, 2005; Winstanley et al., 2018). The link between vocabulary and mental health was examined by Cohen et al. (1998), who found that 40% of 7-14 year olds referred to psychiatric services had significantly low standardised receptive and expressive vocabulary scores. Armstrong et al. (2016) further discovered a significantly increased risk of affective disorders, illegal substance misuse and alcoholism in young adults aged 21 whose low vocabulary profiles had persisted from age 5. A relationship also exists between vocabulary and economic prosperity. A large-scale analysis by Law et al. (2009) reports that 5 year olds with typical non-verbal ability but poor receptive vocabulary are 50% more likely to be unemployed or to have literacy or mental health difficulties at age 34. Blanden (2006) demonstrated that age 5 receptive vocabulary was a main predictor of whether a child reached income above the poverty line at age 30, defined as 60% below the UK average income. The existing literature, based largely on correlational and prediction studies, suggests that strong vocabulary may act as a protective factor against a range of adverse outcomes. Given its wide-ranging influence, vocabulary is an important area to target for instruction. Without the right support, vocabulary difficulties are likely to persist throughout a child's primary education, into secondary school (Spencer et al., 2012) and beyond (Law et al., 2009).

Despite the established importance of vocabulary acquisition, research indicates that it is a neglected area of pedagogy (Biemiller & Slonim, 2001; Marulis & Neuman, 2010). As a rule, little systematic and explicit vocabulary teaching takes place in mainstream education, characterised instead by incidental learning encounters (Blachowicz, et al., 2006; Watts, 1995). As an illustration, data from 23 Canadian upper primary classrooms revealed that just 6% of curriculum time was dedicated to vocabulary learning during literacy lessons and a further 1% in other subjects (Scott et al., 2003). Another example of the dearth and perhaps also the quality of vocabulary instruction can be seen in an investigation of 5-6 year olds by Christian et al. (2000), in which chronological age emerged as a significant predictor of receptive vocabulary outcomes, whilst the amount of time spent in school did not - a finding confirmed by Biemiller and Slonim (2001). Teachers are becoming increasingly aware of the need for professional development in this area, as demonstrated in a workforce survey showing that 100% of teachers recognised the importance of children's oral language, yet only 4% felt sufficiently trained to teach these skills (Communication Trust, 2017). Improvement in the quantity and quality of vocabulary teaching is a valuable goal to enhance outcomes for a range of learners.

Large numbers of children experience restricted vocabulary development. A recent survey of Year One teachers across the UK reported that 49% of pupils were perceived to have limited vocabulary to the extent that it affected school attainment (Oxford University Press, 2018). A range of papers indicate that the highest prevalence occurs in areas of socioeconomic disadvantage. A study of low-income pupils at school entry by Locke et al. (2002) determined that 56% had oral vocabulary at least a standard deviation below the population mean. A similar figure emerged in a preschool investigation by King et al. (2005), showing that 49% of children in poverty had clinically low scores on an oral language composite that included vocabulary.

Pupils with specific language needs, or *Developmental Language Disorder* (DLD) are also vulnerable in terms of their vocabulary acquisition, affecting 7.6% of children at the start of schooling (Norbury et al., 2016). The terms *Specific Language Impairment* (SLI) or *Language Disorder* (LD) were previously favoured in the literature to denote this cohort. To delineate, the more general category of *Speech, Language and Communication Needs* (SLCN) refers to the wider group of pupils with low levels of oral language of varying aetiology.

An additional 2.3% of school-age children have weak oral language associated with Special Educational Needs (SEN), such as hearing impairment, neuro-developmental disorders, Cerebral Palsy and intellectual impairment (ibid.). Research conducted by the University of Sheffield estimates that a further 0.5% of pupils have severe language and vocabulary needs (Communication Matters, 2013).

Pupils learning English as an Additional Language (EAL) represents one in five pupils nationally (Department for Education-DFE, 2020). This group often displays significantly lower vocabulary performance across the school years (Bialystok et al., 2010; Burgoyne et al., 2009; Cameron, 2002), albeit with wide variation (Schmitt, 2010). These figures suggest a sizeable cohort with limited vocabulary who may benefit from vocabulary instruction.

It is not feasible to withdraw such large numbers of pupils for targeted vocabulary intervention, suggesting that an inclusive classroom-based model may be more expedient. A meta-analysis by Marulis and Neuman (2010) reveals that whole-class vocabulary instruction is no less successful than individual and small group models, with large effects shown for all group sizes. High-quality classroom teaching therefore has the potential to provide an efficient and effective model of vocabulary instruction for the large cohort who require this support.

Early intervention is often considered the key to reducing vocabulary disparities. Research has highlighted that children in disadvantaged areas experience a reduced quantity and variety of vocabulary input (Hart & Risley, 2003). Since the vocabulary gap begins early in infancy (Fernald et al., 2013), by the onset of schooling there is already an average differential of 18 months between the highest and lowest economic groups (Waldfogel & Washbrook, 2010). Lower starting points and a slower learning trajectory often lead to the word gap widening over the school years (Biemiller & Slonim, 2001), culminating in deficient outcomes for many children. It is therefore prudent to offer enhanced vocabulary input at an early phase of education to optimise language and literacy attainment for all pupils.

A wealth of experimental evidence demonstrates that explicit instruction is an effective method for improving vocabulary (e.g. Beck & McKeown, 2007; Coyne et al., 2022; Fukkink & de Glopper, 1998). A meta-analysis of 67 vocabulary studies conducted in the early years of primary school (Marulis & Neuman, 2010) found that explicit vocabulary teaching led to significantly higher gains (p<.001) with nearly double the effect size[1] (g=.111) of incidental encounters (g=.62). Explicit instruction thus offers a powerful tool for improving vocabulary outcomes.

To further optimise the impact of vocabulary instruction, it is useful to understand the types of input that contribute to word learning. Vocabulary instruction in schools has traditionally focussed on the *semantic meaning* of words due to its proven effectiveness for enhancing reading comprehension (Clarke et al., 2010; Department for Children, Family and Schools, 2008; National Reading Panel, 2000; Stahl & Fairbanks, 1986). However, experimental results also point to the effectiveness of a *combined sound-meaning approach* that considers both word meaning *and* the sound structure of words, referred to as the *combined approach* in the current thesis. This evidence is predominantly based on pupils with language difficulties (see meta-analysis by Wisenburn & Mahoney, 2009), although emerging mainstream studies indicate superior vocabulary gains compared to semantic instruction or controls (Damhuis et al., 2016; Droop et al., 2005; Janssen et al., 2018; Moran & Moir, 2018; Silverman, 2007).

There is robust evidence to show that vocabulary is a significant predictor of phonemic awareness and phonic reading skill in younger pupils (e.g. Wagner et al., 1997). This has prompted interest by a number of researchers (Dickinson et al., 2003; Duff et al., 2015; Lonigan, 2007; Munro

---

[1] Hedges g, often used in meta-analyses to control for bias in small sample sizes, follows the same conventions for magnitude of effect as Cohen's d: 0.2=small, 0.5=medium, 0.8=large.

et al., 2008) as to whether a combined vocabulary approach emphasising the phonological form of words as well as meaning might stimulate not only vocabulary learning but also wider gains in phonemic awareness and word-level reading. Thus, vocabulary knowledge is conceptualised to underpin both oral vocabulary and literacy development. The term *phonological* in the current thesis conveys its broader psycholinguistic connotation, referring to the sound structure of words (Stackhouse & Wells, 1997), whereas *phonological awareness* denotes the ability to attend to and consciously manipulate the sound structure of spoken words, as distinct from meaning (ibid.).

Teaching vocabulary with an additional focus on sound structure is likely to have its greatest impact during the early primary school years when phonological learning is at its peak (Wagner et al., 1997). Hyde Wright et al. (1993) consider this a “prime phonological time when a child is beginning to learn to read and when phonological coding is perhaps more dominant” (p. 223). It is worth investigating whether combined vocabulary instruction can affect distal outcomes of phonemic awareness and phonic reading in this age group, since this would afford an additional opportunity to acquire and apply these foundational literacy skills.

An approach that delivers equivalent vocabulary gains to traditional meaning-based instruction as well as supplementary gains in phonemic skills would constitute an efficient and effective pedagogy. The fact that the combined sound-meaning method is already being implemented in some primary school settings as part of a whole-school vocabulary approach (e.g. Parsons & Branaghan, 2014) offers a further impetus for timely evaluation.

1.2 Thesis Overview

The current thesis aims to extend the literature and enhance evidence-based practice in the fundamental area of vocabulary instruction. The overarching aims are to:

1. confirm the relationships between vocabulary, phonemic awareness and decoding in the early years of schooling
2. examine the causal impact of whole-class vocabulary instruction on vocabulary and wider outcomes of phonemic awareness and decoding for the full range of mainstream pupils.

To fulfil these aims, a pilot study and two empirical research phases were implemented with Year One pupils (ages 5-6). An initial pilot study was undertaken with 20 children to ascertain the feasibility and sensitivity of assessment measures needed for the empirical studies. A cross-sectional study of 152 pupils examined the predictive relationships between vocabulary, phonemic awareness and phonic reading described in the literature. It also provided an opportunity to trial the assessment measures intended for the experimental research phase - a controlled experiment conducted over a 26 week intervention period with 272 Year One pupils across 11 schools. Using a repeated measures design over three testing points (pre-intervention, post-intervention and maintenance four months later), the impact of vocabulary instruction on outomes of vocabulary, phonemic awareness and decoding was measured in the three groups, i.e. semantic (meaning-based) instruction, combined (sound-meaning) instruction and a waiting control group receiving usual vocabulary teaching.

The content of the thesis falls broadly into four sections. The thesis and its rationale are introduced in the current chapter (Chapter 1). The investigation commences with the cross-sectional study (Chapters 2-8). The first literature review (Chapter 2) appraises the evidence base regarding the role of sound and meaning in vocabulary acquisition, leading to a second literature review (Chapter 3) evaluating the body of prediction studies on the relationships between vocabulary, phonemic awareness and phonic reading. The research questions and hypotheses for the cross-sectional study are then posed (Chapter 4). An overview of the pilot study is briefly presented (Chapter 5), followed by the cross-sectional methods (Chapter 6), results (Chapter 7) and discussion (Chapter 8). The next section relates to the main experimental vocabulary study (Chapters 9-14). The third literature review evaluates the evidence base for the combined sound-meaning approach (Chapter 9), moving on to a fourth literature review of the principles underpinning effective vocabulary intervention (Chapter 10) and the research questions and hypotheses for the experimental phase (Chapter 11). Subsequent chapters set out the method (Chapter 12), results (Chapter 13) and discussion (Chapter 14) of the experimental phase. The thesis culminates in the overall discussion and conclusions of the research study (Chapter 15), including practical implications and suggested avenues for future research in the vital field of vocabulary instruction.

1.3 Chapter Summary

A main driver for the current thesis is the far-reaching impact of vocabulary acquisition on children's attainment and life outcomes. A universal model of classroom instruction will be considered due to the large cohort requiring vocabulary support. Since the vocabulary gap starts early in life, it is optimal to intervene in the initial school years, using an explicit model of instruction to deliver the greatest chance for improvement. The predictive relationships between vocabulary, phonological awareness and decoding documented in the literature (Wagner et al., 1997) indicate that it may be valuable to consider whether vocabulary intervention can promote wider literacy-related outcomes in addition to increased vocabulary. A cross-sectional study will examine these predictive relationships in the Year One age group, with outcomes informing the design of the experimental vocabulary intervention.

We now turn to the literature reviews to evaluate the theoretical and empirical bases that provide the rationale for exploring a combined sound-meaning approach to vocabulary instruction.

Chapter 2: Literature Review One - The Role of Sound and Meaning in Vocabulary Acquisition

2.1 Chapter Introduction

Word learning entails attention to phonological form, meaning and the associative link between the two (Perfetti & Hart, 2002). Using a psycholinguistic frame of reference, the current literature review will evaluate the respective roles of sound and meaning in the process of learning new vocabulary. First, the nature of lexical (vocabulary) representation will be explored, with a particular focus on phonological and semantic components. The process of children's receptive and expressive vocabulary learning will then be discussed, including the vital influence of memory. The final theme considers the impact of lexical characteristics on the rate of word acquisition.

2.2 The Nature of Lexical Representation

The mental lexicon can be conceptualised as an interconnected body of word knowledge. In typical vocabulary acquisition, a range of semantic, syntactic, phonological and orthographic information is stored in memory as a *lexical representation*, (Gupta, 2005; Marslen-Wilson, 1987; Stackhouse & Wells, 1997). The abstract nature of lexical representation creates difficulties for the definition, operationalisation and direct observation of this latent construct. The current discussion will examine three key facets: the *phonological representation*, considered to store information about the word's sound structure, the *semantic representation,* incorporating the meaning of the word and the *associative link* between sound and meaning.

The phonological representation is viewed to contain information about the sound structure of a spoken word. The quality of a phonological representation has been defined in terms of its distinctness and segmental structure (Metsala & Walley, 1998). According to Elbro (1996), *distinctness* refers to the magnitude of difference between a lexical item and similar sounding *phonological neighbours*, i.e. the number of words differing by a phoneme. *Segmental quality* refers to an unconscious sensitivity to the subdivisions in a word, including the whole word, large segments of syllables and rhyme, as well as phonemes (Metsala & Walley, 1998). In practical terms, distinctness and segmental quality are not separable. So, for example, when a child pronounces "particliar" for the word 'particular', the phonological representation could be underspecified in terms of distinctness, segmental quality or both (Elbro & Palleson, 2002). The purpose of implicit phonological representations is to support speech perception and production (ibid); however, it is also thought that phonological representations act as an important precursor to later developing explicit phonological awareness skills needed for literacy (ibid; Werker & Tees, 1999).

Lexical representation also includes a semantic component, i.e. the concept or meaning that a spoken word symbolises. Infants' first words are often labels for whole objects. As vocabulary expands, attention begins to focus on attributes of size, colour, category and function (Clark, 2015; Markman & Wachtel, 1988). Children start to group objects with similar attributes together into categories, which over time develop into complex semantic networks or taxonomies, described by Murphy (2010) as comprising a superordinate category (such as animals), subcategories (farm animals) and members of the category (cow). Syntactic information about how the word functions in a sentence can be considered part of the semantic representation, alongside social and experiential information (ibid).

The associative link, or the connection between phonological and semantic representations, is a vital component of vocabulary learning (Gupta, 2005). The mutual influence of phonological and semantic aspects of representation provides strong cues for word encoding and retrieval (Storkel & Morrisette, 2002). The two aspects typically develop in tandem, although phonological form and semantic meaning can at times be dissociated, for example in the case of word finding difficulties (Alt & Plante, 2006). The integral link between sound and meaning is captured in Perfetti's lexical quality hypothesis: "a name without meaning and a meaning without a name are both low quality" (1985, p. 118). The associative link is incrementally enriched through subsequent encounters with the word (Dollaghan, 1987).

2.3 The Role of Sound and Meaning in Vocabulary Learning

Marslen-Wilson suggests that "to understand spoken language is to relate sound to meaning" (1987, p.72). Vocabulary learning is a complex and demanding task, requiring the child to analyse the phonological and semantic information available, to map one on to the other and to store the linked information in memory.

Word comprehension and expression are generally depicted as separate though inter-related processes. In typical development, receptive understanding tends to precede expressive word use (Schmitt, 2010). *Receptive vocabulary* entails recognition of a word in spoken or written form, alongside some degree of comprehension of its meaning. *Expressive vocabulary* involves the naming of a word, i.e. production of its phonological form, also in relation to meaning. In curriculum terms, receptive vocabulary exerts its greatest influence on listening and reading, whilst expressive vocabulary impacts most on speaking and writing (ibid.). Developmental models of word comprehension and production, largely based on the school-age population, will now be considered as these provide the greatest relevance for the current study sample.

2.3.1 Receptive Vocabulary Learning in Children

The process of understanding language begins with the extraction of words from the continuous stream of speech (Werker & Curtin, 2005). Research indicates that infants use a range of cues to locate word boundaries, including perceptual cues such as stress patterns (Echols & Newport, 1992; Jusczyk et al., 1999) and the position of sounds within a word (Jusczyk et al., 1999). Evidence suggests that lexical cues are also utilised, for example recognising common sound patterns (Cairns et al., 1997; Jusczyk & Aslin, 1995; Mattys & Jusczyk, 2001; Saffran et al., 1996) and using familiar words to anchor down new vocabulary in continuous speech (Bortfeld et al., 2005).

Receptive vocabulary recognition can be theoretically explained in terms of distributed connectionist models (e.g. Gaskell & Marslen-Wilson, 1997), with the proviso that these are largely based on adult samples. Within this paradigm, the speed and accuracy of lexical access would depend on the specificity of the lexical representation. The spoken word activates similar-sounding lexical entries, leading to competition and ultimately inhibition of the unneeded entries.

Oral word recognition is supported by the well-researched process of *fast mapping* (Carey, 1978), in which a phonological word form is recognised and linked to a semantic meaning to forge an initial lexical representation of the spoken word – a similar concept to the previously described associative link. Dollaghan (1985) investigated the nature and quantity of information encoded during fast mapping with 35 typically developing children aged 2-6. After a single exposure to a novel word, 81% of this age group demonstrated receptive understanding by pointing to the matching picture. This is a robust finding that has been replicated in other research studies (Dickinson, 1984; Heibeck & Markman, 1987; Oetting et al., 1995). After two exposures, 45% of Dollaghan's sample could produce the word, suggesting that additional exposures were needed to form a phonological representation adequately distinct and stable for spoken word production. Older children experienced greater success on both tasks than younger participants. A number of quasi-experimental studies indicate that children with language difficulties require more exposures than typically developing youngsters to facilitate fast mapping, particularly for expressive word production (Dollaghan, 1987; Gray, 2005; Rice et al., 1990).

Two semantic principles may serve to narrow the number of preliminary meanings to be activated upon hearing a new word. According to the *whole object principle*, the child will seek out an entire object as the referent (Markman, 1990). The *mutual exclusivity principle* proposes that the child will initially assume that a referent has only one name – a useful process thought to accelerate vocabulary growth (Golinkoff et al., 1994). Mutual exclusivity is also credited with initiating the comprehension of attributes: when the child has a pre-existing label for a presented object and a novel label is introduced, the child may seek the meaning in an attribute or part of an object (Markman & Wachtel, 1988).

After sufficient word encounters, the final stage of lexical consolidation can take place. As opposed to the relatively quick process of fast mapping, researchers generally agree that lexical consolidation requires a more extended time course linked to overnight sleep (Dumay & Gaskell, 2007; Smith et al., 2018). In one investigation, Smith et al. (2017) familiarised participants aged 8-13 with 16 real words, each with a nonword alternative (e.g. dolphin/dolpheg). Testing took place immediately after training and again 24 hours later, alongside sleep monitoring. Significant improvements were seen after sleep on both receptive and expressive vocabulary recall tasks, and this result was maintained one week later.

2.3.2. Expressive Vocabulary Learning in Children

There is a large degree of consensus that the stored receptive representation serves as the basis for the expressive representation (Dollaghan, 1985; Gray, 2005; Gupta, 2005; Schmitt, 2010). Research on word production suggests that it is a multistage process (Dell et al., 1997). Using the example of picture naming, a conceptual meaning would activate the semantic representation, which in turn promotes retrieval of the stored phonological representation and access to the motor programme for word production (Gupta, 2005; Stackhouse et al., 2007).

Most researchers posit a high level of interaction (and mutual support) between the semantic and phonological aspects of word naming (Dell & O'Seaghdha, 1992; Indefry & Levelt, 2004; Rapp & Goldrick, 2000). Whilst comprehension can proceed on the basis of an imprecise or underspecified phonological representation, production is likely to require a more accurately encoded word form (Fowler et al., 2004; Stackhouse & Wells, 1997). An implication for vocabulary instruction is that a concise phonological representation may provide better support for expressive word use.

2.3.3 The Role of Phonological Memory in Vocabulary Learning

Learning new vocabulary is heavily reliant on phonological short-term memory. The prominent working memory model by Baddeley et al. (1998) postulates three components: the *central executive* that co-ordinates two subsidiary processing systems – the *visuo-spatial sketchpad* for handling and storing visual information and the *phonological loop* facilitating the temporary storage of verbal information while it is processed for semantic understanding. Subvocal rehearsal appears to play an important role in maintaining the initial phonological representation in memory long enough for new lexical items to be encoded. This phenomenon is exemplified in a study by Horst and Samuelson (2008), which discovered significantly reduced performance on toddlers' recall of receptive and expressive vocabulary after a short delay of five minutes. Opportunities for overt rehearsal of new vocabulary thus appear to provide valuable support for initial encoding of words in memory.

Nonword repetition is widely considered to be an optimal measure of phonological memory. Nonword items are unfamiliar, so children require accurate phonological encoding and storage to produce a correct response. However, stored lexical knowledge can also affect performance (Roodenrys & Hinton, 2002; Thomson et al., 2005), such that the more a nonword resembles a real word, the greater the likelihood that lexical knowledge can support new word learning. The phonotactic probability of nonwords also affects repetition latency and accuracy (Vitevich & Luce, 1999).

Gathercole and Baddeley (1989) found nonword repetition to be moderately correlated with receptive vocabulary in 5 year olds (r=.57). Even after controlling for word reading, nonverbal intelligence and age in their hierarchical regression equation, nonword repetition still predicted 21% of the variance in vocabulary outcomes. Results highlight the need to measure the influence of phonological memory in terms of its contribution to vocabulary, phonemic awareness and decoding outcomes in the current investigation.

The role of phonological memory in new word learning appears to change over the course of development. A longitudinal study by Gathercole et al. (1992) discovered that nonword repetition yielded significant moderate correlations[2] with receptive vocabulary in children aged 4, 5 and 6 (r=.52 - .56), controlling for age and nonverbal intelligence; but thereafter the relationship was no longer statistically significant. Phonological memory thus remains a relevant variable to assess in the

[2] Pearson's correlations will be interpreted as 0.2=weak, 0.5=moderate, 0.8=strong (Field, 2013) in the current thesis.

current 5-6 year old sample, after which time the increasing store of lexical knowledge is likely to play an increasing role in new word learning (ibid.).

2.4 Variables Affecting Vocabulary Learning

Studies of typically developing children point to a range of variables that have the potential to influence the rate of lexical acquisition. Word-level characteristics such as word length, grammatical part of speech, spoken word frequency and the regularity of sound patterns in our language all influence vocabulary learning. As previously discussed in section 2.3.3, the child's existing lexicon also exerts a powerful effect on new word learning.

2.4.1 The Effect of Word-Level Variables on Learning New Words

Word length is a key characteristic affecting the ease of word learning. Word length is commonly measured by the number of constituent phonemes or syllables. Studies of school-age children indicate that recognition is easier for shorter words than for longer words, a finding replicated in gating tasks involving the incremental presentation of a spoken word until the participant has enough information to recognise the word (Metsala et al., 2009). Experimental paradigms have also confirmed the word length effect (Gathercole et al., 1999).

Word type, or grammatical part of speech, also influences the rate of vocabulary learning. Nouns are simplest to acquire, due in part to their concrete mapping between perception and meaning (Gentner, 1982; Hadley et al., 2016). Lexical principles, including the whole-object constraint and mutual exclusivity, can also help to explain the rapid acquisition of nouns (Oetting et al., 1995). Since actions and attributes have less consistency and concreteness, they are typically learned later and more slowly (Maratsos, 1990).

Concreteness and *imageability* are inter-related characteristics affecting word learning. As stated above, grammatical class affects the level of concreteness, as well as how easily the word can be imaged (Gilhooly & Logie, 1980). Despite the importance of these variables, norms are only available for a limited number of words (Coltheart, 1981), thus restricting their use in the current study.

Spoken word frequency is arguably the most important variable affecting the ease of learning a new lexical item (Luce & Pisoni, 1998). Frequency denotes the number of times a child is likely to be exposed to a word in the communication environment. Higher frequency words occur in conversation and are generally learned sooner than the less common words in our language (ibid; Storkel, 2004). Written frequency norms (Kucera & Francis, 1967) were used in the current study instead of spoken norms due to the larger size of database.

A closely linked though still separable construct (Charles-Luce & Luce, 1995) is *age of acquisition* (AoA). A longitudinal study by Walley and Metsala (1990) examined the effect of this variable on spoken word recognition. Participants were tested three times at chronological ages 5, 8 and adult on a mispronunciation detection task, in which words are either spoken correctly or with minimal differences. Results demonstrated that at age 5 children were significantly more accurate in identifying mispronunciations for words with an early AoA than for mid or later acquired words. The 8 year olds were significantly more sensitive to mispronunciations for early and mid AoA than later acquired words. Adult responses did not differ according to the AoA variable. Findings are consistent with the idea that phonological representations for words with an early AoA may be more robustly specified than later acquired items, making them easier to recognise.

Two further lexical variables relate to the regularity of sound patterns in our language. *Phonotactic probability* reflects the relative frequency of segments and sound sequences in syllables and words (Vitevitch & Luce, 1999). The corollary is phonological *neighbourhood density*, referring to the number of similar-sounding words for a lexical item, generated by the addition, deletion, or substitution of one phoneme in any word position (Luce & Pisoni, 1998). The higher the incidence of a sound pattern (phonotactic probability), the greater the number of words that contain that segment (neighbourhood density).

The phonotactic probability of a word's sound patterns affects performance on vocabulary tasks in children (Storkel, 2001; Storkel & Rogers, 2002) as well as adults (Storkel et al., 2006).

Storkel (2001) investigated whether the effect of phonotactic probability on new word learning changes over the course of early childhood. Nonwords were presented to 3-6 year olds, with targets containing high vs. low probability sound sequences paired with a picture referent. Task performance was significantly better for words with high phonotactic probability across the age range, however there was also an increasing correlation between receptive vocabulary and phonotactic probability, suggesting that phonotactic knowledge accrues in tandem with increases in vocabulary size. Nonword experiments with school age children have similarly concluded that phonotactic probability significantly affects new word learning (Gathercole et al., 1992; 1999).

Phonotactic probability is closely connected to the concept of *neighbourhood density*. In typical development, phonological similarities are detected between new and existing words in the lexicon. As the density of similar-sounding words increases, the child must distinguish between these words, resulting in more detailed representations of the shared segments (Storkel, 2004). Predictably, words with many similar sounding neighbours (dense neighbourhoods) contain commonly occurring phonological sequences (ibid), as well having a greater frequency of occurrence and earlier age of acquisition (Charles-Luce & Luce, 1990), thus highlighting the interplay between these constructs. A study by Coady and Aslin (2003) exemplifies the developmental changes in the lexicon in terms of the density variable. The lexicons of two 3 ½ year olds and their mothers were examined. The children’s lexicons were characterised by few words with many neighbours. In contrast, the adult lexicons contained infrequent words with rare sound patterns and less neighbours. The researchers’ interpretation is that while vocabulary size is small, children acquire relatively more neighbours for existing items in order to facilitate new word learning.

2.4.2 The Effect of the Existing Lexicon on Learning New Words

A person with a large vocabulary can be expected to have considerably more phonological patterns to draw upon than someone with a smaller lexicon. The more words a child knows containing a particular phonological pattern (phonotactic probability), the stronger the basis for using this representation to quickly access the same pattern in other words to scaffold new word learning. This process, known as *redintegration,* has been shown to support children's lexical acquisition. An experiment with school-aged children by Thomson et al. (2005) demonstrates the impact of stored phonological information on vocabulary learning. Words with more shared segments (dense neighbourhoods) displayed a significant recall advantage over words in sparse neighbourhoods. Results of an investigation of 7-8 year olds by Gathercole et al. (1999) revealed significantly higher recognition of real words over nonwords, both matched for phonotactic probability, further underscoring the important contribution of stored lexical knowledge in receptive vocabulary learning.

Evidence also demonstrates the process of redintegration at work in children's expressive vocabulary development. In a study by Turner et al. (2004) a sample of 5-10 year olds was tested on recall and production using stimuli varied for neighbourhood density. Younger children made significantly greater use of words with many neighbours (dense) to aid recognition of novel words than the older children. Older children made more use of words in dense neighbourhoods for production. The authors' line of reasoning is that as phonological representations become more finely detailed, there is less need for these to support word recognition, although specificity may still be required for more complex expressive tasks.

2.5 Chapter Summary

The current chapter has highlighted the integral role of sound and meaning in new word learning - influencing lexical representation and acquisition, as well as memory processes. Important word-level variables and the existing lexicon also affect the rate of vocabulary learning.

Building upon the theory and processes described in this chapter, the literature review in Chapter 3 will now examine the constructs of vocabulary size, phonemic awareness and decoding and then move on to appraise the prediction literature to explore the empirical relationships between these skills.

Chapter 3: Literature Review Two - The Relationships Between Vocabulary, Phonemic Awareness and Decoding

3.1 Chapter Introduction

The previous chapter explored how sound and meaning are integral to new word learning. The current review explores the relationships that underpin the combined vocabulary teaching approach by discussing the theory and evidence regarding vocabulary's ability to *predict* distal outcomes of phonemic awareness and decoding. The purpose of the review is to construct a likely model of the predictive relationships between vocabulary, phonemic awareness and decoding to investigate in the cross-sectional study, which can in turn inform the design of the experimental research phase. The chapter will first describe the main variables of interest in the current thesis.

3.2 Defining Vocabulary, Phonemic Awareness and Decoding

3.2.1 Vocabulary Size and Depth

An important distinction in the literature relates to the size and depth of vocabulary knowledge. Though not mutually exclusive, vocabulary *size* or breadth denotes the number of words known, whereas *depth* reflects how well a word's meaning is understood. Vocabulary size, more than depth of understanding, appears to be associated with phonemic awareness and decoding skills (Duff et al., 2015; Oulette, 2006). Vocabulary size is generally indexed by picture recognition (receptive vocabulary) or naming tasks (expressive vocabulary).

Vocabulary depth can be measured along a continuum from implicit to explicit knowledge, ranging from (1) no knowledge, (2) a general sense of the word, (3) comprehension without expressive use to (4) rich and decontextualised knowledge (Beck et al., 2013). Rich and decontextualised lexical knowledge is the ultimate aim of vocabulary instruction (ibid) to allow flexible use of vocabulary (Stahl & Fairbanks, 1986). Vocabulary depth is typically assessed using a definition task, although an element of size is arguably also being tested. Incorporating a definitions task in the current assessment battery will enable optimal comparison to other studies.

3.2.2 Phonemic Awareness

Phonological awareness denotes the ability to attend to and consciously manipulate the sound structure of a spoken word, as distinct from its meaning (Stackhouse & Wells, 1997). Large-segment awareness of syllables and rhyme is considered to support vocabulary acquisition (Muter et al., 2004), while *phonemic awareness* is needed for word-level literacy (Ehri et al., 2001). It will be useful to measure rhyme in the current test battery, given its dual links with vocabulary and decoding (Muter et al., 2004). Phonemic awareness is measured by tasks requiring the child to reflect upon or manipulate the sound components of spoken words. According to Cassano and Steiner (2016), tasks can vary across several dimensions, including task difficulty (detection, segmentation, deletion, production), response format (verbal, nonverbal) and extraneous cognitive demands, such as memory load.

Phonemic awareness assessments in younger primary school pupils share a significant moderate to high correlation with each other (r=.68-.86) according to Lundberg et al. (1980). Despite measuring the same central construct of phonemic awareness, there are some notable differences between task types, as demonstrated in a factor analysis by Yopp (1988) showing that phonemic awareness tasks loaded onto two separate factors: (1) simple tasks such as alliteration, blending and segmentation and (2) compound tasks that place a greater burden on working memory, such as phoneme deletion. The significant high correlation between the two factors (r=.77) concurs with Lundberg et al.'s analysis that they nonetheless largely measure the same construct. For the current assessment battery, it will be useful to index phonemic awareness at a range of levels.

3.2.3 Phonic Decoding

Disparate reading measures have been used across the literature, including *nonword decoding*, *real word decoding*, irregular *exception words* and *word recognition* (an amalgamation of decodable words and irregular exception words). To provide clarity in the current review, these specific terms will be used, and a combination of measures will be denoted as *word reading*. In a study by Hulme et al. (2007) a range of reading tests were administered to 66 primary-aged pupils, demonstrating high intercorrelations (*r*=.68-.92). This suggests that even though tests measure slightly different word-level reading skills, they still represent the same basic construct.

The reading-related variable of interest in the current study is phonic decoding, as this is implicated in the relationship between vocabulary, phonemic awareness and word reading (Bowey & Patel, 1988). Nonword reading represents the optimal task to index phonic decoding in the current investigation. Since there is no stored representation for a nonword, the task primarily draws upon bottom-up processing based on perceptual information, although stored lexical items could support performance through the process of redintegration (Roodenrys & Hinton, 2002).

3.3 Theoretical Viewpoints on the Relationship Between Vocabulary, Phonemic Awareness and Decoding

There are two main perspectives regarding the role of vocabulary in the development of phonemic awareness and indirectly in the development of word decoding - the emergent and accessibility positions.

The emergent position, typified by the lexical restructuring hypothesis (Metsala & Walley, 1998), holds that phonological representations are initially holistic in form but become progressively more distinct and segmental as vocabulary size increases. Children gradually develop more sensitivity to word segments, allowing new vocabulary to be stored separately from existing items in the lexicon, first in terms of larger segments and later as separate phonemes (ibid). According to Metsala (1997), this is particularly the case for words that have a similar sound structure to previously stored lexical items, i.e. words in denser neighbourhoods. Perfetti & Hart (2002) point out that lexical representations may also become increasingly delineated as children encounter words that differ minimally in sound, meaning or orthography, ultimately leading to higher lexical quality.

Phonological sensitivity can be described as developing along a continuum from an implicit to explicit level of analysis (Stackhouse & Wells, 1997). Under lexical restructuring theory (Metsala &

Walley, 1998), if implicit representations are specified as separate phonemes, the school age child may begin to consciously use phonemic awareness skills to support the connection to print (Elbro, 1998; Metsala, 1997; Swan & Goswami, 1997; Walley et al., 2003). On a reciprocal basis, letters can also help to refine phonemic awareness ability (Castles & Coltheart, 2004; Metsala & Walley, 1998; Perfetti et al., 1987). Experimental research with children suggests the likelihood that phonological representations are restructured on a word-by-word basis as new items are added to the lexicon. This stands in contrast to explicit phonemic awareness ability, which appears to operate system-wide (Bowey & Hirakis 2006; Cutler, 2008; Elbro & Pallesen, 2002; Hogan, 2010; Metsala, 1997).

The alternative accessibility position postulates perceptual skills (rather than vocabulary size) as the main driver for increased phonological sensitivity, with phonological representations stored as separate sounds from early infancy for the purpose of speech perception (Bailey & Plunkett, 2002; Swingley & Aslin, 2000). Evidence from preferential-looking experiments indicates that infants have the capacity to discriminate between minimally contrasting items, for example baby/vaby – an ability that some researchers present as evidence of phonemically segmented representations (Swingley & Aslin, 2000; Werker et al., 2002). Other researchers offer an opposing viewpoint, suggesting that the task could be completed using perceptual skills to discriminate between two whole word forms, consequently not requiring a phonemically specified representations (Charles-Luce & Luce, 1995; Walley, 1993).

Consistent with the lexical restructuring model (Metsala & Walley, 1998), findings from a range of research paradigms confirm that phonological representations undergo a gradual process of refinement over the course of childhood. Gating experiments demonstrate that with increasing age children require less phonetic information to recognise a word (Metsala, 1997; Garlock, et al., 2001). Mispronunciation detection and production tasks have similarly shown a cumulative effect of phonemic context over the course of childhood, often lasting up to age 7 or 8 (Nittrouer & Studdert-Kennedy, 1987; Treiman & Breaux, 1982). Further evidence for restructuring is provided by

Ainsworth et al. (2015) showing that preschoolers with larger vocabularies have more highly specified representations than those with a smaller lexicon.

In the accessibility view, vocabulary size would not be a unique predictor of phonemic awareness development, since representations are considered to be phonemically specified from infancy. The emergent tradition predicts the opposite, that increased vocabulary is related to higher performance on phonemic awareness tasks. Results of correlational and prediction studies presented in the next section provide convincing evidence of the empirical links between vocabulary size and explicit phonemic awareness, endorsing the emergent position.

3.4 Empirical Studies on the Relationship Between Vocabulary, Phonemic Awareness and Decoding

This section will evaluate concurrent and longitudinal research regarding vocabulary's ability to predict phonemic awareness and word decoding. The empirical relationship between phonemic awareness and decoding will also be briefly described. Parameters include a focus on mainstream pupils in the initial years of schooling and studies that include vocabulary as a dependent variable.

3.4.1 The Relationship Between Vocabulary and Phonemic Awareness

Significant moderate correlations between vocabulary size and phonemic awareness are regularly found in the 5-6 year old age group. An investigation by Metsala (1999) of 61 preschoolers aged 4-5 yielded a moderate association between receptive vocabulary and phoneme blending (r=.52), controlling for age. Bowey and Patel (1988) derived a similar magnitude of correlation (r=.48) between receptive vocabulary and a phonemic awareness composite in a sample of 60 pupils aged 6-7. Confirmatory results were provided in a large-scale study (N=533) of 4-5 year olds by Dickinson et al. (2003) in which receptive vocabulary correlated with a composite of rhyme and phonemic tasks (r=.51). Longitudinally, Schatschneider et al. (2004) also demonstrated that age 5 receptive vocabulary correlated with a phonemic awareness composite at age 7 (r=.49), again in a large sample of 540 pupils. Moderate correlations are thus the norm between receptive vocabulary size and a broad spectrum of phonemic awareness tasks in younger pupils.

Fewer studies have assessed the correlation between expressive vocabulary and phonemic awareness. Cassano and Schickedanz (2015) administered a picture naming task to 61 children aged 4-5 years that yielded moderate correlations to a range of phonemic awareness tasks (*r*=.39 to .55). In a study by Wagner et al. (1997) with a 6 year old sample (*N*=216), a definitions task (arguably also a depth measure) yielded correlations to phonemic awareness (*r*=.34-.48) similar to those found by Cassano and Schickedanz. Based on the limited evidence available, moderate correlations are the norm between expressive vocabulary size and phonemic awareness, albeit of a potentially smaller magnitude than correlations between receptive vocabulary and phonemic awareness.

A wealth of research has demonstrated that vocabulary is a unique predictor of phonemic awareness (Bowey & Patel, 1988; Chaney, 1998; Duff et al., 2015; McBride-Chang et al., 1997; Oulette & Beers, 2010; Senechal et al., 2006; Torgesen & Davis, 1996; Wagner et al., 1997). Cross-sectional evidence includes a hierarchical regression analysis by Dickinson et al. (2003) with 4-5 year olds (*N*=533), revealing that receptive vocabulary accounted for 15% of the variance in phonological awareness (a composite of rhyme and phonemic awareness) after controlling for socioeconomic background. A hierarchical regression by Senechal et al. (2006) with a 5-6 year old sample (*N*=84) demonstrated that receptive vocabulary added 4% unique variance to phonemic awareness outcomes, controlling for a large number of variables including age, prior attainment and parent literacy. A likely reason for the divergent results may be the composite phonological awareness measure (including rhyme) employed by Dickinson et al. (2003), which was developmentally appropriate for the preliterate sample and could therefore evidence a stronger relationship than if a pure phonemic awareness measure had been used. Another potential factor in the discrepant outcomes is the differing number of variables entered into the regression models. A larger set of predictors, often with overlapping variance, can lessen the observed impact on the outcome (Howitt & Cramer, 2014).

Longitudinal studies equally highlight vocabulary as a significant predictor of phonemic awareness outcomes. A hierarchical regression analysis on 41 children by Chaney (1998) discovered that receptive vocabulary at age 3 predicted 12% of the variance in age 7 phoneme segmentation and 16% of phoneme deletion, controlling for age and socioeconomic background. Results of a structural equation model by Duff et al. (2015) on data from 300 children demonstrated that 2 year old receptive vocabulary accounted for 4% of variance in later phoneme deletion skill at school age (4-9 year olds). The amalgamation of results across such a wide age range in the Duff et al. study limits the conclusions that can be drawn and may also have reduced the magnitude of the effect, in contrast to Chaney's more homogeneous 7 year old sample. The higher upper age limit in the Duff et al. (2015) sample is another plausible reason for the conflicting results given the dwindling relationship between vocabulary and phonemic awareness.

A variety of research designs have confirmed that vocabulary's influence on phonemic awareness diminishes as phonological skills reach maturity. The previously mentioned longitudinal investigation by Wagner et al. (1997) found that vocabulary definitions remain a stable predictor of phonemic awareness until around age 8, with decreasing returns thereafter. This timeframe resonates with gating experiments indicating the need for significantly less acoustic input as children get older (Garlock et al., 2001; Metsala, 1997). Additional evidence derives from developmental norms documenting that phonemic awareness skills reach ceiling performance by around age 8 (Liberman et al., 1974; Treiman & Zukowski, 1996). Year One (age 5-6) therefore provides an optimal age to benefit from this relationship, linking also with the intensive phonics learning already taking place at this time.

Several preschool investigations did not detect vocabulary as a predictor of phoneme-level awareness. In line with developmental norms (ibid), the phonological representations of this preliterate age group may still be based on larger units of syllables and rhyme. This assertion was supported in a longitudinal study by Carroll et al. (2003), who tested 67 children at age 3-4 and then again at three time points over the next 12 months after starting school. The test battery included large-segment awareness (syllables and rhyme), phonemic awareness and receptive vocabulary. Structural equation modelling indicated that T1 (Time 1) vocabulary predicted T2 large segment awareness; T2 large segments in turn predicted phonemic awareness at T3. The conclusion drawn by the researchers was that large segment sensitivity may be a skill that grows out of normal language development, and that phonemic awareness can build on the foundation of earlier large-segment awareness.

Similarly, a longitudinal study of 90 children by Muter et al. (2004) did not find that receptive vocabulary at school entry (age 4-5) predicted phonemic awareness two years later (age 7-8). However, when floor performance scores on the phonemic measures were excluded, vocabulary did emerge as a significant predictor of phonemic awareness. This can perhaps be explained by the weaker relationship between vocabulary and phonemic awareness in the lower performers, who mirrored younger learners. As previously seen in Carroll et al. (2003), vocabulary was not yet predictive of phonemic awareness in prereaders, although it significantly predicted the earlier-developing skill of rhyme.

Ambiguous results were produced in a large-scale investigation by Lerner and Lonigan (2016) with 358 children aged 4-5. Control variables of age, initial phonemic performance and letter names were included in their growth curve analysis. In this model, vocabulary predicted phoneme blending but not the more difficult phonemic awareness tasks. The theme again arises of a preliterate age group not yet ready for phonemic segmentation, potentially explaining why vocabulary only emerged as a predictor of early blending. Letter name knowledge was a significant predictor of all phonemic awareness tasks for the preschool sample. Studies of school age pupils tell a different story, however, indicating that letter names are established early and hence do not necessarily predict variation in phonemic awareness skill beyond preschool (Byrne & Fielding-Barnsley, 1989; Clayton et al., 2020; Ortiz et al., 2012).

It is generally agreed that the influence of vocabulary on phonemic awareness is not a direct one, but rather that high-quality segmental phonological representations form the basis for explicit phonemic awareness development (Elbro, 1996; Metsala & Walley, 1998; Werker & Tees, 1987). Speech perception tasks, such as the commonly used mispronunciation detection task (Claessen et al., 2009) are considered a valid way to index phonological representation to avoid the confound of speech output. In a principle components analysis by Antony et al. (2010) a mispronunciation detection task administered to 4-5 year olds (*N*=175) showed a high correlation to speech perception, both for changes of consonants (*r*=.91) and vowels (*r*=.93). The relationship of mispronunciation detection to receptive vocabulary was demonstrated in a study of 38 typically developing preschoolers by Ainsworth et al. (2015), who discovered a moderate correlation between receptive vocabulary and mispronunciation detection (*r*=.44). Longitudinal research by Fowler et al. (2004) indicates that the correlation between vocabulary and speech perception was significant at age 6-7 (*N*=93) but not at age 9-10 (*N*=67), echoing the previously discussed diminishing relationship between vocabulary and phonemic awareness.

3.4.2 The Relationship Between Phonemic Awareness and Decoding

Explicit phonemic awareness, in turn, exerts a causal impact on word level literacy, according to a meta-analysis of 52 phonemic awareness intervention studies for the US National Reading Panel (Ehri et al., 2001). The meta-analysis revealed a significant medium effect of phonemic awareness on nonword reading (*d*=.49), word recognition (*d*=.32) and spelling (*d*=.47), with the highest effect sizes for interventions linked to print. Comparable medium effect sizes of intervention were discovered in a meta-analysis by Bus and van IJzendoorn (1999) incorporating 33 studies - for word recognition (*d*=.44), and spelling (*d*=.61). Phonemic awareness explained about 12% of the variance in word recognition, leading the authors to argue that phonemic awareness linked to print is an important but insufficient strategy to support word reading, with vocabulary knowledge promoted as a further vital ingredient.

Phonemic awareness appears to play a mediating role in the vocabulary-decoding relationship (Russell et al., 2018). Clear evidence of mediation can be seen in a structural equation model by Hulme et al. (2015), revealing a significant path from preschool (age 3) naming vocabulary to phonemic awareness at school entry (age 4-5), and then from phonemic awareness to word recognition once literacy instruction was underway (age 5-6).

3.4.3 The Relationship Between Vocabulary and Decoding

Moderate correlations are consistently reported between receptive vocabulary and word reading (Bowey & Patel, 1988; Dickinson et al., 2003; Garlock et al., 2001; Lee, 2011; Kim et al., 2013; Scarborough, 2001; Wagner et al., 1997), analogous to the correlational results between vocabulary and phonemic awareness discussed in section 3.4.1.2.

Fewer investigations have considered correlations for expressive vocabulary and decoding. Using a definitions task with a sample of 67 pupils aged 6-7, Oulette and Beers (2010) found a significant moderate correlation (r=.31) between these skills. A similar magnitude of correlation (r=.25) was discovered by Wagner et al. (1997) in a larger sample (N=216) of 5-6 year olds. Based on the limited research available, expressive vocabulary appears to have a weaker relationship to decoding than the previously described relationship between receptive vocabulary and decoding.

Consistent with results for phonemic awareness, numerous cross-sectional and longitudinal investigations have demonstrated that vocabulary size is a significant predictor of word reading, accounting for a modest amount of variance in performance (Bowey & Patel, 1988; Catts et al., 1999; Dickinson et al., 2003; Duff, et al. 2015; Garlock, et al., 2001; Oullette, 2006; Russell et al., 2018; Schatschneider et al., 2004; Verhoeven et al., 2011; Wagner et al., 1997).

A cross-sectional study by Garlock et al. (2001) serves to highlight vocabulary's ability to predict word recognition. The performance of preschoolers in the US (aged 4-6) was compared with school-age children (aged 7-8) and adults. The sample consisted of 64 participants in each age group. Tests were administered for receptive vocabulary, phonemic awareness, word recognition and nonword repetition (phonological memory). Hierarchical regression analysis showed that receptive vocabulary added 5.5% unique variance to word recognition after controlling for age. Phonemic awareness and phonological memory predicted additional variance in word recognition when entered in later steps in the model. Using this order of entry, it is not possible to say whether the contribution attributed to vocabulary overlapped with the contribution of phonemic awareness, had it been counted prior to vocabulary. It is important for the current study to determine whether additional phonemic awareness during combined instruction will yield greater gains than traditional semantic vocabulary training.

Research by Duff et al. (2015) with 300 British children exemplifies receptive vocabulary's longitudinal ability to predict later word reading (composite of nonword decoding, real word decoding and exception words). First testing around age 2 was followed by a subsequent single testing point somewhere between the ages of 4-9. Structural equation modelling indicated that infant receptive vocabulary accounted for 11% of the variance in a word reading composite at school age. The contribution of vocabulary to word reading in this study is higher than in other papers, perhaps linked to the diverse reading measures in the battery appealing to a wide age range. This makes sense in terms of research showing that the relative contribution of different reading skills changes over time (Gough & Tunmer, 1986; Wagner et al., 1997). The broad age range in this investigation limits interpretation of the developmental relationship between receptive vocabulary and decoding, corresponding to the phonemic awareness outcomes discussed in section 3.4.1.

Numerous researchers have investigated whether vocabulary and phonemic awareness provide shared or separate contributions to decoding outcomes. This is a key consideration for the current intervention design, i.e. to determine whether additional phonemic awareness input in combined instruction could deliver extra gains. Supporting evidence is provided in a hierarchical regression analysis by Dickinson et al. (2003) investigating a large cross-sectional sample of 533 low-income children aged 4-5. Background variables were entered in the first regression step, followed by a phonological awareness composite in the second step and receptive vocabulary in the third step. In this model, both phonological awareness and vocabulary each accounted for about 7% of the total variance in a literacy composite that included word decoding. Whilst the composite assessments for phonological awareness and literacy offer a less precise estimate than a pure decoding measure, results support the possibility that vocabulary and phonemic awareness each contribute discretely to decoding outcomes at the start of schooling.

The unique contributions of both vocabulary and phonemic awareness were further confirmed in a large-scale longitudinal investigation of 604 pupils by Catts et al. (1999) that examined vocabulary (in an oral language composite) and phonemic awareness at age 5 to ascertain word reading outcomes at age 7. Although the oral language composite is an indicative rather than precise measure of vocabulary size, the high correlation between receptive vocabulary and sentence-level skills (r=.73) documented by Lonigan et al. (1998) affirms that combined language scores still strongly represent vocabulary skill.

The predictive role of random automised naming (RAN) in reading outcomes was also explored in the Catts et al. (1999) study. Two hierarchical regression models were fitted, each with three steps. Model one entered variables in the following order: phonemic awareness, RAN pictures and an oral language composite of vocabulary and syntax. After accounting for the effect of phonemic awareness (33%), RAN contributed a further 5%, and then oral language added another 5% of the variation in word reading outcomes. The second model entered the order as oral language, phonemic awareness and then RAN. Accounting for oral language in the first step, phonemic awareness contributed an additional 8% of the variance in decoding outcomes, and RAN added a further 2%. Oral language and phonemic awareness at age 5 thus emerged as unique predictors of word reading at age 7. The study also underscores the importance of RAN as a unique predictor of early word reading (Araújo et al., 2015; Lervag et al., 2009; Wagner et al., 1997). Two other significant predictors of word reading to emerge in the model were nonword repetition and nonverbal intelligence, commending these as further measures to include in the current battery.

Schatschneider et al. (2004) evidenced a small additional contribution of vocabulary to decoding performance in a longitudinal study of 540 American children. Pupils were tested on receptive vocabulary, a phonemic awareness composite and word reading at school entry (age 5-6) and again at age 7-8. Their dominance analysis (a form of multiple regression) entered 14 predictors in the model, with the result that earlier receptive vocabulary accounted for just 1% additional variance in word reading at age 7-8. Unfortunately, the larger number of variables in the model creates difficulties for comparison to other studies and may have influenced the smaller amount of additional variance in word reading.

Equal contributions of vocabulary and phonemic awareness to decoding outcomes are suggested in a cross-sectional investigation by Bowey and Patel (1988) of 60 pupils aged 6-7. In the first regression step, they alternately entered an oral language composite (receptive vocabulary and syntax) and phonemic awareness, each accounting for around 40% of the variance. When the alternate variable was entered as the second step, neither oral language nor phonemic awareness contributed any further variance to word recognition, implying that the measures explained similar variation in the reading outcome for this age group.

The age variable may be an important factor in reconciling the differences between study results. Investigations with younger samples (Dickinson et al., 2003; Catts et al., 1999) demonstrate separable influences of vocabulary and phonemic awareness on decoding performance, whereas they explain the same variance in slightly older pupils with more developed phonemic awareness skills (Bowey & Patel, 1988).

Not every investigation has revealed vocabulary size as a significant predictor of word decoding in the age group under investigation. As previously noted, Hulme et al. (2015) found an indirect relationship via phonemic awareness. In a cross-sectional sample of 304 children aged 6-7, Kim et al. (2013) administered an expressive naming test, which did not add to the prediction of word recognition, although a significant moderate correlation was found (r=.47), commensurate with the extant literature. The multilevel model contained several intercorrelated linguistic variables known to affect word reading performance (ibid), including morphological awareness, orthographic awareness, phonemic awareness, which may have exerted a stronger or shared influence on word recognition. Muter et al. (2004) similarly did not report receptive vocabulary as a significant predictor of word recognition in a longitudinal study of 90 children at school entry (age 4-5), Year 1 (age 5-6) and Year 2 (age 6-7). Despite significant moderate correlations between receptive vocabulary and word recognition at T1 and longitudinally at all time points (r=.40-.71), vocabulary only predicted word recognition at T2 when floor scores on phonemic awareness measures were excluded, thus indicating a more typical 5-6 year old distribution. Once again, the shared or stronger contribution of other predictors in the model (i.e. rhyme, phonemic awareness, letter recognition and grammatical awareness) may also have affected results.

There is scant research to address the issue of whether expressive vocabulary provides analogous results to receptive measures. Oulette and Beers (2010) concurrently measured the contribution of both receptive vocabulary and expressive definitions on nonword reading in a sample of 67 pupils at age 6-7. Using a hierarchical regression model, phonemic awareness was entered as the first step, followed by decoding, irregular word reading and listening comprehension each in separate steps. In the fifth step, neither receptive nor expressive vocabulary entered on an alternate basis demonstrated any further unique contribution to decoding performance. These results differ from previously described papers, almost certainly due to the strength of the decoding predictors entered prior to vocabulary, which unfortunately does not provide comparable information to models where vocabulary was entered much earlier. Since the role of expressive vocabulary in the prediction of decoding remains unclear, it is prudent for the current investigation to extend the literature by including both receptive and expressive size measures in the current test battery.

Robust research confirms a decreasing association between vocabulary and decoding, mirroring its time-limited influence on phonemic awareness. In a longitudinal study of 626 children, Storch and Whitehurst (2002) tested children at six time points from nursery to age 9. Structural equation modelling demonstrated that an oral language composite was a significant predictor of word reading until age 7-8. Wagner et al. (1997) explored a similar age span, assessing 216 pupils over the course of their literacy development from ages 5-10. The definitions measure significantly predicted subsequent word reading up until age 8, after which no further significant effect was detected. A highly powered longitudinal design (*N*=1073) by Lee (2011) equally discovered a lessening influence of vocabulary on word reading over time. Vocabulary naming was assessed at age 2, followed by a battery of tests at three time points (6-7, 8-9, 10-11) for receptive vocabulary, word recognition and nonword reading. MANCOVA was used to compare reading outcomes for those with the largest (>460 words) and smallest (<230 words) lexicons at age 2. After controlling for gender, birth order, ethnicity and socioeconomic status, Lee showed that earlier expressive size continued to predict nonword reading up to age 8-9, a result broadly in line with Storch and Whitehurst (2002) and Wagner et al. (1997). Given the magnitude and rigour of the above studies, there is a high degree of certainty that for typically developing children the influence of vocabulary on decoding is limited to the early school years.

Once phonic decoding becomes established by around age 8 (Lee, 2011; Wagner et al., 1997), there is strong evidence that vocabulary increasingly begins to predict irregular exception word reading (Bowey, 2001; Lee, 2011; Mitchell & Brady, 2013; Nation & Cocksey, 2009; Nation & Snowling, 2004; Oullette, 2006; Oullette & Beer, 2010; Ricketts et al., 2007). Exception word reading, for example the word ‘yacht’, requires access to a stored lexical representation, since words cannot be decoded directly (Ricketts et al., 2007). This shifting relationship between vocabulary and word reading aligns with research showing that as decoding skills mature, oral vocabulary predicts comprehension outcomes more strongly than word reading (Nation & Snowling, 1998; Oulette & Beers, 2010).

To recap, studies of younger school children tend to find that vocabulary and phonemic awareness provide separable contributions to decoding (Catts et al., 1999; Dickinson et al., 2003; Garlock et al., 2001). However, by age 7 the evidence is more controversial, with some researchers observing overlapping contributions (Bowey & Patel, 1998; Schatschneider et al., 2004) and others detecting no contribution of vocabulary to decoding outcomes (Hulme et al., 2015; Muter et al., 2004). A reasonable interpretation is that over the early school years vocabulary’s influence on decoding decreases whilst the role of phonemic awareness increases. This process appears to continue until around age 7-8 when phonological skills typically reach maturity.

3.4.4 The Relationship Between Vocabulary, Phonemic Awareness and Spelling

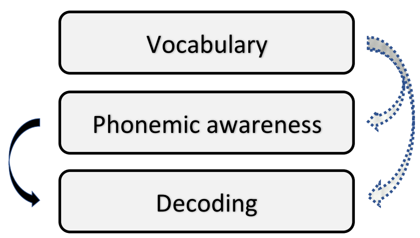
Although not the prime outcome under investigation, the dearth of research whether vocabulary also predicts spelling raises this as an avenue for exploration. Existing studies have yielded moderate correlations between these variables (Kim et al., 2013; Sénéchal & LeFevre, 2002). In a three-year longitudinal study of 153 pupils aged between 5-7 year by Caravolas et al. (2001), a path analysis revealed that receptive vocabulary did not uniquely predict variance in spelling when stronger predictors were entered into the model, including phonemic awareness, letter-sound correspondence, prior reading skill and prior spelling. The cross-sectional investigation of 6-7 year olds by Kim et al. (2013) likewise did not find expressive naming to be a significant predictor of real word spelling once phonemic awareness and morpheme awareness were entered into the model. Based on the limited evidence, the outcome is uncertain. However, vocabulary is unlikely to emerge as a significant predictor of spelling outcomes.

3.5 Chapter Summary

The body of research evaluated in sections 3.4.1 and 3.4.2 confirms that vocabulary, phonemic awareness and decoding are moderately correlated variables. Studies additionally demonstrate that vocabulary uniquely predicts phonemic awareness and word reading in the age group of the current study and in models containing fewer variables. These predictive relationships are depicted by a dotted line in Figure 3.1. It is well established that phonemic awareness exerts a causal influence on decoding (Ehri et al., 2001), indicated by the solid line. Reciprocal relationships have not been included, as these were not the focus of the current review.

Figure 3.1

Relationships Between Vocabulary, Phonemic Awareness and Decoding in the Literature



The relatively modest additional contribution of vocabulary to decoding outcomes depicted in the literature may be partly explained by the covariance between vocabulary and phonemic awareness variables (Bowey & Patel, 1988). Converging evidence exists that vocabulary begins to lose its value in predicting phonological memory, phonemic awareness and decoding outcomes after the initial school years as these skills are becoming more established.

Existing studies have incorporated a range of significant predictors of decoding, most commonly receptive vocabulary, expressive vocabulary, phonemic awareness (and sometimes earlier developing rhyme), nonverbal ability, nonword repetition (phonological memory), RAN and mispronunciation detection (speech perception). These may therefore be useful variables to measure in the current cross-sectional assessment battery. A nonword reading task provides the optimal measure of decoding.

Based on the reviewed literature describing vocabulary's likely influence on phonemic awareness and decoding, Chapter 4 will present the aims, research questions and hypotheses to concurrently test these relationships in a cross-sectional investigation.

Chapter 4: Aims, Research Questions and Hypotheses for the Cross-Sectional Study

4.1 Aims of the Cross-Sectional Study

A study that investigates the relationships between vocabulary, phonemic awareness and decoding will help to validate and further clarify variables influencing word reading. The relative contributions of vocabulary and phonemic awareness to decoding outcomes still needs to be established. Findings would inform an experimental study to test whether these relationships can be harnessed in a vocabulary teaching intervention to improve vocabulary and perhaps broader outcomes of phonemic awareness and decoding. The cross-sectional study seeks to address several gaps in the prediction research literature reflected in the aims below.

4.1.1 Cross-Sectional Research Aim 1

To Incorporate Specific Rather Than Composite Testing Measures

Existing prediction studies have often used amalgamated assessments that do not provide concise information about the relationships between vocabulary, phonemic awareness and nonword literacy. Separate measures will help to fine-tune our knowledge on the role of vocabulary (receptive and expressive), phonemic awareness (alliteration and segmentation) and decoding (nonword reading). Although not a main goal of this thesis, a spelling task will also be included to enhance the scarce evidence base and to enable comparison to the literature on decoding outcomes.

4.1.2 Cross-Sectional Research Aim 2

To Trial Assessment Tools Needed for the Experimental Phase

Importantly, this first empirical study provides a vital opportunity to determine the validity and sensitivity of the assessment battery with the Year One age group in preparation for the main experimental investigation.

4.2 Research Questions and Hypotheses for the Cross-Sectional Study

Research questions (RQ) 1-10 will be addressed through multiple regression analyses with simultaneous entry of variables. Question 1 investigates the predictors of phonemic awareness gleaned from extant research, questions 2-5 explore the predictors of nonword reading, and questions 6-9 examine the predictors of nonword spelling in a comparable regression model. Questions 10-12 utilise a three-step hierarchical regression model to consider the potential influences of independent variables on the proposed intervention groups.

4.2.1 Predictors of Phonemic Awareness

4.2.1.1 Cross-Sectional Research Question 1

Does receptive vocabulary uniquely predict phonemic awareness (alliteration and phoneme segmentation) in 5-6 year olds, accounting for rhyme awareness and speech perception?

Hypothesis 1: Receptive vocabulary is expected to be a significant predictor of phonemic awareness in this age group, even when the other variables are entered simultaneously into the regression model. The literature regularly detects a relationship between receptive vocabulary and phonemic awareness in school-age children (Bowey & Patel, 1988; Hulme et al., 2015; McBride-Chang et al., 1997; Russell, et al., 2018; Senechal et al., 2006; Torgesen & Davis, 1996; Wagner et al., 1997), particularly in younger age groups and in models with fewer predictors. The predictors selected for the first regression equation are precursors to phonemic awareness, i.e. rhyme (Carroll et al., 2003; Muter et al., 2004) and speech perception (Elbro, 1996; Metsala & Walley, 1998).

4.2.2 Predictors of Nonword Reading

4.2.2.1 Cross-Sectional Research Question 2

Does receptive vocabulary uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?

Hypothesis 2: Receptive vocabulary is anticipated to be a unique predictor of decoding in this age group, in line with prediction studies of a similar age group (Duff et al., 2015; Garlock et al., 2001; Russell et al., 2018; Verhoeven et al., 2011).

4.2.2.2 Cross-Sectional Research Question 3

Does expressive vocabulary uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?

Hypothesis 3: No firm prediction can be offered based on the available literature, although given the strong correlation between receptive and expressive size measures (Oulette, 2006), expressive vocabulary could conceivably emerge as a predictor of decoding in this age group.

4.2.2.3 Cross-Sectional Research Question 4

Does alliteration uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?

4.2.2.4 Cross-Sectional Research Question 5

Does segmentation uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?

Hypothesis 4 and 5: It is likely that phonemic awareness will uniquely predict nonword reading, as per numerous prediction and experimental studies (Ehri et al., 2001). There is less certainty as to whether outcomes will differ between alliteration and segmentation. Some literature suggests alliteration to be more aligned with reading and segmentation as a precursor for spelling (Perin, 1983). Studies that have included phoneme segmentation tasks (e.g. Muter et al., 1998) indicate that it is a unique predictor of decoding, however it remains unclear whether it would still be a unique predictor if alliteration were also entered into the model.

4.2.3 Predictors of Nonword Spelling

4.2.3.1 Cross-Sectional Research Question 6

Does receptive vocabulary uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

4.2.3.2 Cross-Sectional Research Question 7

Does expressive vocabulary uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

Hypothesis 6 and 7: It is not possible to predict whether either vocabulary dimension will explain nonword spelling outcomes given the scant evidence available. Although receptive vocabulary and real word spelling are moderately correlated (Kim et al., 2013; Sénéchal & LeFevre, 2002), there is minimal information regarding the ability of vocabulary size to predict phonic spelling in the intended age group. Receptive vocabulary did not emerge as a significant predictor of spelling in Caravolas et al. (2001), however the strength and number of predictors in the model are not comparable to the decoding predictors reviewed in Chapter 3. The lexical restructuring process could theoretically impact on phonic spelling, given the reliance on the same phonological representation as for reading. However, this is unlikely, since spelling requires a higher level of specificity in the phonological representation than reading (Stackhouse & Wells, 1997) and may draw upon more explicit and advanced phonemic awareness skills, such as segmentation.

3.2.3.3 Cross-Sectional Research Question 8

Does alliteration uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

3.2.3.4 Cross-Sectional Research Question 9

Does segmentation uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

Hypothesis 8 and 9: It is highly likely that phonemic awareness will significantly predict spelling performance (Ehri et al., 2001). However, it is not clear whether outcomes will differ for the two phonemic awareness variables. Segmentation, the ability to hear the individual sounds in words, is a potential predictor of spelling skill given its clear and established links to the spelling process (Perin, 1983). As an earlier developing skill, alliteration may also significantly predict spelling outcomes.

4.2.4 Contributions of Phonemic Awareness and Vocabulary to Decoding Outcomes

4.2.4.1 Cross-Sectional Research Question 10

How much contribution do control variables (speech perception, RAN, nonword repetition, rhyme and nonverbal ability) provide to nonword reading in 5-6 year olds?

Hypothesis 10: The chosen background control variables, drawn from the existing literature, are predicted to account for a significant and sizeable amount of variance in nonword reading. Speech perception predicts phonemic awareness (Elbro & Palleson, 2002; Werker & Tees, 1987), which in turn causally impacts decoding (Ehri et al., 2001). RAN is a correlate and predictor of word reading, as demonstrated in numerous studies (Araújo et al., 2015; Lervag & Hulme, 2009; Wagner et al., 1997). Nonword repetition (as an index of phonological memory) is a significant predictor of vocabulary learning (Gathercole & Baddeley, 1989) and decoding performance (Catts et al., 1999). As a precursor to phoneme-level awareness, rhyme has the potential to indirectly influence nonword reading performance (Carroll et al., 2003; Muter et al., 2004). Nonverbal ability often emerges as a significant predictor of word reading (e.g. Catts et al., 1999).

4.2.4.2 Cross-Sectional Research Question 11

How much additional contribution does vocabulary provide to nonword reading in 5-6 year olds, accounting for control variables?

4.2.4.3 Cross-Sectional Research Question 12

How much additional contribution does phonemic awareness provide to nonword reading in 5-6 year olds, accounting for the control variables and vocabulary?

Hypothesis 11 and 12: It is predicted that vocabulary and phonemic awareness will each provide a small yet significant amount of the additional variance in decoding outcomes, in accordance with several hierarchical regression analyses carried out in the intended age group (Garlock et al., 2001; Dickinson et al., 2003). Dickinson et al. (2003) alternately entered a phonological awareness composite or receptive vocabulary in the first regression step, each accounting for 7% of variance in decoding. The use of a pure phonemic awareness measure in the current study will place these outcomes under closer scrutiny. Garlock et al. (2001) entered phonemic awareness and phonological memory into the hierarchical regression equation after vocabulary. A more valid approach for the current thesis will be to control for background variables in the first step before determining the influence of vocabulary and phonemic awareness. However, since Muter et al. (2004) did not find receptive vocabulary to be a significant predictor of decoding in their path analysis with multiple variables, the outcome remains uncertain.

4.3 Chapter Summary

The current chapter specified the aims, research questions and hypotheses for the cross-sectional study. Chapter 5 will now briefly report the pilot study to trial the assessment measures and procedures to determine their sensitivity and feasibility for use in the cross-sectional study with Year One pupils. Thereafter, the method chapter for the cross-sectional research will be presented in Chapter 6.

Chapter 5: The Pilot Study

5.1 Chapter Introduction

A pilot study was conducted with 20 Year One pupils to trial the assessments prior to use in the cross-sectional and experimental investigations. The purpose of the pilot phase was: (1) to choose assessments to sensitively measure the variables under investigation in the intended age group and (2) to determine the suitability and timing of the tasks for 5-6 year old children.

5.2 Method

5.2.1 Study Design

The pilot was conducted in October 2016 to indicate how well the measures would be able to index performance at the start of the school year, in line with planned baseline testing for the intervention study. *Ethical Approval* in Appendix A was granted for the pilot and cross-sectional studies by the ethics review panel in the Department of Human Communication Sciences, University of Sheffield following the University's ethics review procedures.

5.2.2 Participants

A convenience sample of 20 pupils was recruited to receive the assessment battery, drawn from a Year One class in a school known to the researcher - an Advisory Teacher for speech, language and communication needs. The primary school is located in a rural Devon village.

The sample contained 8 girls and 12 boys with a mean age of 5;7 (range 5;2-6;1). Three pupils had Special Educational Needs at the level of School Action, and none had an Education, Health and Care Plan (EHCP). These figures were in line with the national average (DFE, 2016). Two children were in receipt of Free School Meals (FSM), slightly below the average (ibid). Pupils all spoke English as the main language; there were no bilingual or multilingual pupils in the class.

5.2.3 Sampling

The headteacher was approached regarding the recruitment of pupils to the pilot. Upon agreement, an outline of the pilot study and a consent form were sent out to parents/carers of all Year One children. Signed forms were returned for 20 pupils out of the total of 23 in the class.

Parents were not asked for a reason for denying permission. No exclusionary criteria were implemented prior to testing to support the research aim of gathering data across the full ability range in the mainstream classroom.

The Index of Multiple Deprivation (IMD; GOV.UK., 2015) provided a proxy of the socioeconomic status (SES) of the catchment area, based on school postcode. The IMD is measured in deciles representing neighbourhoods in England, ranked from the most deprived (decile 1) to the most affluent (decile 10). The IMD metric relates to seven domains - income, employment, education, health, crime, housing and barriers to services, and was chosen to provide a more rounded index of deprivation than measures focussing solely on income (Crawford & Greaves, 2013). The school is in IMD decile 8, representing a mid to high level of SES.

5.2.4 Materials

The pilot test battery set out in Table 5.1 included a range of variables characterising the relationship between vocabulary, phonemic awareness and decoding skills (see Chapter 3). A nonword spelling task was added to enhance the evidence base. The *Test Descriptions* in Appendix B provides further information on the reliability and validity of the selected assessment tools. The Peabody Picture Vocabulary Test 4 (PPVT4; Dunn & Dunn, 2007), Expressive Vocabulary Test 2 (EVT2; Williams, 2007) and Phonological Assessment Battery 2 Fluency subtest (PhAB2; Gibbs & Bodman, 2014) are not included in the appendix, since these were removed from the test battery after the pilot phase.

Standardised measures were chosen if available for the participant age group, using the newest, most recently standardised version. Apart from the vocabulary measures, tests were developed and standardised in the UK to increase validity and to minimise cultural bias. A further important criterion was for the test to be able to measure progress beyond age 6 without ceiling effects, as children would reach this age by the end of the intervention phase. Standardised scores and raw scores (number of correct items) were calculated for use in analyses. The bespoke spelling task and adapted QPR task provided raw scores only.

Table 5.1

Assessment Battery – Pilot Study

Variables	*Test*	*Age Norms*
Receptive Vocabulary	PPVT4 Peabody Picture Vocabulary Test	2;6-90
Expressive Vocabulary	EVT2 Expressive Vocabulary Test	2;6-90
Rhyme Awareness	PhAB2 Rhyme	5-6
Rhyme Production	PhAB2 Rhyme Fluency	5-11
Phonemic Awareness	PhAB2 Alliteration	5-11
Phonemic Awareness	PhAB2 Phoneme Segmentation	5-6
Nonword Reading	PhAB2 Nonword Reading	5-11
Nonword Spelling	Bespoke Nonword Spelling task	none
Speech Perception	QPR Mispronunciation Detection task	5-6
Memory Retrieval (RAN)	PhAB2 Naming Speed Digits	5-11
Nonword Repetition	PhAB2 Phonological Working Memory	5-11

Note. QPR=Quality of Phonological Representations (Claessen et al., 2009), PhAB2=Phonological Assessment Battery 2.

5.2.4.1 Vocabulary Assessments

A receptive picture pointing task was chosen to provide the optimal index of vocabulary size, since it determines whether the receptive form-meaning link has been established, arguably the minimum specification for knowing a word (Schmitt, 2010). In the standardised PPVT-4 (Dunn & Dunn, 2007), a plate of four pictures is displayed, and the child is asked to point to the item spoken by the tester. Testing starts with the set corresponding to the child's age, proceeding through sets of increasing difficulty until the set where eight or more errors are made. This forms the 'critical range' of correct items that are totalled (one point each) for the raw score.

An expressive naming task was incorporated to minimise the depth of understanding required. The standardised EVT-2 (Williams, 2007) requires the child to name a series of pictures of increasing difficulty (up to 190 items), scoring two points for each correct answer, or one point if partially correct.

The PPVT-4 and EVT-2 were replaced in the cross-sectional study for reasons to be outlined in section 5.4.1.

5.2.4.2 Phonological Awareness, RAN and Nonword Repetition Assessments

The PhAB2 (Gibbs & Bodman, 2014) is a series of tests standardised for children aged 5-11 measuring phonological awareness and other phonological processing skills. *Phonological processing* refers to the set of abilities used to analyse the sound structure of words when processing oral and written language (Wagner & Torgesen, 1987). The PhAB2 was chosen due to its child-friendly format and suitable age norms, although some ceiling effects were anticipated on the rhyme and phoneme segmentation subtests in the age range of the sample. Across tasks, a discontinuation rule applies after four consecutive errors.

A large-segment phonological awareness task (rhyme) was deemed necessary for the battery, as research suggests that different grain sizes may differentially affect vocabulary and literacy performance (Carroll, et al., 2003; Muter et al., 2004). Rhyme awareness was measured by the ten item PhAB2 Rhyme subtest, in which the tester says aloud a set of three words accompanied by pictures. The child is asked to state the two that rhyme. The PhAB2 Fluency subtest contains a measure of rhyme production as part of a composite, in which the pupil is asked to name as many rhyming words to an orally presented target as possible in 30 seconds. This subtest was discontinued after the pilot for reasons outlined in section 5.4.1.

The PhAB2 Alliteration and PhAB2 Phoneme Segmentation subtests were chosen to index phonemic awareness. Alliteration assesses the child's ability to detect whether words share the same beginning sound. The PhAB2 Alliteration subtest is presented in three parts. Parts one and two (10 items) are standardised for children aged 5;0-6;11, whereas part three (12 items) is standardised for ages 7;0-11;11 and was therefore not used. In parts one and two, children listen to three words corresponding to the presented pictures and then say the two items that share the same start sound (e.g. *r*oad, light, *r*ain).

Segmentation refers to the ability to separate a spoken word into its constituent phonemes. The PhAB2 Phoneme Segmentation subtest is based on 10 words of increasing length, containing two, three and four phonemes. A picture is presented alongside the word spoken by the assessor. The child attempts to say each of the phonemes in sequential order.

The PhAB2 also assesses RAN and nonword repetition, suggested by the literature as relevant to the relationship between vocabulary and phonic reading (Wagner & Torgesen, 1987). As discussed in Chapter 3, RAN predicts fluency of word reading (see meta-analysis by Araújo, et al., 2015). The PhAB2 Naming Speed (digits) subtest was selected to assess random automatised naming, or the fluency with which a child can access a lexical representation and generate a verbal response. If phonological, semantic and the linked lexical representations are well-specified, then access to the word should be quick and efficient (Gibbs & Bodman, 2014). A random array of 50 single digits is presented. A timer is used to measure the time taken to say the digits aloud as quickly as possible. The time is then converted to a standardised score.

The PhAB2 Phonological Working Memory subtest is a nonword repetition task of 14 items requiring the pupil to repeat nonwords of increasing syllable length, thus mirroring the task of expressive word learning (Gathercole et al., 1997). The PhAB2 manual states that the test was “designed to have minimal similarity to real words” (Gibbs & Bodman, 2014, p. 14), although measures of phonotactic probability or neighbourhood density in the manual would help to support this assertion. This task was audiotaped for use in inter-rater reliability checks.

5.2.4.3 Nonword Reading Assessment

A nonword reading task provides the optimal measure of phonic decoding skill, minimising the use of stored lexical representations to support performance (Roodenrys & Hinton, 2002). The PhAB2 Nonword Reading subtest provides a list of nonwords to assess the ability to decode phonemes and to blend these together. The test is presented in two parts. Part one includes single syllable words of increasing phonic complexity and phoneme length (3, 4 and 5 phonemes). Part two uses multisyllabic words and nonwords, some with illegal phonotactic patterns in English (e.g. mrints). The task was audiotaped to enhance scoring reliability through an inter-rater check.

5.2.4.4 Speech Perception Assessment

The Australian-normed *Quality of Phonological Representations* task (QPR; Claessen et al., 2009) was chosen to assess speech perception. A receptive mispronunciation detection task is considered a potentially more valid index of phonological representation than measures involving speech production (Antony et al., 2010; Claessen & Leitao, 2012; McNeill & Hesketh, 2010). The QPR task is based on 10 multisyllabic words (helicopter, telescope, dominoes, crocodile, television, hippopotamus, binoculars, microphone, rhinoceros, spaghetti), using six correct and four incorrect pronunciations for each word (100 items in total).

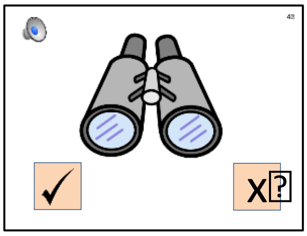
Mispronunciation detection tasks tend to have ceiling effects after the age of 4 (Stackhouse et al., 2007), however the QPR stimuli provided better scope to detect variation in this age group by including items of higher difficulty, including multisyllabic words, AoA from 4;11-7;0 (Kuperman, et al., 2012) and a combination of consonant and vowel modifications.

Two adaptations were made to the standardised QPR test for the pilot study: (1) the number of practice items was reduced and (2) the content was converted into digital format to enhance consistency of presentation and administration and to enable automatic scoring. Although an Australian test, the items were found to be mainly appropriate to the UK context, apart from one practice item (boomerang) that was not in the experience of UK pupils. The item was therefore deleted, leaving just one practice item – ambulance. Prior to testing, pupils were asked to name the images, and a score of 0 or 1 was given. If a word was not known, the tester said it aloud before the child proceeded to the test.

The computerised presentation was developed using symbols from WidgitOnline (2007) with kind permission of the publisher. Audio was recorded by a female native English speaker using a Samson sound deck microphone with noise reduction and Sennheister 280 pro-64 headphones. Wav. files were used to minimise data loss and to deliver high audio quality, and these were then edited using Audacity software. Piloting indicated that 0.5 seconds lead in time was optimal and 0.4 seconds transition after hearing the item. Recording volume was set between 0.5 and 1. The audio and video files were embedded into a Powerpoint slide for each item. The MacBook Air was used to provide optimal sound quality. The order of 100 items was randomised online using random.org (2016). The slide format is shown in Figure 5.1.

Figure 5.1

Example QPR Slide Presented on a Computer Screen



Instructions were delivered verbatim: *You will hear some words on the computer. If it sounds right, click the tick (√); if it sounds wrong, click the cross (x).* The child heard the pronunciations through high-quality headphones and then pressed the tick or cross attached to a keyboard. The next slide appeared immediately. Response time data was not collected. The child was not allowed to return to previous items or to hear the item again, and the tester was instructed to say: *That's OK, let's skip that one*.

DMDX code was written to provide automatic scoring and to reduce error. The total number of correct responses was tabulated, as well as the main data of interest, which was the number of *correct rejections*, i.e. how many times a child correctly identified a mispronounced word. Stackhouse et al. (2007) suggest that "accepting similar sounding non-words as the real word implies that the child's phonological representation of that word is fuzzy or inaccurate" (p. 74). In the study by Claessen et al. (2009), correct rejections displayed a more normal distribution than the other error types, indicating this to be a more sensitive measure to detect individual variation. There was no discontinuation rule, necessitating children to complete all items. If a pupil was not willing to finish the task (two instances), the tester was asked to draw a line under the last item completed, and the rest of the test was marked as incorrect (see record sheet in Appendix C: *Quality of Phonological Representations Task*).

5.2.4.5 Bespoke Nonword Spelling Test

Scant literature exists regarding the relationship between vocabulary and spelling in the intended age group. However, the established links between vocabulary and nonword reading suggest the possibility that vocabulary may also be related to spelling. To test this hypothesis, a graded nonword spelling task was created based on phases 2-5 of the Letters and Sounds programme (Department for Education and Skills-DFES, 2007), which forms the basis of children's phonics instruction at primary school. Where possible, nonwords were extracted from the teaching manual. If none was available for the target pattern, then a nonword was generated using a real word from the manual and changing a consonant phoneme (see Table 5.2). Nonword formation was based on the protocol used for items on the Phonics Screening Check (Standards and Testing Agency, 2012), which suggested including all letters of the alphabet and ensuring that letter strings were acceptable in the English language.

Table 5.2

Nonword Spelling Test Based on Letters and Sounds (DFES, 2007)

Phase 2	*Phase 3/5*	*Phase 4*
cag	hish	grib
reb	chee	bant
pim	zight/zite	pronk
nud	woat/wote	spunch
lis/s	yain/yane	
	gorb	

The child was provided with wide-lined paper and a pencil, and the adult dictated each nonword as soon as the child was ready (generally administered in a small group). If the child felt unable to spell it, s/he was asked to give a best try or leave it blank. No time restrictions were imposed, and the child could ask for multiple repetitions. A discontinuation rule of four consecutive errors was implemented in line with most of the other tests in the battery. Items were scored in two ways: (1) they were marked as either correct or incorrect with a score of zero or one, and (2) a point was awarded for each phoneme correct to provide a measure of emerging spelling skill.

In developing the scoring protocol (see *Nonword Spelling Task Marking Grid* in Appendix D, consideration was given to developmentally appropriate letter formation (Tremain et al., 2014) and phonically plausible alternative spellings (Standards and Testing Agency, 2012).

5.2.5 Procedure

Three testers carried out the assessments over the course of one week. The first tester was the researcher, and the two other testers were Specialist Support Assistants within the team. Each tester had an education degree and considerable experience in working with children. Prior to testing, adults spent time in the classroom to familiarise themselves with the pupils. Participants took part in three individual assessment sessions, each of around 20 minutes duration, spread over two separate days to minimise fatigue. Assessments were administered in a quiet space. Children were allowed a break at any point and could cease testing at any time, although no pupils required either adaptation. The researcher carried out session one, and the other testers administered sessions two and three. Verbatim instructions and administration procedures were practised during a training session and then kept alongside assessors during testing sessions.

5.3 Results

Given the small number of participants, it was appropriate to confine analysis to the descriptive data in Table 5.3 and correlations in Table 5.4.

5.3.1 Descriptive analysis

Table 5.3

Descriptive Statistics (Raw Scores) - Pilot Study

	N	*M*	*SD*	*Min*	*Max*
PPVT4 Receptive Vocabulary	20	101.2	18.828	50	126
EVT2 Expressive Vocabulary	17	84.29	16.567	53	114
QPR Mispronunciation Detection (CR out of 60)	19	45.75	7.511	24	56
PhAB2 Alliteration (out of 10)	20	7.45	2.012	4	10
PhAB2 Phoneme Segmentation (out of 10)	20	7.70	3.895	1	12
PhAB2 Rhyme (out of 10)	20	7.80	2.505	1	10
PhAB2 Rhyme Fluency	20	4.65	3.183	0	10
PhAB2 Nonword Reading (out of 28)	20	8.15	4.320	1	14
Nonword Spelling (out of 15)	19	7.63	2.733	4	12
PhAB2 Naming Speed Digits (standard score)	20	102.90	14.732	74	119
PhAB2 Phonological Working Memory (out of 14)	20	9.70	3.164	1	14

Note. CR=correct rejections.

Scrutiny of the minimum and maximum scores, as well as the standard deviations, indicates suitable variation in the pilot cohort performance. Skewness values between -1 and +1 and kurtosis values below 7 were considered acceptable (West et al., 1995). Inspection of histograms and skewness and kurtosis values indicated that all variables were normally distributed, apart from rhyme, which demonstrated a slight negative skew (-1.270) due to some pupils performing at ceiling level.

5.3.2 Bivariate Correlations

Significant bivariate correlations in the moderate range were found between the main variables of interest (vocabulary, phonemic awareness and decoding). Significant moderate correlations between expressive vocabulary and alliteration were also evident. A significant moderate correlation was seen between nonword repetition and expressive vocabulary but not between nonword repetition and receptive vocabulary. RAN was only weakly correlated to nonword reading in this sample.

Table 5.4

Pearson's Correlations - Pilot Study

		1	2	3	4	5	6	7	8	9	10
1	PPVT-4	-									
2	EVT-2	.649**	-								
3	QPR	-.185	.649**	-							
4	Allit	.392	.601*	.481*	-						
5	Segment	.050	.176	-.372	.092	-					
6	Rhyme	.185	.485*	-.611**	.155	.139	-				
7	Rhyme prod	.201	.557*	-.151	.412	.029	.334	-			
8	NWRdg	.497*	.619**	-.435	.718**	.156	.402	.437	-		
9	NWSpg	.575	.517	.366	.561	-.054	-.246	.426	.237	-	
10	NWRep	.409	.605*	-.204	.510*	.061	.324	.485*	.589**	.887	-
11	RAN Digits	.090	.005	.008	.358	-.358	-.046	.192	.275	.432	.127

Note. * $p \leq 0.05$; ** $p \leq 0.01$ (two -tailed). Raw scores for all variables, except standard scores for RAN Digits. QPR=Mispronunciation Detection, Allit=PhAB2 Alliteration, Segment=PhAB2 Phoneme Segmentation, Rhyme=PhAB2 Rhyme, Rhyme prod=PhAB2 Fluency, NWRdg=PhAB2 Nonword Reading, NWSpg=Nonword Spelling, NWrep=PhAB2 Phonological Working Memory, RAN Digits=PhAB2 Naming Speed Digits.

5.4 Discussion and Conclusions

5.4.1 Modifications to the Test Battery After the Pilot Study

The pilot study met its aim to trial the assessment battery to ascertain its sensitivity and suitability in the intended age group, leading to several modifications for the cross-sectional phase.

5.4.1.1 Changes to Standardised Assessments

The PPVT4 and EVT2 contained US-based vocabulary items, not in the experience of UK children, and were therefore not considered a valid measure of their knowledge. The standardised vocabulary measures were therefore replaced in the cross-sectional study by the BPVS3 and CELF4 Expressive Vocabulary test.

The PhAB2 measures provided appropriate variation in performance, apart from the rhyme task which showed a slight ceiling effect. The rhyme fluency measure, asking children to name as many items in a category as possible in 30 seconds, did not appear to be a reliable test of rhyming ability for this cohort, since its high demand on attention and motivation often led to children giving up quickly. This subtest was part of a composite fluency measure that included alliteration and semantics, rather than measuring rhyme discretely. Due to these drawbacks, it was removed from the battery. The other PhAB2 subtests were retained.

Materials for assessing nonverbal ability were not available at the time of the pilot, so these were added to the cross-sectional test battery and will be discussed in the next chapter.

5.4.1.2 Changes to Speech Perception Task

Pupils engaged well with the QPR Mispronunciation Detection task on the computer, however there were far too many items for the testing time available. Consequently, this scale was subjected to item reduction from 100 items down to 50 for the cross-sectional study. To preserve the integrity of the scale, the original 10 words were retained, using the original 3:2 ratio, resulting in 30 incorrect pronunciations and 20 correct pronunciations in the revised scale. The final list of incorrect pronunciations is shown in Appendix C.

Several criteria guided decisions regarding which of the original 60 incorrect pronunciations to retain. Firstly, items were evaluated in terms of their degree of difficulty and excluded if they did not discriminate between pupils' performance (M=1.00; 9 items). Items were mainly retained on the basis of mean scores, aiming for the middle range to provide optimal discrimination (Ebel & Frisbie, 1986). Extreme means ($M \geq .95$; $M < .20$) were excluded, although two items (M=.20; M=.95) had to be retained since no others were available to fill the slots. Secondly, the remaining 51 items (after removal of the nine ceiling items) were entered into an item analysis. The 'corrected item-total correlation' was used to decide between items with the same mean, as well as consulting the 'Chronbach's Alpha if deleted*'* to retain items that upheld a high alpha whilst discarding items that would reduce it. Selection was thirdly informed by theoretical principles affecting item difficulty, based on the type of change (consonant or vowel) and the syllable where the change occurs (discussed in Claessen et al., 2009). The final scale aimed to include a balance of these features. – The *Item Reduction of the QPR Scale* in Appendix C displays the information supporting the choices of the final 30 mispronunciations.

The fact that both the original and revised versions of the QPR task shared the same good level of reliability (α=.84) attests to the similarity of the scales.

5.4.2 Timing Considerations

Conducting three testing sessions was deemed inconvenient for teachers, requiring too many extractions from the classroom. Since the pupils were able to engage with the length of testing quite easily, the assessments were condensed into two 20-minute testing sessions for the cross-sectional study.

5.4.3 Discussion of Results

Bivariate correlations were mainly but not entirely in agreement with the reviewed literature, which suggested significant moderate correlations between receptive vocabulary, phonemic awareness and nonword reading. Receptive vocabulary was moderately correlated with nonword reading but not phonemic awareness in this sample; however, expressive vocabulary correlated with both alliteration and nonword reading. Both vocabulary measures were moderately correlated to nonword repetition as expected (Gathercole & Baddeley, 1989). Unusually, RAN digits did not significantly correlate to the reading measure (Araújo et al., 2015), however it was retained in the battery due to its documented importance in the literature. Whilst phoneme segmentation did not relate to any other measures, it was preserved as a potentially useful predictor of nonword reading and nonword spelling. The QPR task was moderately associated with rhyme and alliteration, suggestive of the role speech perception plays in the formation of phonological awareness (Elbro, 1996).

5.5 Chapter Summary

The pilot study led to a modified assessment battery and procedures, which were taken forward to the cross-sectional investigation. The method for this study will now be described in Chapter 6.

Chapter 6: Method for the Cross-Sectional Study

6.1 Chapter Introduction

The piloting phase and subsequent modifications resulted in a sensitive battery of assessments to measure vocabulary, phonological awareness, phonic literacy and several associated phonological processing skills in 5-6 year old children. The revised battery was administered to a cross-sectional sample of 152 participants to validate the relationships between vocabulary size, phonemic awareness and decoding in the current age group.

6.2 Ethical Approval

This study was granted ethical approval by the ethics review panel in the Department of Human Communication Sciences, University of Sheffield, in accordance with the University's ethics review procedures (see Appendix A).

6.3 Research Design

The cross-sectional investigation was designed to ascertain concurrent relationships at a specific timepoint. Data collection took place midway through the academic year over a four week period between February and March 2017.

6.4 Participants

6.4.1 Sample Size and Power

An a priori power analysis for multiple linear regression was conducted using G*Power 3.1 (Faul et al., 2007). The calculation was based on an alpha level of 0.5, a medium effect size commensurate with studies using similar measures and participants (Bowey & Patel, 1988; Garlock et al., 2001; Wagner et al., 1997) and the inclusion of 11 predictors in the model. A sample size of 136 pupils was calculated. To ensure adequate power after parental consent, a sample of 152 Year One pupils was recruited from five mainstream primary schools in Devon.

6.4.2 Participant Characteristics

The mean age of pupils was 6;0 years (range 5;6- 6;7). The gender ratio was well-balanced: 48% girls (*N*=73) and 52% boys (*N*=79). The Free School Meals entitlement of 21.7% (*N*=33) was higher than the national average of 14.3% (DFE, 2016). The sample was mainly of white British origin with English as their first and only language. Just 8.6% (*N*=13) of pupils were identified as EAL learners, well below the national average of 31.4% (ibid), comprising seven Polish, two Romanian, one Malayalam, one Arabic and one Italian speaker. In total, 20.4% (*N*=31) of pupils had identified special needs – higher than the 14.4% UK average. This total included 28 pupils at School Action level and a further three with an Education, Health and Care Plan – one for SLCN, one for behavioural difficulties and one for Autism. Hearing was recorded as normal for 94.7% of pupils (*N*=144), with a further eight pupils having hearing difficulties diagnosed by a medical practitioner. Participant characteristics are set out for each school in Table 6.1 below. As a measure of SES, schools tended to use Pupil Premium, which is granted if a child receives free school meals, is in the care of the local authority or has a parent in the armed forces (ibid).

Table 6.1

Participant Characteristics by School – Cross-Sectional Study

	School A	*School B*	*School C*	*School D*	*School E*
Number of Pupils	49	27	8	46	22
Mean Age	6;0 (3.8)	5;11 (3.6)	5;9 (2.0)	6;0 (3.5)	6;1 (3.3)
Girls	20	12	4	26	11
Boys	29	15	4	20	11
Pupil Premium	18	6	2	5	4
EAL	4	5	0	3	0
SEN	9	8	2	5	6

Note. Mean age=years; months (standard deviation in months); EAL=English as an Additional Language on the School-Level Annual Census; SEN=Special Educational Needs at School Action or EHCP.

6.4.3 School Characteristics

School-level information in Table 6.2 was drawn from the most recent Office for Standards in Education (OFSTED, 2015) report and the IMD website (GOV.UK., 2015).

Table 6.2

School Characteristics – Cross-Sectional Study

	Setting	*School Size*	*Number on Roll*	*IMD Decile*	*FSM*	*OFSTED Rating*
School A	urban	large	297	3	Well above	Outstanding
School B	urban	large	329	3	Above	Good
School C	rural	small	84	6	Well below	Good
School D	rural	large	411	6	below	Requires improvement
School E	suburban	large	227	6	below	Outstanding

Note. FSM=free school meals for the school; size=large, average, small; number on roll excludes nursery.

6.5 Sampling

6.5.1 Recruitment of Schools

A standard invitation letter and information sheet (see *Research Information for Schools – Cross-Sectional Study* in Appendix E) was sent to Devon infant and primary schools located within an hour's drive of the researcher's work base.

Since the literature highlights socioeconomic status as a key factor in vocabulary attainment (Hart & Risley, 1995), it was deemed important to include schools from a wide SES spectrum, even though this is not a variable for analysis in the current study. As a result, respondent schools were placed into a high- or low-SES block upon entry to the research project. Low-SES schools were defined as those with an IMD decile from 1-4, and high-SES schools were at decile 5 or above. No schools higher than decile 6 responded to the invitation. The pupils' home postcodes were collected and analysed at a later stage after parental consent was gained. Paired sample t-tests showed no significant difference between the average decile based on school versus home postcode for the high-SES category, $t(74)=1.567$, $p=.12$, nor for the low SES category, $t(74)=.129$, $p=.90$. Therefore, the school postcode was considered a valid indicator of IMD.

Responses were received from nine schools, one of which was located outside the defined geographical area and was therefore not eligible. The eight remaining candidate schools fell equally into the two SES categories. Schools were accepted until a total of at least 90 children was reached in the high- and low-SES blocks. In mixed-age classes, only the Y1 children were invited to join the study. Five schools took part, comprising 10 classes (see Table 6.3). The remaining three headteachers were thanked for their interest.

Table 6.3

Year One Classes and Pupil Numbers – Cross-Sectional Study

School	*Class*	*Number of Pupils*
School A	Class 1	17
	Class 2	18
	Class 3	17
School B	Class 4	12
	Class 5	22
School C	Class 6	8
School D	Class 7	26
	Class 8	25
School E	Class 9	14
	Class 10	13
Totals		*172*

6.5.2 Recruitment of Participants

The Head teacher at each participating school was initially contacted by telephone to provide an overview of the study and to grant permission for the researcher to liaise with the Year One teacher. Teachers sent home the *Research Information for Schools* (in Appendix E) and *Parent/Carer Consent Form - Cross-Sectional* Study (in Appendix F). The researcher explained the project to the pupils and teachers in the classroom and to parents/carers on the playground, also providing opportunities to ask questions. In summary, the Head teacher, Year One teacher, parents and children gave informed consent prior to the start of testing.

Overall, 88% of consent forms were signed and returned, granting permission for 152 pupils to take part in the study. In total, 20 parents did not return the consent form. The correspondence stated that no reason would be solicited if parents/Carers withheld permission for their child to take part in the study.

A sample of adequate power was reached (N=152). There was no attrition, as testers returned to schools if pupils were absent during the week of testing. No exclusionary criteria were implemented prior to testing, in support of the research aim to examine the full range of pupils within a mainstream classroom. There were no further exclusions after testing (see Chapter 7, section 7.2). All pupils with hearing impairment were included, since they either wore hearing aids, or the hearing loss was of an intermittent nature.

6.6 Materials

6.6.1 The Assessment Battery

The literature review in Chapter 3 highlighted the importance of certain variables in explaining the relationship between vocabulary, phonemic awareness and decoding. These measures were trialled in the pilot study, leading to the modifications discussed in Chapter 5, section 5.2.4.1. Children were assessed with the battery of tests presented in Table 6.4. Appendix B contains reliability and validity information for all tests.

Table 6.4

Assessment Battery - Cross-Sectional Study

Variables	*Test*	*Age norms*
Receptive Vocabulary	BPVS3 British Picture Vocabulary Test	3-16
Expressive Vocabulary	CELF4 Expressive Vocabulary subtest	5-9
Speech Perception	QPR Mispronunciation Detection task	5-6
Phonemic Awareness	PhAB2 Alliteration	5-11
Phonemic Awareness	PhAB2 Phoneme Segmentation	5-6
Rhyme Awareness	PhAB2 Rhyme	5-6
Nonword Reading	PhAB2 Nonword Reading	5-11
Nonword Spelling	Bespoke Nonword Spelling task	none
Memory Retrieval	PhAB2 Naming Speed Digits	5-11
Nonword Repetition	PhAB2 Phonological Working Memory	5-11
Nonverbal Ability	BAS3 Matrices (early years and primary)	3-18

Three additional tools were included in the cross-sectional study - the BPVS3 (Dunn et al., 2009), CELF4 (Semel et al., 2003) and BAS3 (Elliot & Smith, 2011).

6.6.2 Replaced Vocabulary Assessments

The BPVS3 is an assessment of receptive vocabulary widely used in research studies of school-age children. The assessor displays a plate of four coloured pictures, says the target word and asks the child to point to the corresponding picture to demonstrate understanding. The BPVS3 includes 168 items in total, broken down into 14 sets of increasing age level, with 12 items in each set.

Testing starts with the set corresponding to the child's age. The basal set, containing no more than one error, must first be established. Testing proceeds until the ceiling set is reached, i.e. the set with eight or more incorrect responses. One point is awarded for each correct response between the start of the basal set and the end of the ceiling set, so that testing only takes place over the critical range of items at an appropriate level of difficulty for the individual.

The *CELF4* is an instrument used by Speech and Language Therapists to identify language impairment. The CELF4 Expressive vocabulary subtest is a referential naming test incorporating items such as objects, actions and people. The test contains 27 items of increasing difficulty, based on age of acquisition. Pupils are shown a single picture per page, which they are asked to name. Testing starts at the child's chronological age and ceases after seven consecutive incorrect responses. Points are awarded for each item according to a scoring metric in the manual: two points for a correct answer, one point for a related response and zero points if the answer does not demonstrate understanding or if the child does not provide a response. Scores on all items are added together for a total raw score out of 54 points. The raw score is translated into a scaled score (*M*=10, *SD*=3), which was then converted by the researcher to the standardised metric of the other assessments (*M*=100; *SD*=15) to enable comparison.

6.6.3 Additional Nonverbal Skills Assessment

The *BAS3* is designed to assess general cognitive ability. The *Matrices* subtest was administered as an index of nonverbal cognition. Although no oral response is required, it is widely accepted that verbal mediation is a key process that children use in solving matrices problems (ibid). Two versions of the assessment are required to cover the intended age range. The Early Years version is standardised for use with ages 3;0-5;11, and the School Age test is for ages 6-18. If the child reaches criterion at the end of the early version, then testing progresses to the school age version. In the Matrices task, a grid of either four (preschool version) or nine (school version) squares is presented, with the bottom right square left blank for the answer. A choice of four possible solutions is presented underneath. The child points to the square that fits the pattern presented. One point is awarded for a correct response and zero for an incorrect response. The test commences at different start points according to age. Decision points are included at various stages, where the test can be discontinued if the child makes three or more mistakes. Otherwise, the test continues until the next decision point. The number of correct items in the critical range becomes the raw score, which is converted to an ability score. The ability score was then converted by the researcher to match the other standardised scales in the battery (*M*=100; *SD*=15).

6.7 Procedure

6.7.1 Ethical Considerations

The risk of harm or distress to participants was estimated to be low for the cross-sectional study. However, it was still deemed appropriate to address the unequal relationship between researchers and participants to ensure that pupils did not feel coerced into participation, and that they felt in control of how much to participate. Prior to testing, the researcher visited the classroom to discuss the project with the children and to show the general task formats. The aim of the project was conveyed as helping the researcher to try out the games, rather than about children getting the right answers. Children were informed that they could opt in or out of the tasks without repercussions. All children who had parental permission decided to complete the test battery. One pupil asked if she could bring a friend, which was duly permitted. The tester ensured that the friend had previously completed the battery, so as not to contaminate results through seeing the stimuli in advance. The decision was taken not to offer stickers or rewards for participation to avoid discrimination against that those not taking part.

Testers were professionals with a high level of experience in working in schools with young children (see section 6.7.4). Each assessor spent a short period in the classroom before the testing sessions started to become familiar with the children. For the same reason, a brief slot was allocated before each individual testing session for the assessor to have a general chat with the child. Pupils were reminded again at the start of the session that they could have a break or discontinue at any time. Whilst several took a toilet break, none refused to complete the battery. In fact, most seem to enjoy the adult attention and the short multi-sensory tasks.

Generic positive reinforcement was given throughout the testing sessions, and testers were instructed not to make any negative comments. Corrective feedback was only given for practice items.

6.7.2 Data Management and Confidentiality

Confidentiality of pupil data was ensured from the outset by assigning a participant number once consent was received. Prior to the start of testing, teachers were asked to submit a range of nominal pupil-level data for entry into the Statistical Package for Social Sciences 24 (SPSS) database linked to the anonymous participant number: (1) age, (2) gender, (3) SEN status, (4) EAL status plus the child's DFE proficiency category A-E (DFE, 2016a) and (5) whether the child was in receipt of the Pupil Premium. The participant number was attached to all test data and audio recordings, with a master list of pupil names and identification numbers held by the researcher in password-protected digital format and in a locked filing cabinet at the researcher's place of work. The anonymised test records were also stored in a locked filing cabinet. Audio recordings were held in a password protected folder on the researcher's laptop, which has the additional security features of a bitlocker code, username and password.

6.7.3 Pupil Testing Sessions

During testing, pupils were assessed individually in a quiet area of the school. Each child participated in two testing sessions, each of around 20 minutes duration, administered on separate days to minimise fatigue. Children were reminded at the beginning of every session that they could have a break or stop at any point. Both testing sessions occurred during the same week, unless the child was absent, in which case the tester returned several weeks later to complete testing.

Testing sessions were arranged to optimise performance and enjoyment. The assessment battery included short multisensory tasks to maintain interest and attention. Difficult assessments were not placed at the start or end of a session to maximise pupil motivation. The PhAB2 tasks were administered in the order suggested in the manual.

Table 6.5

Pupil Testing Sessions - Cross-Sectional Study

Session One	*Session Two*
PhAB2 Rhyme	PhAB2 Phonological Working Memory
PhAB2 Phoneme Segmentation	QPR Mispronunciation Detection task
PhAB2 Alliteration	Nonword Spelling task
PhAB2 Nonword Reading	CELF4 Expressive Vocabulary
PhAB2 Naming Speed Digits	BAS3 Matrices
BPVS3 Receptive Vocabulary	

6.7.4 Assessor Background and Training

Four testers were recruited from the Communication and Interaction Team in Devon, including two Specialist Support Assistants (testers C and D) who administered session one, and two Advisory teachers (tester A; the researcher was tester B) who delivered testing session two. All staff had a full DBS (Disclosure and Barring Service) check and worked under the direction of the main researcher. Assessors received training and practice in the delivery of all assessments, in addition to having manualised instructions next to them. The first two sessions for each tester were observed by the researcher, with feedback given.

6.7.5 Data Analysis

Assessments were marked first by the tester, then by the researcher, and then checked a third time by another member of the Communication and Interaction Team. The data was entered into SPSS software for analysis. The SPSS data sheet was double-checked by the researcher, and an administrative assistant carried out a third check.

The data was intended for quantitative analysis, including (1) descriptive presentation of participant data, (2) correlations to confirm the relationships suggested in the literature using the current assessment battery and (3) regression analyses to determine significant predictors of phonemic awareness and decoding outcomes.

6.8 Chapter Summary

A sample of adequate size was reached for data collection to analyse of the relationships between vocabulary, phonemic awareness and phonic reading. The results of the cross-sectional analyses will now be described in Chapter 7.

Chapter 7: Results of the Cross-Sectional Study

7.1 Chapter Introduction

A range of analyses were performed on the cross-sectional data to investigate the relationship between vocabulary, phonemic awareness and phonic literacy in Year One pupils. Descriptive analysis indicated the extent to which tests enabled adequate variation in performance, an important consideration for the experimental phase. Bivariate correlations and multiple regression analyses facilitated comparison of current results to the existing literature. A final hierarchical regression analysis incrementally entered the independent variables to be manipulated in the experimental study to consider their potential influence on decoding outcomes. An inter-rater reliability (IRR) check was completed for the nonword repetition and the nonword reading tasks, since verbal pupil responses necessitated tester judgment rather than having a clear distinction between correct and incorrect answers.

7.2 Descriptive Analyses

Table 7.1 provides an overview of mean performance on the standardised tests within the battery. Standardised scores were available for all measures, except the QPR task and the bespoke spelling task.

Table 7.1

Standardised Test Scores (M=100; SD=15) – Cross-Sectional Study

	N	*M*	*Min*	*Max*	*SD*
BPVS3 Receptive Vocabulary	152	92.80	69	113	11.183
CELF4 Expressive Vocabulary	152	97.08	60	135	15.094
PhAB2 Alliteration	152	94.76	69	110	12.648
PhAB2 Phoneme Segmentation	152	99.80	73	110	10.717
PhAB2 Rhyme	152	95.20	69	110	14.208
PhAB2 Nonword Reading	152	97.87	72	131	13.613
PhAB2 Naming Speed Digits	139	99.75	69	130	14.489
PhAB2 Phonological Working Memory	152	91.55	69	110	13.550
BAS3 Matrices	147	88.35	55	128	15.100

Note. CELF4 uses a different metric (*M*=10, *SD*=1) and was translated to the above scale.

Pupil performance in Table 7.1 falls within the average range on all measures. Table 7.2 presents raw scores, which will be used from this point forward in all analyses, since these typically produce more correct estimates in multivariate analyses (Cramer & Howitt, 2004; Kline, 1998).

Table 7.2

Descriptive Statistics (Raw Scores) - Cross-Sectional Study

	N	*M*	*SD*	*Min*	*Max*
BPVS3 Receptive Vocabulary (raw points score)	152	80.37	14.847	14	107
CELF4 Expressive Vocabulary (46 points)	152	21.08	8.436	2	46
QPR Mispronunciation Detection (30 items)	152	14.43	6.979	0	27
PhAB2 Alliteration (10 items)	152	6.20	3.022	0	10
PhAB2 Phoneme Segmentation (12 items)	152	8.71	3.657	0	12
PhAB2 Rhyme (10 items)	152	6.74	3.205	0	10
PhAB2 Nonword Reading (28 items)	152	8.22	5.890	0	24
Nonword Spelling (15 items)	152	6.04	3.851	0	13
PhAB2 Naming Speed Digits (time in seconds)	139	102.99	36.185	48	286
PhAB2 Phonological Working Memory (14 items)	152	10.02	3.167	0	14
BAS3 Matrices (raw points score)	147	6.03	3.016	0	17

Initial data screening was conducted to identify outliers and to check normality of distributions. Data points three standard deviations from the mean were deemed as outliers (Field, 2013). Six outlying scores were identified: PhAB2 Naming Speed (two outliers), BPVS3 (one outlier) and BAS3 (three outliers). In each case, removing the outliers did not noticeably alter the distribution or the mean, so the scores were retained in the dataset.

Normality of distributions was tested through visual inspection of histograms, Q-Q plots and analysis of skewness and kurtosis values. Using limits of -1 to +1 for skewness (Kim, 2013), histograms displayed a slight skew for PhAB2 Phoneme Segmentation (-1.073), PhAB2 Naming Speed (1.711), BPVS3 (-1.088), PhAB2 Phonological Working Memory (-1.029) and BAS3 Matrices (1.544). The remaining variables had skewness values within the stated limits. All variables had kurtosis values below 7 (ibid). Since Q-Q plots for all variables depicted normally distributed outcomes, and skewness values did not deviate considerably from the -1 to +1 limit, variables were considered appropriate to take forward to regression analysis (Field, 2013).

7.3 Bivariate Correlations

A correlational analysis was carried out to investigate the relationships between key variables for later comparison to the extant literature. Results are presented in Table 7.3.

Table 7.3

Pearson's Correlations – Cross-Sectional Study

		1	2	3	4	5	6	7	8	9	10
1	BPVS3	-									
2	CELF4	.736**	-								
3	Allit	.483**	.410**	-							
4	Segment	.420**	.346**	.451**	-						
5	Rhyme	.485**	.410**	.657**	.409**	-					
6	NWRdg	.493**	.364**	.587**	.420**	.561**	-				
7	NWSpg	.549**	.485**	.566**	.498**	.516**	.666**	-			
8	QPR	.544**	.427**	.454**	.262**	.494**	.398**	.512**	-		
9	NSDigits	-.296**	-.302**	-.374**	-.279**	-.361**	-.483**	-.515**	-.400**	-	
10	PhonWM	.404**	.321**	.339**	.209**	.283**	.369**	.344**	.293**	-.095	-
11	BAS3	.242**	.261**	.196*	.148	.271**	.280**	.276**	.213**	-.110	.132

Note. * $p \leq 0.05$; ** $p \leq 0.01$ (two -tailed). Allit=PhAB2 Alliteration, Segment=PhAB2 Phoneme Segmentation , Rhyme=Phab2 Rhyme, NWRdg=PhAB2 Nonword Reading, NWSpg=Nonword Spelling, NSDigits=PhAB2 Naming Speed Digits (seconds), PhonWM=PhAB2 Phonological Working Memory.

For the purpose of comparison to existing literature, the outcomes of interest are the highly significant moderate correlations between receptive vocabulary (BPVS3) and phonemic awareness (alliteration r=.48), as well as between receptive vocabulary and nonword reading (r=.49). Nonword reading and nonword spelling had highly significant moderate correlations with all independent variables in the model (r=.36-.59), except for the weaker relationship to BAS3 Matrices (r=.276-.280). Overall, expressive vocabulary displayed a lesser magnitude of correlation than receptive vocabulary to the other variables in the model. It is notable that both dimensions of vocabulary are more strongly correlated to nonword spelling than to nonword reading in this sample. Nonsignificant correlations were linked to the BAS3 Matrices, RAN and nonword repetition.

7.4 Multiple Linear Regression Analyses

Multiple regression analyses were carried out to verify and further investigate the predictive relationships between vocabulary, phonemic awareness and nonword reading, controlling for other relevant variables. A similar exercise was performed for the nonword spelling variable, although as previously stated, this is not the main outcome of interest in the current study.

7.4.1 The Ability of Receptive Vocabulary to Predict Phonemic Awareness

Separate multiple regression equations were created for alliteration and segmentation, since the pattern of results differed between these two variables. Predictors were entered simultaneously.

7.4.1.1 Alliteration Outcome

<u>Cross-sectional RQ 1</u>: *Does receptive vocabulary uniquely predict phonemic awareness (alliteration and phoneme segmentation) in 5-6 year olds, accounting for rhyme awareness and speech perception?*

Assumptions for multiple linear regression were assessed first. Scatterplots confirmed a linear relationship between alliteration and the collective set of independent variables (BPVS3, Phab2 Rhyme, QPR), as well as between alliteration and each variable separately. Homogeneity of error variances (homoscedasticity) was determined through visual inspection of a residuals plot (studentised residuals versus unstandardised predicted values), demonstrating an even spread of residuals. One multivariate outlier was detected for alliteration (*SD*=3.306), marginally outside the stated threshold of 3 SD. Since leverage values were all within conventional limits, i.e. below 0.2 (Huber, 1996), and Cook's distance values were below 1.0 (ibid), this data point was retained. No multicollinearity was detected between predictors, since variables displayed moderate correlations (see Table 7.3), and VIF (Variance Inflation Factor) values were well below the acceptable limit of 10 (Hair et al., 2014). Visual inspection of a histogram and P-P plot depicted normally distributed residuals. Results for alliteration performance are presented in Table 7.4.

Table 7.4

Multiple Regression Analysis - PhAB2 Alliteration Outcome (N=152)

Predictor		*Unstandardised Coefficient* B	*Standardised Coefficient* β	*p value*
(Model	R^2=.473			<.001
(Constant)		-.606		.550
BPVS3 (receptive vocabulary)		.036	.175	.020
PhAB2 Rhyme		.492	.521	<.001
QPR Mispronunciation Detection		.044	.102	.177

The model was highly significant, $F(3,148)=44.294$, $p<.001$, explaining 47% of the overall variance in alliteration performance. Significant predictors in order of their relative contributions are rhyme (β=.521) and receptive vocabulary (β=.175). Speech perception was not a unique predictor of alliteration skill, when controlling for the variance provided by the other predictors in the model.

7.4.1.2 Segmentation Outcome

Regression assumptions for the segmentation outcome were checked using the procedure above. All assumptions for linearity, homoscedasticity, multicollinearity and normally distributed residuals were met, and no outliers were detected. Results for segmentation are shown in Table 7.5.

Table 7.5

Multiple Regression Analysis - PhAB2 Phoneme Segmentation Outcome (N=152)

Predictor		*Unstandardised Coefficient* B	*Standardised Coefficient* β	*p value*
(Model)	R^2=.233			<.001
(Constant)		.781		.598
BPVS3 (receptive vocabulary)		.076	.308	<.001
PhAB2 Rhyme		.322	.282	.001
QPR Mispronunciation Detection		-.024	-.045	.616

The model was again highly significant, $F(3,148)=15.008$, $p<.001$, this time explaining 23% of the variance in segmentation performance. The significant predictors in order of their relative contributions are receptive vocabulary (β=.308) and rhyme (β=.282). Once again, the speech perception task did not make a significant contribution to segmentation outcomes.

7.4.2 The Ability of Vocabulary and Phonemic Awareness to Predict Nonword Reading

A multiple regression analysis was conducted to explore the predictors of decoding, including a set of background control variables (speech perception, rhyme, RAN, nonword repetition, nonverbal skill), vocabulary (receptive and expressive) and phonemic awareness (alliteration and segmentation). The main objective was to ascertain whether vocabulary and / or phonemic awareness would predict nonword reading, controlling for the influence of the other independent variables in the model.

Cross-sectional RQ 2: Does receptive vocabulary uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?

Cross-sectional RQ 3: Does expressive vocabulary uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?

Cross-sectional RQ 4: Does alliteration uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?

Cross-sectional RQ 5: Does segmentation uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?

Regression assumptions were investigated for the nonword reading outcome. All criteria for linearity, homogeneity of variances and normal distribution of residuals were fully met. No outliers were detected. There was no multicollinearity between variables based on VIF values, which fell well below threshold. The high correlation between BPVS3 and CELF4 (r=.74) prompted further exploration of possible collinearity. The regression model was repeated, separately entering receptive *or* expressive vocabulary. Entered alone, each vocabulary dimension was not a significant unique predictor of decoding (BPVS3 p=.09; CELF4 p=.83). The R^2 for models with the separate vocabulary entries was lower, indicating that entering *both* variables provided a better model fit to explain nonword reading performance. This outcome and the need to further explore the unique contribution of each vocabulary dimension led to both variables being retained in the model. Results are shown in Table 7.6.

Table 7.6

Multiple Regression Analysis - PhAB2 Nonword Reading Outcome (N=136)

Predictor		*Unstandardised Coefficient* B	*Standardised Coefficient* β	*p value*
(Model)	R^2=.470			<.001
(Constant)		-2.756		.422
BPVS3 (receptive vocabulary)		.115	.265	.015
CELF4 Expressive Vocabulary		-.117	-.174	.080
PhAB2 Alliteration		.514	.270	.004
PhAB2 Phoneme Segmentation		.104	.062	.399
PhAB2 Rhyme		.322	.173	.055
QPR Mispronunciation Detection		-.101	-.122	.150
PhAB2 Naming Speed Digits (RAN)		-.043	-.267	.001
PhAB2 Phonological Working Memory		.288	.150	.030
BAS3 Matrices		.202	.109	.110

The model was highly significant, accounting for 47% of the variation in nonword reading performance, $F(9,126)=12.411$, $p<.001$. The significant predictors in order of importance are alliteration (β=.270), naming speed (β=.267), receptive vocabulary (β=.265) and phonological working memory (β=.150). The main results of note are that receptive vocabulary and alliteration continued to predict nonword reading, even after all independent variables were entered into the equation. Expressive vocabulary and segmentation were not significant predictors after controlling for the other variables.

7.4.3 The Ability of Vocabulary and Phonemic Awareness to Predict Nonword Spelling

A further multiple linear regression model was conducted to predict nonword spelling outcomes based on the same independent variables. Regression assumptions were met for linearity, homogeneity of variances, normal distribution of residuals, outliers and multicollinearity. Results for nonword spelling are shown in Table 7.7.

<u>Cross-sectional RQ 6</u>: *Does receptive vocabulary uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?*

<u>Cross-sectional RQ 7</u>: Does expressive vocabulary uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

<u>Cross-sectional RQ 8</u>: Does alliteration uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

<u>Cross-sectional RQ 9</u>: Does segmentation uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

Table 7.7

Multiple Regression Analysis– Nonword Spelling Outcome (N=136)

Predictor		*Unstandardised Coefficient* B	*Standardised Coefficient* β	*p value*
(Model)	R^2=.475			<.001
(Constant)		.119		.054
BPVS3 (receptive vocabulary)		.034	.122	.258
CELF4 Expressive Vocabulary		.027	.063	.525
PhAB2 Alliteration		.243	.198	.031
PhAB2 Phoneme Segmentation		.199	.184	.013
PhAB2 Rhyme		.007	.006	.949
QPR Mispronunciation Detection		.055	.103	.218
PhAB2 Naming Speed Digits		-.026	-.252	<.001
PhAB2 Phonological Working Memory		.083	.067	.325
BAS3 Matrices		.100	.084	.215

The model was highly significant, $F(9,126)=12.652$, $p<.001$, predicting 48% of variance in the spelling performance. In order of importance, the three significant predictors are RAN (β=-.252), alliteration (β=.198) and segmentation (β=.184). Within this model, neither receptive nor expressive vocabulary was a significant predictor of nonword spelling. However, both phonemic awareness measures explained significant variation in spelling performance.

7.5 Hierarchical Regression Analysis

Having established that both receptive vocabulary and phonemic awareness are unique predictors of nonword reading, hierarchical regression was used to quantify the variance that could be explained by the incremental entry of these predictors. All multivariate assumptions were met for this analysis.

Cross-sectional RQ 10: How much contribution do control variables (speech perception, RAN, nonword repetition, rhyme and nonverbal ability) provide to nonword reading in 5-6 year olds?

Cross-sectional RQ 11: How much additional contribution does vocabulary provide to nonword reading in 5-6 year olds, accounting for control variables?

Cross-sectional RQ 12: How much additional contribution does phonemic awareness make to nonword reading in 5-6 year olds, accounting for control variables and vocabulary?

Table 7.8 reports a three-step hierarchical regression model designed to mirror the variables involved in the experimental conditions of the forthcoming intervention. Step one comprises background control variables (rhyme, RAN, nonword repetition, speech perception and nonverbal ability). The vocabulary variables were entered in the second step to determine how much additional variance could be explained by vocabulary, thus emulating the effect of the additional vocabulary learning in the semantic teaching group. In step three, the phonemic awareness variables were entered to ascertain their additional contribution to decoding outcomes, beyond control variables and vocabulary, thereby mirroring the proposed influences on the combined sound-meaning group.

Table 7.8

Hierarchical Multiple Regression Predicting Phab2 Nonword Reading (N=136)

Variable	*Step 1* Unstandardised Coefficient B	*Step 1* Standardised Coefficient β	*Step 2* Unstandardised Coefficient B	*Step 2* Standardised Coefficient β	*Step 3* Unstandardised Coefficient B	*Step 3* Standardised Coefficient β
Constant	4.865 *		-1.895		-2.756	
Control variables						
PhAB2 Rhyme	.635 **	.341 **	.574 **	.308 **	.322	.173
PhAB2 Naming Speed	-.050 **	.312 **	-.051 **	-.316 **	-.043 **	-.267 **
PhAB2 PhonWM	.372 *	.194 *	.337 *	.176 *	.288 *	.150 *
QPR Mispr Detec	-.012	-.015	-.079	-.095	-.101	-.122
BAS3 Matrices	.185	.099	.201	.108	.202	.109
Vocabulary						
BPVS3 (Rec Vocab)			.135 *	.309 *	.115 *	.265 *
CELF4 Expr Vocab			-.117	-.174	-.117	-.174
Phonemic awareness						
PhAB2 Phon Segm					.104	.062
PhAB2 Alliteration					.514*	.270 *
R^2	.388 **		.423 *		.470 *	
F	16.459 **		13.398 **		12.411 **	
ΔR^2			.035 *		.047 *	
ΔF			3.906 *		5.592 *	

Note. * $p≤.05$; **$p≤..001$; PhonWM=Phonological Working Memory; Rec Vocab=Receptive Vocabulary; Expr Vocab=Expressive Vocabulary; Phon Segm=Phoneme Segmentation; QPR Mispr Detec=QPR Mispronunciation Detection.

The full model was highly significant, $F(9,126)=12.411$, $p<.001$, explaining 47% of the variance in nonword reading outcomes. Background control variables accounted for 39% of the variance in decoding. Adding vocabulary in the second step resulted in a statistically significant increase of 3.5% variance in nonword reading, $F(2,128)=3.906$, $p=.02$. Entering phonemic awareness into the third step provided a further significant contribution of 4.7% to nonword reading, $F(2,126)=5.592$, $p=.005$.

7.6 Inter-Rater Reliability Checks

Oral pupil responses were audiorecorded for two tasks to enhance consistency through IRR checks. The PhAB2 Phonological Working Memory task was audiotaped using the internal microphone and the Quicktime app on an Apple MacBook Air laptop. An IRR check was carried out between testers A and B. Only Advisory Teachers administered this task, since it requires skill in discerning speech errors. A random 10% of the sample (*N*=16) was selected for review. None of the recordings were lost, however some of the selected files were unsuitable for reliability testing due to poor recording quality. In these cases, the recording was discarded and another one randomly selected until the full complement was reached. Pupil responses were marked on the test record during the testing session, and then the alternative tester was asked to blind mark the pupil responses from the audio recording without knowledge of the pupil's identity or the original marking.

The IRR check originating with tester A revealed 94.12% agreement, Cohen's Kappa=.784. Cohen's Kappa is a correlation that can be used between two testers rating the same content using the following magnitudes of agreement (Cohen, 1988): values up to .20=small, .21–.40=fair, .41–.60=moderate, .61–.80=substantial, and .81–1.0=almost perfect. In this case the kappa coefficient represents a substantial magnitude. The IRR check relating to the pupils of tester B displayed 83.33% accuracy, Cohen's Kappa=.619, also in the substantial category. Overall, testers assigned the same nonword repetition score with 91.1% agreement.

The PhAB2 Nonword Reading test was also audiotaped and subjected to IRR checks to monitor agreement between testers C and D, who were Specialist Support Assistants in the team. These testers used a handheld Olympus WS-853 MP3 Digital Stereo Voice Recorder to trial another alternative for audio recording. A random 10% of the sample (*N*=16) was again selected for second marking. A total of 11 audio recordings was lost from the digital dictation device, six by one tester and five by the other tester, indicating that this was a less dependable method of data collection. The reason for the data loss was tracked back to the additional step required to rename files on the digital recorder compared to the laptop. Responses were marked by the tester during the session, and the selected set of pupils was marked by the alternative tester. The anonymised digital files were played back through a HP Probook laptop using Quicktime software.

The IRR check originating with tester C revealed 84.31% agreement, Cohen's Kappa=.664. The IRR check relating to the pupils of tester D displayed 86.27% agreement, Cohen's Kappa=.646. Both kappas equate to a substantial magnitude. Overall, 85.4% agreement was reached between testers regarding the accuracy of pupils' nonword reading.

Discrepancies between the IRR outcomes were mainly attributed to children's quiet voices and extraneous background noise. Lessons learned from the IRR check informed procedures for the experimental study. Specifically, testers needed an even higher level of practice, particularly when discriminating between similar-sounding items. Improved audio recording quality, including less background noise, would also enhance analysis and inter-rater accuracy.

7.7 Chapter Summary

Results of the cross-sectional study yielded mainly moderate correlations between variables. Receptive vocabulary and phonemic awareness (alliteration) significantly predicted nonword reading in multiple regression models. Both phonemic awareness measures were significant predictors of nonword spelling, but vocabulary was not. Hierarchical regression indicated that each step in the model (control variables, vocabulary and phonemic awareness) added unique variance to nonword reading performance.

The discussion in Chapter 8 will consider the implications of these results, along with further modifications to the assessment battery.

Chapter 8: Discussion of the Cross-Sectional Study

8.1 Chapter Introduction

Results from the previous chapter allow initial conclusions to be drawn from the cross-sectional study. Chapter 8 will initially discuss findings with reference to the existing literature that suggests vocabulary as a significant predictor of phonemic awareness and phonic reading. Confirmation of these relationships in the 5-6 year old age group can provide a strong rationale for experimental testing of the causal impact of vocabulary intervention on vocabulary and distal outcomes of phonemic awareness and decoding. The second section moves on to evaluate and further improve the assessment battery in preparation for the intervention study. The final section provides an opportunity to reflect on the strengths and limitations of the cross-sectional study.

8.2 Discussion of Results

8.2.1 Predictors of Phonemic Awareness

<u>Cross-sectional RQ 1</u>: *Does receptive vocabulary uniquely predict phonemic awareness (alliteration and phoneme segmentation) in 5-6 year olds, accounting for rhyme awareness and speech perception?*

Hypothesis 1 stated that receptive vocabulary would predict phonemic awareness, based on extensive research evidence (e.g. Wagner et al., 1997). The hypothesis was upheld, since receptive vocabulary significantly predicted both alliteration (p=.02) and segmentation (p<.001) in the current sample, controlling for precursor skills of rhyme and speech perception. The use of separate phonemic awareness tasks extended the evidence base by demonstrating receptive vocabulary's contribution to *both* types of phonemic awareness skill, a question that could not be addressed through the composite measures found in previous studies.

The role of the precursor control variables in the above model should also be explored. Rhyme contributed significantly to phonemic awareness outcomes, consistent with previous research (Bryant et al., 1990; Duncan et al., 1997; Goswami, 1999; Muter et al., 2004), and this was the case for both alliteration (p<.001) and segmentation (p=.001). Although the question of whether rhyme can causally influence phonemic awareness is not clear (Bryant et al., 1990), its predictive power suggests a direction of influence that could be usefully incorporated in the sound-meaning vocabulary intervention. On the other hand, mispronunciation detection did not make a significant contribution to either phonemic awareness variable, despite literature highlighting the role of speech perception in the development of explicit phonemic awareness (Carroll et al., 2003; Elbro, 1996; Stackhouse & Wells, 1997). It is conceivable that speech perception plays a weaker role in predicting phonemic awareness than the other variables in the model, and / or its contribution may overlap with other variables in the model.

The first regression model explained 47% of the variance in alliteration performance, signifying that the chosen predictors accounted for a sizeable proportion of this skill. The same variables predicted segmentation, but the model only accounted for 23% of the variance, suggesting a weaker influence of the precursor skills on segmentation. The predictors in the model, including vocabulary, therefore appear to explain individual differences in alliteration more than segmentation.

The fact that vocabulary size predicted both alliteration and segmentation performance sits well with the hypothesised influence of lexical restructuring (Metsala & Walley, 1998), in which increases in lexical size are suggested to prompt increasingly segmental phonological representations, providing a basis for explicit phonemic awareness skills.

8.2.2 Predictors of Nonword Reading

<u>*Cross-sectional RQ 2*</u>*: Does receptive vocabulary uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?*

<u>*Cross-sectional RQ 3*</u>*: Does expressive vocabulary uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?*

<u>*Cross-sectional RQ 4*</u>*: Does alliteration uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?*

<u>Cross-sectional RQ 5</u>: *Does segmentation uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?*

Hypotheses 2-5 are based on a second multiple regression analysis examining whether four key variables drawn from the literature (receptive vocabulary, expressive vocabulary, alliteration and segmentation) predict nonword reading outcomes when simultaneously entered alongside a range of other predictors. If so, it would be useful to understand this influence prior to experimental testing of the impact of vocabulary intervention. The fact that the model was significant and explained a sizeable 47% of the variance in decoding lends credibility to a potential causal effect of these variables.

Hypothesis 2 stated that receptive vocabulary would predict performance in nonword reading. The hypothesis was upheld since receptive vocabulary emerged as a significant unique predictor of decoding (p=.15), concurring with other investigations of young pupils (Duff et al., 2015; Garlock et al., 2001; Russell et al., 2018; Verhoeven et al., 2011). Based on known predictive and causal research (Ehri et al., 2001; Muter et al., 2004), vocabulary's contribution to decoding appears to be mediated by phonemic awareness skill. Lexical restructuring provides a possible explanatory mechanism, since representations specified as separate phonemes can underpin explicit phonemic awareness skills, which in turn support literacy development (Elbro et al., 1998; Metsala, 1997; Swan & Goswami, 1997; Walley et al., 2003).

Hypothesis 3 did not state a firm prediction regarding whether expressive vocabulary would predict nonword reading due to the limited evidence base. Current cross-sectional evidence suggests that it may not contribute in the same way as receptive vocabulary. When both receptive and expressive vocabulary were entered in the current regression model, expressive vocabulary was not a significant predictor of decoding in this age group (p=.08). Nor was expressive vocabulary significant when entered into the regression model without receptive vocabulary (p=.83). The fact that a higher correlation was found between receptive vocabulary and decoding (r=.49) than between expressive vocabulary and decoding (r=.36) is a further sign of a weaker relationship. Given the high correlation between the vocabulary dimensions (*r*=.74), another reasonable account is that expressive vocabulary did not reach significance due to its overlapping contribution with receptive vocabulary.

Hypothesis 4 correctly predicted that alliteration would be a significant predictor of decoding in the model (*p*=.004), reflecting a wide evidence base (e.g. Wagner et al., 1997). This is an unsurprising result given the experimental evidence documenting the causal impact of phonemic awareness on phonic reading (Ehri et al., 2001).

Hypothesis 5 stated that segmentation would predict nonword reading outcomes, however this was not supported by the data (*p*=.40). The inclusion of separate phonemic awareness measures challenges previous results by Muter et al. (1998) that demonstrated segmentation as a significant predictor of decoding. When both phonemic awareness variables were entered in the current regression model, only alliteration uniquely predicted decoding. A conceivable explanation, suggested by Perin (1983), is that alliteration is more closely aligned with phonic reading than with segmentation.

In summary, two of the key variables under scrutiny emerged as significant predictors of decoding: receptive vocabulary and alliteration. The outcomes closely mirror the main literature (e,g, Wagner et al., 1997). Expressive vocabulary and segmentation did not uniquely contribute to decoding in this model. However, the inclusion of these variables provides valuable evidence by highlighting that these were *not* significant predictors of decoding in the Year One age group. RAN and nonword repetition made further significant contributions, however mispronunciation detection did not significantly predict decoding performance.

8.2.3 Predictors of Nonword Spelling

Cross-sectional RQ 6: Does receptive vocabulary uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

Cross-sectional RQ 7: Does expressive vocabulary uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

Cross-sectional RQ 8: Does alliteration uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

Cross-sectional RQ 9: Does segmentation uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

Hypotheses 6-9 were addressed through a third regression model to explain performance on the nonword spelling task. Once again, the chosen predictors, analogous to those for the nonword reading regression analysis, provided a good fit, accounting for 47.5% of the variance in spelling.

The scarcity of evidence limited predictions for hypotheses 6 and 7 regarding whether receptive or expressive vocabulary would explain spelling outcomes. Current regression results indicate that neither vocabulary measure uniquely explained variation in nonword spelling, a result in accordance with the limited available evidence base (Caravolas et al., 2001; Kim et al., 2013). One possibility is that vocabulary size may be a stronger driver of phonic reading than for spelling. Whilst both reading and spelling are considered to rely on the same phonological representation (Stackhouse & Wells, 1997), spelling additionally draws upon later developing phonemic awareness skills such as segmentation (ibid; Perin, 1983).

Hypotheses 8 and 9 correctly posited that alliteration and segmentation would predict significant variance in phonic spelling. This result concurs with a research synthesis by the US National Reading Panel (Ehri et al., 2001) demonstrating the causal effect of phonemic awareness instruction on spelling outcomes. The more complex skill of nonword spelling thus appears to draw upon both types of phonemic awareness skill, whereas nonword reading was only predicted by alliteration. Findings add to the current limited prediction literature for phonic spelling outcomes.

8.2.4 Relative Contributions of Background Variables, Vocabulary and Phonemic Awareness for Nonword Reading Outcomes

Cross-sectional RQ 10: How much contribution do control variables (speech perception, RAN, nonword repetition, rhyme and nonverbal ability) provide to nonword reading outcomes in 5-6 year olds?

<u>Cross-sectional RQ 11</u>: How much additional contribution does vocabulary provide to nonword reading outcomes in 5-6 year olds, accounting for control variables?

<u>Cross-sectional RQ 12</u>: How much additional contribution does phonemic awareness provide to nonword reading outcomes in 5-6 year olds, accounting for the control variables and vocabulary?

Since results highlighted receptive vocabulary and phonemic awareness as unique predictors of nonword reading, a hierarchical regression analysis was conducted to determine the additional variance provided by each variable to give an indication of their potential effect through vocabulary intervention.

Hypothesis 10 stated that control variables of speech perception, RAN, nonword repetition, rhyme and nonverbal ability would contribute to nonword reading outcomes, based on the existing literature. The model accounted for 39% of the variance in decoding performance, illustrating the influence of maturation, teaching and natural lexical restructuring on vocabulary outcomes. The use of a business-as-usual control group in the vocabulary intervention will help to account for these influences.

Hypothesis 11 was confirmed, since adding the expressive and receptive vocabulary variables in step two further contributed to nonword reading outcomes. This step was designed to emulate the effect of semantic vocabulary instruction on decoding outcomes, without explicit attention to phonological awareness. The 3.5% additional variance accords with the significant but modest contributions documented in other papers (Catts et al., 1999; Dickinson et al.; 2003; Garlock et al., 2001; Schatschneider et al., 2004), thus providing a degree of confidence that additional vocabulary could enhance decoding performance. It is suggested that lexical restructuring linked to additional vocabulary learning may be a potential driver of this process.

Hypothesis 12 anticipated that adding the phonemic awareness variables in step three would provide a further significant contribution to decoding outcomes. The hypothesis was supported, with results adding another 4.7% variance in decoding. This important finding lends weight to the suggestion that the additional explicit phonemic cues in combined vocabulary instruction could provide a supplementary boost to decoding beyond the influence of restructuring alone.

Results of this regression model give further impetus to the argument that vocabulary gains might causally support both phonemic awareness and decoding, and that enhanced and explicit phonemic awareness input would produce additional gains in these distal skills.

8.3 Modifications to the Assessment Battery After the Cross-Sectional Study

Descriptive results in Table 7.1 showed that test performance was in the average range with adequate variation in pupil outcomes, suggesting that the battery could sensitively measure performance in the Year One sample.

Regarding the BAS3, there was greater discrepancy from the standardised mean of 100 compared to the other measures. This could be linked to the fact that two test versions were needed to cover the sample age range. The Early Years version is standardised up to age 5;11, and the School Age test is for ages 6+. If the child reaches criterion at the end of the early version, then testing progresses to the school age version. Scrutiny of initial items in the school age test showed a large jump in difficulty, potentially producing a lower score for those who moved onto the higher version. However, since no other measure of nonverbal cognition was located to cover the required age span, and since performance was still broadly in the average range, the BAS3 was retained in the experimental test battery. Several tasks were nonetheless substituted to further improved the battery in preparation for testing in the intervention phase.

8.3.1 Changes to Phonemic Awareness Assessment

The CTOPP2 Elision task with its higher upper age norms was selected over PhAB2 Segmentation to alleviate the ceiling effects found during the cross-sectional study. The cross-sectional study took place in February, midway through the academic year, with some pupils already achieving top scores on the phonemic awareness subtests, so by the end of the academic year ceiling effects would be likely to increase further. Both tasks tap phonemic awareness, however at different levels of complexity. A factor analysis by Yopp (1988) confirmed elision as a more complex phonemic awareness skill than segmentation, representing an additional task demand and greater burden on working memory than simpler phoneme segmentation (ibid).

Despite ceiling effects, PhAB2 Alliteration was nonetheless retained in the battery, as it is commonly included in the research literature and therefore useful for comparison.

8.3.2 Changes to the Speech Perception Task

The QPR task did not make a significant contribution to phonemic awareness or decoding in the current model, in contrast to results of other studies (Elbro, 1996; Metsala & Walley, 1997; Werker & Tees, 1987). Presumably, after controlling for the contributions of vocabulary and rhyme, no additional influence of speech perception could be detected. The QPR task was therefore not taken forward to the experimental test battery.

8.3.3 Changes to Nonword Spelling Task

Minor adjustments were made to the nonword spelling task to better reflect the guidance on nonword test creation (DFE, 2012) to include all phonemes. To incorporate the /f/ and /v/ phonemes, the pseudoword 'liss' was changed to 'loff', and 'gorb' was adapted to 'vorb'.

8.3.4 Other Changes to the Test Battery

The PHAB2 Phonological Working Memory and PhAB2 Naming Speed (digits) tests were no longer required for testing in the intervention study, as these were not outcomes of interest. A bespoke measure of taught vocabulary was needed for the experimental phase. The development of the Taught Vocabulary Definitions probe will be described in Chapter 12.

8.4 Strengths and Contributions of the Cross-Sectional Study

Predictors for the cross-sectional study were drawn from a thorough review of the literature, thereby providing a strong test of vocabulary's ability to predict distal outcomes of phonemic awareness and phonic decoding.

The assessment battery added several novel measures to extend the evidence base. The inclusion of an expressive vocabulary test highlighted that vocabulary dimensions exerted a differential influence on phonemic awareness and decoding outcomes. Results confirm the literature that receptive vocabulary significantly predicts both phonemic awareness and decoding in the intended age group, although expressive vocabulary did not predict these outcomes. Importantly, having an expressive measure in the test battery helped to discount its contribution to decoding performance, although the shared variance between receptive and expressive vocabulary measures may be an influential factor in this outcome. The inclusion of a novel spelling measure adds to the early and emerging evidence base in this area. Current results indicate that vocabulary size predicts decoding outcomes for reading in the Year One age group but not spelling performance.

A further contribution of the current study stems from its use of separate rather than composite assessments, leading to greater specificity of results. Separating phonological awareness into discrete measures to index rhyme, alliteration and segmentation helped to fine-tune results beyond existing studies. The current homogenous assessment tools were able to delineate that alliteration was a significant predictor of decoding, however segmentation was not. Separate assessments also demonstrated that both alliteration and segmentation were significant in predicting phonic spelling skill.

8.5 Limitations of the Cross-Sectional Study

Two potential study limitations will now be explored. Firstly, a measure of letter-sound knowledge was not included in the regression model, despite its potential contribution to phonic reading outcomes. A key aim of the first investigation was to confirm that the documented relationships in the literature also applied to the Year One age group. Since this variable was not typically included in existing prediction studies, a fairer comparison was achieved by excluding it. Another reason for not indexing letter-sound knowledge is that despite the important role of phoneme-grapheme associations in initial reading, these skills are established early (Ball & Blachman, 1991) and do not consistently predict variation in phonemic awareness and word reading outcomes by age 5-6 (Byrne & Field-Barnsley, 1989; Carroll et al., 2003; Clayton et al., 2020; Lerner & Lonigan, 2016; Muter et al., 2004; Ortiz et al., 2012; Schatschneider et al., 2004). The dwindling relationship between letter-sound knowledge and decoding in the Year One cohort is seemingly due to ceiling performance (ibid) and the shared variance between letter knowledge and phonemic awareness skills (Lerner & Lonigan, 2016).

A second conceivable drawback of the cross-sectional study is linked to the representativeness of the sample. Even though all schools in the geographical area were invited to participate, the final sample fell in the low- to mid-SES range from decile 3 to 6, thus lacking schools in the higher-SES deciles. A consequence of this sampling issue is a higher number of free school meals and special educational needs than the national average. Results could therefore be construed as less representative of affluent pupil cohorts. The number of pupils learning English as an Additional Language was well below the national average, creating the possibility that results could also be less applicable to this cohort.

8.6 Chapter Summary

The cross-sectional study confirmed the relationships between vocabulary, phonemic awareness and decoding espoused in the literature. Using additional measures for expressive vocabulary and spelling added to the evidence base, alongside the more specific phonemic awareness tests rather than amalgamated measures.

The unique contribution of receptive vocabulary to phonemic awareness and phonic reading lends credibility to the idea of using vocabulary instruction to enhance these distal outcomes, in addition to the expected vocabulary gains. This result was demonstrated in the multiple regression models. They were confirmed in the hierarchical regression, which also showed that vocabulary and phonemic awareness made discrete incremental contributions to decoding outcomes.

The influence of receptive vocabulary on phonemic awareness and decoding supports predictions of the lexical restructuring hypothesis (Metsala & Walley, 1998) that increases in vocabulary size support increasingly segmental phonological representations. The data suggests that this developmental effect would be seen in all pupils but could theoretically be higher in groups receiving a boost to their vocabulary through intervention. Specific phonemic awareness input linked to combined vocabulary instruction could exert a further incremental effect on decoding outcomes (Ehri et al., 2001), potentially producing higher decoding gains than traditional semantic instruction.

The role of other significant predictors was also explored. Rhyme emerged as a significant contributor to phonemic awareness outcomes but not as a direct predictor of decoding skill. If it is possible to harness the contribution of rhyme in sound-meaning intervention, the effect would most likely be seen on phonemic awareness results.

Several further improvements were made to the test battery prior to experimental testing. The QPR task did not make a unique contribution to phonemic awareness or decoding outcomes and was therefore taken out of the test battery. A more difficult phonemic awareness measure was needed to avoid ceiling effects on PhAB2 Phoneme Segmentation, leading to the selection of the CTOPP2 Elision test.

Current cross-sectional results imply a direction of influence but not causality. Taken together, they provide a rationale for further experimental testing to manipulate vocabulary and phonemic awareness input in the vocabulary intervention described in the next section. Prior to chapters on the experimental vocabulary study, two literature reviews will be undertaken. Chapter 9 evaluates the impact of combined vocabulary instruction, while Chapter 10 considers the evidence base on effective vocabulary instruction to provide the pedagogical underpinnings for the intervention design for both teaching groups.

Chapter 9: Literature Review Three - The Impact of Combined Sound-Meaning Vocabulary Instruction

9.1 Chapter Introduction

Results of the cross-sectional study in Chapter 7 highlighted vocabulary as a significant predictor of phonemic awareness and phonic reading in line with existing research evidence (e.g. Wagner et al., 1997). Together, findings affirm a rationale to test whether these relationships can be utilised in an intervention to improve children's vocabulary and wider literacy outcomes, as suggested by several researchers (Duff et al., 2015; Munro et al., 2008). Although experimental studies have demonstrated the impact of vocabulary instruction on reading comprehension (Clarke et al., 2010; Fricke et al., 2013), there has been a dearth of enquiry on its causal effects on word-level literacy.

This review appraises the experimental literature on combined sound-meaning vocabulary intervention to understand the current state of knowledge regarding outcomes for vocabulary, phonemic awareness and phonic reading. Peer-reviewed studies comparing semantic to phonological or combined approaches were selected as relevant to the present investigation. Despite categorisation of interventions as phonological or semantic, it is not feasible in practice to work on phonological form or meaning in isolation (Janssen et al., 2018); the reviewed studies therefore are categorised as having a *main* focus on meaning, sound structure, or a combination of the two.

Existing papers fall into four main themes, the first three focussed on vocabulary outcomes and the fourth additionally measuring distal outcomes of phonemic awareness and word reading: (1) intervention studies stemming from the field of speech and language therapy to treat word-finding difficulties in clinic and specialist language support centres, (2) interventions for pupils with language disorder delivered in mainstream settings by Speech and Language Therapists (SALTs), often alongside education staff, (3) vocabulary instruction with the full range of pupils in the mainstream setting and (4) a small number of quasi-experimental studies that have supplemented the assessment battery with measures of phonemic awareness and phonic reading. This review will report the terminology used in the research studies to denote specific language needs, i.e. SLI, DLD, LD or more generic *SLCN*. More specific *Details of the Reviewed Phonological-Semantic Intervention Studies* is provided in Appendix G.

9.2 Vocabulary Outcomes for Pupils in Specialist Language Settings

A sizeable set of papers compares the effectiveness of word-finding treatments in clinical settings or specialist language units. The evidence base comprises a meta-analysis, quasi-experimental studies and case-series designs, offering information regarding the impact of these approaches for the SLCN cohort who require a specialist level of vocabulary intervention.

A meta-analysis of interventions for word-finding difficulties by Wisenburn and Mahoney (2009) determined that interventions that included phonological awareness cues improved expressive vocabulary significantly more than semantic approaches. The meta-analysis included 44 papers that divided equally into semantic, phonological and mixed therapies. Semantic therapies focussed on meaning or grammar, while phonological therapies highlighted oral word segmentation. The mixed category in this meta-analysis was quite heterogeneous in comparison, including combined phonological-semantic treatment, but also functional approaches such as using vocabulary in conversation and in role-play. The diversity of the mixed category makes comparison more difficult than if the combined and functional approaches had been presented separately.

Large average effect sizes were calculated on picture naming tasks for trained words immediately after intervention in all conditions. Phonological-based therapies had the highest magnitude of effect (d=2.43), compared to semantic cueing (d=1.41) and the mixed category (d=1.32). This scenario changes at follow-up testing 1-3 months later, when phonological approaches showed a heightened effect (d=6.87), and mixed therapy (d=2.26) now exceeded semantic-based interventions (d=1.71). One caveat is that phonological approaches were also linked to higher standard deviations, suggesting that the outcome was not consistent across all participants. Whilst results indicate that all methods were efficacious, it appears that interventions involving phonological awareness (phonological and mixed conditions) had the greatest post-intervention impact, with increasing effects over time. The focus of the meta-analysis on expressive outcomes may also have favoured treatments with a phonological emphasis.

A further noteworthy result is that whilst post-treatment gains for trained words was consistently higher across all therapies than for untrained words, semantic therapy provided the largest effect size for generalisation (d=.57) when compared to phonological (d=.37) and mixed approaches (d=.39). Input based on phonological form may thus provide superior benefits for encoding and retrieving spoken words, whilst semantic input is more likely to support connections to new words. On this basis, a pedagogy that merges both phonological and semantic information may have the greatest potential to benefit initial word learning *and* wider generalisation.

A number of small-scale investigations carried out with children in specialist language settings have similarly demonstrated better expressive vocabulary outcomes when phonological form is included in vocabulary teaching. Gray (2005) used a within-groups design with 24 participants with SLI aged 4;0-5;11 years to evaluate whether a phonological or semantic approach led to superior word learning compared to controls matched for age and gender who did not receive intervention. Oral vocabulary was taught individually by a SALT for five 30-minute sessions over the course of three weeks (total of 7.5 hours), including counterbalanced teaching of one word set in the phonological condition (syllables, rhyme and phonemic cues) and another word set using semantic cues (category, function, parts, association). Assessment of comprehension took the form of a picture pointing task, and expression was assessed through picture naming. Nonparametric analyses linked to small sample size showed that the intervention group *comprehended* significantly more words in the semantic condition (p=.007) but *produced* significantly more in the phonological condition (p=.03), prompting the author to conclude that oral phonological awareness cues may help children to create more segmental phonological representations that can enhance retrieval and spoken production. The study highlights the distinct effects that sound and meaning input might bring to receptive and expressive dimensions of vocabulary learning.

A quasi-experimental investigation by Wing (1990) also evidenced significantly better vocabulary naming in the phonological condition. The within-groups design involved 10 six year old children with SLI allocated to two intervention groups, one receiving phonological treatment and the other semantic treatment, plus a control group matched for age which did not receive intervention. The phonological input involved oral segmentation of words into syllables, rhyme and phonemes, but children were also asked to mentally visualise the image and choose it from an array. The semantic group focussed on features such as category and colour. Words were presented verbally alongside visual images for both groups. Instruction occurred three times per week for 25 minutes per session over 10 weeks (total of 12.5 hours). Results favouring the phonological treatment (reported as $p<.05$) should be interpreted with caution due to the additional visualisation cues, which could potentially have led to the higher scores. With such a small sample, a within-groups design might have afforded greater statistical power than comparing two smaller groups of five. A further methodological concern is that the Test of Word Finding (German, 1986) used to assess expressive vocabulary fell outside the standardisation ages for several pupils, potentially reducing the validity of the outcome data. Bearing these issues in mind, the study provides suggestive evidence that an approach embedding phonological and visual cues can support vocabulary learning in the language-impaired cohort.

A within-groups repeated measures design study by German et al. (2012) compared combined and semantic treatment, demonstrating significantly higher expressive vocabulary gains using a combined approach. The intervention was given to 10 pupils with SLI aged 7;11-8;9 in the lowest quartile on the Test of Word Finding (German, 2000). Pupils were taught 10 subject words in small groups by a SALT in clinic over four 30-minute sessions across two weeks using semantic cues (definitions, examples/nonexamples and sentences) and then for the following two weeks using a combined phonological-semantic approach (meaning, syllables, similar-sounding words, clear speech production), yielding a total of 2 hours in each condition. Nonparametric Wilcoxon signed rank tests were chosen due to small sample size. Significantly higher picture naming performance was seen in the combined condition than in the semantic condition (*p*=.01), however there were no significant group differences for receptive vocabulary (*p*=.33). Maintenance testing three weeks later showed a minimal decrease in expressive vocabulary from T2 to T3, although no significance values were specified.

A contradictory outcome was obtained in a between-groups design by Hyde-Wright et al. (1993), which discovered significant improvement in the semantic but not the phonological group. The sample included 15 pupils with severe SLI, spanning a wide age range from 8;1 to 14;6 years old, alongside an age-matched control group of 15 pupils not receiving the vocabulary training. Following group assignment to ensure equivalent ages in both interventions, 15 vocabulary sessions were delivered individually by a SALT over five weeks. Pupils in the semantic teaching condition (*N*=8) learned words with semantic facilitation (category, function, context, description, synonym, association) in 30-minute session. The phonological treatment (*N*=7 reduced to *N*=6 at post-test) taught vocabulary using phonological awareness cues (word length, syllables, rhyme, phonemes) in shorter sessions of 15-20 minutes. Both conditions incorporated visual images of target items. Nonparametric tests appropriate to the small sample showed significantly higher progress directly after intervention for the semantic group over controls (*p*=.01) but similar gains between the phonological group and controls (2.5 vs. 2.2 words). The authors maintain that the older age range may have favoured semantic instruction, as opposed to younger pupils who may benefit more from phonological input for vocabulary learning. The unequal intervention dosage received by the two groups (5.0 vs. 7.5 hours) introduces bias in these results.

Steele et al. (2013) also discovered that a semantic approach produced the best performance using a within-groups design with 15 pupils aged 9-11 with language impairment (SLI) compared to 11 age matched control pupils who did not learn the target words. The intervention was delivered individually by a SALT teaching five words linked to a reading text in four conditions – semantic (definitions, synonyms, categories), phonological (blending, segmentation, speech production), combined (blending, segmentation, definition) and control (one verbal presentation of the word to mirror storybook reading). The exact dosage was not specified. Only the semantic condition improved significantly more than controls (*p*=.002). The combined condition had significantly better performance (*p*=.007) than the phonological condition. A definitions probe was used to index target vocabulary learning in this study, in contrast to the expressive naming task typical of the reviewed literature for specialist settings. It is likely that the result was influenced by the vocabulary task, since definitions measure semantic meaning more than picture naming, making it difficult to compare the results to other studies. Outcomes nonetheless highlight the role of semantics in new word learning, as well as the potential benefit of combined instruction to support linked representations based on sound and meaning. A phonological focus alone was less successful at boosting word definitions. The small sample size and older cohort may also be relevant factors in the divergent outcomes.

Mixed results were discovered in a recent quasi-experimental design by Best et al. (2021). Twenty children with DLD aged 6-8 with associated word finding difficulties were randomly assigned to two groups receiving the same interventions in counter-balanced order, with a non-intervention phase between treatments. The therapist delivered weekly individual 45-minute sessions for six weeks, equating to 4.5 hours in each condition. Intervention targeted semantic attributes (category, appearance, location, use, action, association) and phonological features (syllables, rhyme, alliteration, written graphemes). No significant group difference was found on a picture naming task for target vocabulary immediately post-intervention (*p*=.18), but at maintenance testing six weeks later the semantic treatment improved significantly more than the phonological intervention with a medium effect size (*p*=.01, *d*=.49). Subsequent case series analysis identified that response to intervention differed according to individual language profiles. The study highlights how sound and meaning features differentially support pupils with language difficulties, which could conceivably apply to others in the mainstream population.

Bragard et al. (2012) also discovered a mixed picture in a case series design comparing semantic and phonological therapy with four Belgian French speakers with SLI and word-finding difficulties aged 9;6-13;9. Five 30-minute sessions (total of 2.5 hours) were delivered to individuals by a SALT, with each session consisting of 15 minutes of phonological facilitation (phoneme segmentation, link to graphemes), followed by 15 minutes of semantic input (associations, definitions). Picture naming assessments highlighted individual patterns of outcomes: two pupils benefitted most from phonological-based instruction and another from the semantic approach. One pupil did not make significant improvement in either condition, possibly owing to his better vocabulary skills within a profile of general learning difficulties. All students found phonological segmentation cues the most difficult, evidenced by the inability of any of the participants to orally segment the first phoneme of target words even after intervention. This result, mirrored in several other papers (Nash & Donaldson, 2005; Zens et al., 2009), may indicate the benefit of phonological segmentation cues to support accurate lexical representation in pupils with language difficulties, which could feasibly also pertain to younger mainstream children whose phonological awareness is still developing.

An entirely different set of results was obtained by Zens et al. (2009), who found that neither phonological nor semantic facilitation significantly improved word learning. Nineteen children with SLI aged 6;2-8;3 years received counterbalanced phonological and semantic interventions and were compared to a control group not receiving intervention matched for chronological age and gender. Individual treatment was provided twice weekly by a SALT for an hour per session over a six week period (total of 12 hours in each condition). The phonological treatment taught phonemic awareness linked to print, and the semantic intervention focussed on word attributes, association and categories. Importantly, the interventions did not focus on the target vocabulary, resulting in phonological awareness and semantic skills improving with no concomitant effect on receptive or expressive word learning, assessed by a picture pointing and picture naming task. This finding corroborates other studies showing poor generalisation to untrained vocabulary (Best, 2005; Bragard et al., 2012; German, 2002; McGregor, 1994; Wisenburn & Mahoney, 2009) and highlights the need to explicitly link sound and meaning cues to the target vocabulary.

Though not a comparative study, a relevant case series investigation by Easton et al. (1997) measured the impact of combined vocabulary instruction with four 10-11 year olds with SLI and word finding difficulties in a language unit. The pupils received twice weekly group teaching sessions by a SALT lasting half an hour over five weeks (5 hours in total). Vocabulary was taught with semantic facilitation cues (category, function, synonym) and phonological awareness cues (first sound, rhyme, syllables) in the same session. Results indicate expressive naming improvements for each pupil immediately post-intervention and at maintenance testing nine weeks later. Standardised gains were seen on a word finding test (Renfrew, 1988) for all but one pupil, whose pretest score was already at ceiling. Whilst results of a combined approach appear promising, the lack of comparison and control conditions seriously restricts interpretation of whether this approach would be superior to semantic input or to no intervention at all.

On balance, the body of evidence from specialist settings supports the view that additional phonological input during vocabulary instruction is more beneficial than semantic treatment for increasing expressive vocabulary. Opposing results by Hyde-Wright et al. (1993) should be evaluated in light of the greater dosage of semantic intervention, and Steele et al. (2013) indexed gains differently to other reviewed studies through a definition task that largely measured semantic knowledge. Investigations demonstrating efficacy in younger pupils with SLI/DLD (German et al., 2012; Gray, 2005; Wing, 1990) as well as older pupils (Bragard et al., 2012; Easton et al., 1993) suggest that phonological facilitation may be effective for a range of age groups.

Several further lessons can be drawn from this first body of literature. Gray's (2005) finding that semantic and phonological input promotes different aspects of word learning confirms the need to incorporate both types of learning cues for pupils with DLD, with potential application to younger pupils more generally. Zens et al. (2009) indicates that phonological and semantic knowledge needs to be explicitly linked to target vocabulary.

Some notable constraints are inherent in the literature on specialist settings. Enquiry was mainly focussed on expressive vocabulary outcomes for pupils with specific word finding difficulties, whereas investigations into both receptive and expressive dimensions would better support instructional planning for mainstream education. Secondly, research generally compared separate semantic and phonological approaches, although three papers contributed to investigation of the combined approach (Easton et al., 1997; German et al., 2012; Steele et al. 2013). Thirdly, results of the reviewed papers do not necessarily translate into results for the mainstream context, due to potential extra demands on the child's attention, as well as the likelihood of lower staff expertise than in a specialist setting. The next set of papers based in a mainstream context therefore adds an important dimension to evaluate the effectiveness of combined vocabulary teaching.

9.3 Vocabulary Outcomes for Pupils with Language Difficulties in the Mainstream Setting

The literature has shifted emphasis in recent years to consider vocabulary support for a wider range of pupils with SLCN in the mainstream, often as a collaboration between SALTs and education staff. The primary school evidence base includes several quasi-experimental and case-series investigations. The secondary school literature contains more robust designs, incorporating randomised control trials (RCT) as well as quasi-experimental studies. Papers typically concentrate on individual and small-group interventions, although several evaluate class-based models of delivery. Most investigations aim to study the combined vocabulary approach using a repeated measures design, in some cases compared to a control group or condition.

A classroom investigation by St John and Vance (2014) evaluated combined phonological-semantic instruction linked to curriculum vocabulary. A heterogeneous sample of 18 Year One pupils (ages 5-6) with low BPVS scores took part, including children with SLCN, EAL and other special needs. Teachers chose 10 topic words and 10 untaught control words. Guided by an Advisory SLCN Teacher, teachers delivered daily instruction to groups of six pupils for 10–15 minutes over the course of 3 to 4 weeks (approximately 4 hours in total). Taught vocabulary improved significantly more than untaught control words with a medium effect size ($p<.001$, $r=.65$). Despite methodological limitations such as small sample size, absence of a comparison intervention and unmatched word sets, the study has value in indicating the feasibility of the approach for teacher delivery within the primary classroom. The disparate composition of the sample reflects the range of mainstream pupils who may require vocabulary support. The paper additionally shares useful practical teaching formats to inform the current intervention.

A case-series study by Parsons et al. (2005) provides the first suggestion that the combined approach may also improve receptive vocabulary. Two boys (ages 8;10 and 9;5) with SLI received three individual half hour sessions over 7–8 weeks (total of 10-12 hours) from a SALT in a quiet area outside the classroom. Intervention highlighted phonological and semantic features of 18 maths words compared to a control set of untaught vocabulary. Using a pretest post-test design, significant receptive vocabulary improvements were found for both pupils on taught versus control words (p<.01; p<.001). Clearly, only preliminary evidence can be derived from this study given the small sample size and lack of a comparison condition.

Another case series investigation by Clegg (2014) offers initial evidence of the effectiveness of the combined vocabulary approach for mainstream pupils with behavioural needs. A repeated measures within-subjects design was used to examine five pupils (ages 6-8) with clinically low oral language in a primary pupil referral unit. Intervention was provided individually by a SALT in school, who taught six curriculum words over three weeks in twice weekly 30-minute sessions (total of 3 hours). A control list of six curriculum items was also developed. Both lists consisted of nouns differing in length and complexity, however as in the studies by St John and Vance (2014) and Parsons et al. (2005), a word list matched according to lexical features was not achieved. Significant post-intervention gains were made by all pupils on intervention versus control words, which were maintained four weeks later. The Parsons et al. (2005) learning resources were used in the intervention, further endorsing their usefulness across SLCN and behaviour cohorts in primary school. One potential confound is the inclusion of four sessions of phonological awareness training immediately preceding the vocabulary intervention, creating uncertainty as to whether this variable positively affected performance. Results of Zens et al. (2009) would suggest that phonological awareness teaching should not invoke transfer to target vocabulary, however this remains a question for further empirical enquiry.

Although outside the age range of the current research project, a number of investigations have been conducted with mainstream secondary pupils with SLCN, including two RCTs and several quasi-experimental studies. An RCT by Murphy et al. (2017) evaluated the Vocabulary Enrichment Programme (Joffe, 2011) for 11-13 year olds in socially disadvantaged schools. A whole class model of teaching was chosen as a result of screening that identified 37%–60% of students with vocabulary difficulties. The intervention group (N=128) received the programme delivered by their English teacher with SALT support twice a week for 12 weeks (16 hours in total). Results were compared to a waiting control group (N=75) receiving usual vocabulary instruction. Following intervention, the treated group improved significantly more than controls on standardised measures of receptive vocabulary (p=.02, η_p^2=.03) and expressive vocabulary (p=.05; η_p^2=.02), both with a small effect size. Partial eta squared (η_p^2) will be interpreted using parameters by Cohen (1988): small=.01, medium=.06 and large=.14 effect. Gains on standardised testing in this study represent an unusual result in the vocabulary intervention literature.

Confirmatory results were obtained in an RCT by Joffe et al. (2019) using the same programme with 358 pupils aged 11-12 across 21 secondary schools with language difficulties (described as below average English attainment plus performance of one SD below the mean on two or more language tests or 1.5 SD below on any one language test). This time Teaching Assistants (TA) delivered the programme under the direction of a SALT to small groups within the classroom for 6 weeks (total of 4.5-6 hours). Combined vocabulary instruction demonstrated significantly greater improvement than controls not receiving the intervention with small to medium effect size on a range of vocabulary measures, including target word definitions (p=.003, d=.37), expressive naming (p=.02, d=.27) and receptive vocabulary (p=.01, d=.27). Whilst providing an indication of effectiveness compared to usual instruction, results of both RCTs would be further strengthened through the inclusion of an alternative teaching intervention.

Several well-conducted quasi-experimental designs further extend the evidence base on the combined approach for adolescents with SLCN in the mainstream setting. A SALT-delivered intervention by Spencer et al. (2017) included 35 low-attaining teenagers aged 12-14 across three schools. Pupils were taught 10 tier two words over 10 weeks (total of 10 hours input) in small groups, resulting in significant improvement on word definitions (p<.001; η_p^2=.42) with large effect size. No significant improvement was noted for control words (p=.87) matched on relevant lexical variables. These results were confirmed by Lowe et al. (2019) using a within-groups quasi-experimental design with 78 adolescents aged 11-14 with language disorder (specified as at least one SD below the norm on at least one language measure). Pupils were taught 10 science vocabulary items through whole-class instruction by the teacher using sound-meaning cues and usual pedagogy, compared to a control list of 10 untaught words, all matched for phonological complexity, concreteness and frequency. At post-intervention testing, definitions performance in the combined condition was significantly higher than usual practice with large effect size (p=.02; η_p^2=.25) and also significantly higher than control words (p<.001). Nonparametric tests chosen due to floor effects also demonstrated significantly better expressive use of the vocabulary in a sentence for the combined condition compared to usual teaching (p<.001) and compared to control words (p<.001). At follow-up maintenance testing, the combined condition still demonstrated significantly higher expressive vocabulary use than usual teaching (p=.01) as well as control words (p<.001), however the definitions advantage of the combined over usual teaching approach was not maintained (p=.22).

Investigations in primary and secondary schools overwhelmingly concur on the advantage of the combined approach for pupils experiencing language disorder or SLCN compared to typical teaching or controls. The inclusion of an active comparative intervention in these designs would further strengthen the evidence base. The primary literature is less robust and can be enhanced through a design with greater power and experimental control, in line with the secondary literature. Evidence points to the effectiveness of mainstream models supported by an SLCN professional to provide teachers with the relevant expertise and training to deliver the combined approach (Lowe, et al., 2019; Murphy et al., 2017; St John & Vance, 2014) and in one study also for TA delivery (Joffe et al., 2019). Studies also provide practical evidence-based resources and approaches to consider for the current intervention study.

The literature reviewed in this section indicates the effectiveness and practical application of the combined approach in the mainstream school context. However, it still falls short of addressing whether superior vocabulary outcomes resulting from the combined pedagogy would apply to the range of learners in the classroom. Methodological rigour can also be enhanced through stronger experimental designs in this literature. The next section reviews the emerging mainstream evidence base.

9.4 Vocabulary Outcomes for All Pupils in the Mainstream Setting

In recent years, the combined vocabulary approach has been applied as a universal model to teach the full range of pupils in the mainstream classroom. This research represents a considerable leap from its initial application as an intervention for word-finding difficulties. It nonetheless remains an under-researched area, consisting of just a few quasi-experimental studies in nursery and primary classrooms.

In a pretest post-test within-subjects design, Moran and Moir (2018) evaluated the school-based Wordaware programme (Parsons & Branaghan, 2014) in three nursery classes located in areas of social deprivation. A sample of 91 children aged 3-5 received daily combined vocabulary instruction by their teacher of a new word each week linked to a storybook over the course of four weeks (total dosage not specified). Vocabulary definitions performance improved significantly across schools and classes ($p<.001$ in each case, no effect size reported). Descriptive results depict growth in clapping syllables, use of target vocabulary in a sentence and definitional skill at the end of intervention. The research could ideally be enhanced by the inclusion of a control group, comparison intervention, maintenance test and a primary classroom context.

Research by Silverman (2007) offers an improved design that incorporated a semantic comparison intervention and a primary setting delivered by the class teacher. Six US kindergarten classes across two schools took part, comprising 94 pupils aged 5-6 years old. A six-week intervention was implemented three times per week (dosage not specified) to teach a word a day linked to the class storybook. Within each school, three classes were randomly allocated to the three different teaching conditions. *Contextual instruction* focussed on discussion and making links to pupils' experience, *analytical instruction* concentrated on word meanings (semantic condition), and *anchored instruction* divided the time between word meanings and a phonological-orthographic programme (analogous to a combined sound-meaning approach but with additional letter cues). Target vocabulary was measured with a receptive pointing task and an expressive definitions task developed by the researcher.

Post-intervention (T2) ANOVA results revealed a significant difference in receptive target vocabulary between the three groups, $F(4,90)=5.84$, $p=.01$, with contrasts showing equal gains for the semantic and combined groups, both significantly better than the contextual group. A larger effect size was found for the combined group ($d=1.02$) than the semantic group ($d=.67$) in relation to the contextual group. A significant time x group interaction also occurred for target definitions, $F(4,90)=15.24$, $p<.001$, with contrasts showing equivalent gains between the combined and semantic groups, as well as a significant difference between taught groups and the contextual group. Large effect sizes were found for the combined group ($d=1.19$) and semantic group ($d=.85$) compared to the contextual condition.

Maintenance testing six months later (T3) with a reduced sample ($N=50$) revealed no significant interaction for receptive target vocabulary, $F(8,92)=2.19$, $p=.08$, however post hoc results reported by the researcher indicate that the combined group performed significantly better than both the semantic group ($p=.02$) and the contextual group ($p=.004$). Semantic and contextual instruction yielded similar results ($p=.84$). In terms of target definitions, a significant interaction was found, $F(8,92)=7.36$, $p<.001$, with post hoc analysis indicating that the combined group results were significantly better than contextual instruction (*p*=.01). The semantic group did not outperform contextual group (p=.11) and had similar results to combined instruction (*p*=.10). Large effect sizes were once again discovered for the combined group (*d*=.94) and semantic group (*d*=.58) compared to the contextual condition.

Overall, the Silverman (2007) study demonstrates that the combined strategy (incorporating semantics and phonological-orthographic input) conveyed an advantage over the less robust contextual intervention, as well as slightly higher effect sizes than semantic instruction. However, several questions remain unanswered. Whilst the study compared the combined and semantic conditions, a control group experiencing usual teaching was not included. This would provide valuable information regarding the true impact of the experimental interventions over and above customary practice and maturation. It is possible that the allocation of classes to different intervention groups within each school could have contaminated results through an exchange of teaching approaches, although the scripted interventions might have mitigated this effect. In this study, the additional influence of orthography (letters) on the superior combined condition result was not measured.

Access to orthographic forms, or the *orthographic facilitation effect*, promotes oral vocabulary acquisition, as demonstrated by experimental investigations comparing vocabulary outcomes with and without print (Rosenthal & Ehri, 2008; Silverman & Crandell, 2010). Minimising the effect of orthography in the current investigation will enable more specific testing of the effect of oral vocabulary on the dependent variables. A future investigation could subsequently add in the orthographic variable to demonstrate its supplementary effect.

Another quasi-experimental investigation by Damhuis et al. (2016) included a sample of 87 Dutch kindergarten children aged 4-6 compared to an age-matched control group (*N*=115) receiving usual instruction, but not a semantic comparison condition. Pupils received a teacher-delivered storybook intervention using combined vocabulary teaching plus orthography, comparable to the anchored instruction in Silverman (2007). Teaching took place for 30 minutes per day, four days per week over the course of 12 weeks (24 hours of instruction). Since receptive vocabulary levels differed at first testing, this was included as a covariate in the analysis. Immediately after intervention, target vocabulary (indexed through a receptive pointing task) showed a significant advantage for combined instruction over usual vocabulary practice with a small to medium effect size (p=.004; η_p^2=.06). In terms of standardised vocabulary measures, significantly higher progress was observed for expressive vocabulary definitions in the combined group above controls (p<.001; η_p^2=.28), with large effect size. Besides the omission of a semantic condition and the confound of orthography, the lack of documentation about the control (typical practice) condition further limits interpretation, particularly considering the dissimilar Dutch and UK educational systems. With these restrictions in mind, results hint towards the combined vocabulary approach as an effective whole-class strategy in comparison to usual instruction.

Janssen et al. (2018) carried out a quasi-experimental study comparing a phonological and semantic approach with 85 children aged 4-6, also in the Netherlands. Two schools (each with two classes) were randomly allocated to conditions. Teachers delivered storybook intervention four days per week over the course of four weeks for 35-40 minutes per day (10 hours of instruction). Schools were well matched in terms of SES and EAL variables, but the semantic condition had significantly lower standardised receptive vocabulary scores, leading to the decision to include this as a covariate in analysis. Phonological teaching cues included alliteration, the written word, detection of similar sounding words and rhyme. Semantic input included definitions, context, examples, acting out the word, drawing/craft and semantic webs. Target receptive vocabulary (a pointing task) improved equally in both groups (p=.30). Expressive definitions of taught words improved significantly more in the combined group with a large effect size (p=.001, η_p^2=.14). The combination of form and meaning once again exerted the greater effect on vocabulary overall. Despite several robust design features in this study, the lack of a follow-up maintenance test prevents evaluation of long-term impact, and a control group is still needed to compare progress to usual teaching and development.

Two intervention studies by Droop et al. (2005) investigated the combined vocabulary approach through storybooks with 5 year olds in the Netherlands. Unfortunately, only the abstract was made available in English, thus curtailing the level of detail available for review. The first study included 263 children split into intervention and control classrooms, with vocabulary instruction delivered by the teacher over the course of 16 weeks. In the second study, 442 pupils were allocated to experimental and control classrooms, each receiving instruction for 20 weeks. Both cohorts had teaching input for an hour each day with activities focussed on phonological form and meaning. On Monday, the teacher read the storybook and discussed word meanings. On Tuesday, sound-based activities were carried out with target vocabulary. Wednesday focussed on story comprehension. On Thursday, pupils took part in writing activities. Friday was devoted to evaluation activities. The researchers were interested in whether the combined sound-meaning programme provided differential outcomes for first (L1) and second (L2) language learners. L2 learners improved significantly more than controls on target receptive vocabulary, although L1 learners learned at a similar rate to controls. This raises a caveat for vocabulary investigations in relation to the equity of outcomes across different pupil cohorts. Although a control condition was incorporated in the design, the lack of a comparative semantic condition continues to constrain interpretation.

There is consensus in the existing classroom-based literature that combined instruction significantly improves target vocabulary more than either a semantic condition or control group, although no study has included both. A mainstream primary study that compares the combined approach to best practice semantic instruction, as well as to a teaching-as-usual group can provide more robust evaluation of which approach is more efficacious. Since the combined approach is now starting to be utilised with the full ability range of pupils in the classroom (e.g. Parsons & Branaghan, 2014), timely evaluation is critical to determine whether it provides better results than traditional semantic instruction.

Several studies took place with a similar age range and mainstream sample to the proposed investigation (Damhuis et al., 2016; Droop et al., 2005, Janssen et al., 2018; Silverman, 2007), providing evidence that this could be an effective vocabulary teaching option in the early primary school years. The current study seeks to extend this evidence base in terms of sample size, experimental control features and effect size reporting. Since the majority of papers stem from the Netherlands, further investigations in the UK context are also needed. A study that minimises the effect of orthography (for the purpose of experimental control) will demonstrate more clearly the influence of oral phonemic awareness facilitation on vocabulary outcomes.

9.5 Wider Outcomes of Phonemic Awareness and Phonic Reading

In addition to vocabulary outcomes, several researchers have asked whether the combined approach could also improve distal skills of phonemic awareness and word reading (Duff et al., 2015; Munro et al., 2008). It is theoretically plausible that an explicit focus on phonological segmentation of target vocabulary could help to create more segmental phonological representations (Metsala & Walley, 1998), which may in turn prepare children for explicit phonemic awareness and decoding of words for reading. This fledgling field of research incorporates a small number of relevant classroom studies in the primary age range. Several other papers have implemented the approach with specific populations, although these are outside the scope of the current review. Multi-component programmes with separate vocabulary and phonemic awareness instruction are similarly not included, since these do not necessarily reflect vocabulary's contribution to phonemic outcomes (e.g. Bowyer-Crane et al., 2008; Fricke et al., 2013; Justice et al., 2003; Lonigan, 2007; Pallante & Kim, 2013; Porta & Ramirez, 2019).

Several classroom investigations have measured phonemic awareness outcomes resulting from vocabulary instruction with a focus on phonological form. The previously reviewed paper by Janssen et al. (2018) assessed rhyme awareness and phoneme blending as part of the test battery. The form-based group improved significantly more on both skills with a medium effect size (p=.02, η_p^2=.08). Droop et al. (2005) found significantly higher post-intervention scores for the combined classrooms over controls on measures of phonological awareness and letter knowledge, although the lack of an English translation, effect size calculation and alternative intervention group impedes precise evaluation.

Preliminary evidence of the causal impact of combined vocabulary instruction on phonemic awareness can be found in a small pretest post-test design by Munro et al. (2008). A combined sound-meaning vocabulary programme, termed a 'hybrid' approach by the author, was delivered by SALTs in clinic to 17 children with SLI aged 4-6 in Australia. Six weekly sessions were delivered, each of an hour's duration (6 hours in total). A storybook context was used alongside phonological (alliteration and rhyme) and semantic (features and networks) facilitation cues for 18 high-frequency nouns. A control set of high-frequency nouns was also developed, though not matched on other lexical characteristics. Post-intervention results showed significant gains (reported as p<.05 in each case) on standardised tests of expressive naming (η_p^2=.27), rhyme awareness (η_p^2=.47) and alliteration (η_p^2=.57), each with a large effect size. Whilst the findings point towards distal phonemic awareness gains, a comparative intervention condition would strengthen the evidence considerably. It is also important to explore the long-term effectiveness of the combined approach by including a maintenance testing point. Potential application to the mainstream setting requires further enquiry, since the intervention was delivered in clinic. Assessments used in the study do not coincide with existing literature, thus limiting opportunities for comparison. The mainstream literature tends to index target vocabulary learning through a definitions task, whereas the Munro et al. study used a standardised expressive naming measure reminiscent of the earlier word-finding intervention studies. The preschool age range precluded a word reading measure in the test battery, which is vital for determining whether phonemic awareness gains exert an impact on literacy decoding. The intervention provides useful practice games to aid application and consolidation.

No peer-reviewed investigations have yet measured the effect of combined vocabulary instruction on phonic reading skill. A range of experimental studies drawing on connectionist models (e.g. Harm & Seidenberg, 2004) with typically developing children have demonstrated that learning semantic meanings alongside orthography improves word reading more than orthography alone (Dyson et al. 2017; Laing & Hulme, 1999; Nation & Cocksey, 2009; Oulette & Fraser, 2009). The relationship appears to be even stronger for exception word reading (Dyson et al., 2017; Tunmer & Chapman, 2012). Several investigations carried out with special populations have produced similar findings, i.e. for EAL learners (Vadasy & Saunders, 2016) and dyslexic pupils (Austin et al., 2021). Whilst studies confirm the supplemental effect of semantics on orthography to support word reading, they do not illuminate the impact of oral segmentation on reading. Also, the focus on exception words in the literature represents a different skill to the phonic decoding outcome of interest in the current investigation.

The literature on distal outcomes resulting from combined vocabulary instruction is still in its infancy. Indicative evidence suggests that the combined approach has the potential to facilitate phonemic awareness, although the lack of control and semantic comparison groups limits interpretation of whether this is specifically due to the additional sound-based input. Enquiry into phonic reading outcomes resulting from vocabulary instruction is at an even earlier stage, with no known papers published to date.

9.6 Semantic and Phonological Cues to Support Vocabulary Learning

The literature highlights a range of potentially useful facilitation cues to support vocabulary learning. Commonly used semantic cues include definitions, use of the word in a sentence, categorisation, visual images and actions. In terms of phonological awareness facilitation cues, investigations have highlighted phonemic segmentation and rhyme as the main aspects for instruction. Clear articulation has also been raised as a relevant phonological factor in vocabulary intervention studies (Beck & McKeown, 2007; Brackenbury & Pye, 2005; German et al., 2012; Vadasy et al., 2016), although the multi-component nature of the programmes does not allow precise impact to be determined. Two experiments that did separate out the impact of oral pronunciation both showed that clear articulation facilitated nonword vocabulary recall (Duff & Hulme, 2012; McKague et al., 2001). An experiment by Rosenthal and Ehri (2012) with 62 pupils aged 9-10 tested the effect of oral pronunciation versus silent reading on receptive vocabulary definitions and spelling. A one-week delayed maintenance test indicated no significant difference in vocabulary recall (*p* value not stated), possibly due to floor effects. However, the pronunciation group spelled significantly more target words correctly than the silent reading group ($p<.001$; $d=.95$) with a large effect size. These results signal that clear word production may be a useful learning cue to include in combined intervention.

9.7 Chapter Summary

Results of the reviewed studies suggest that the explicit sound-based cues in combined instruction have the potential to support new word learning, arguably by forming more accurate phonological representations of words that create a strong connection between word form and meaning (Metsala, 1999). The combined vocabulary approach has been adapted over time from a clinic-based model for pupils with word-finding difficulties to a prospective mainstream inclusive instructional approach. Small-group and classroom studies generally conclude that the combined approach significantly improves target vocabulary with a large effect size when compared to a semantic approach and to controls. Interest is emerging in the supplementary impact of sound-meaning instruction on phonemic awareness, with several preliminary small-group and class-level studies indicating significant improvements in the combined condition over semantic or control groups with medium to large effect size. No vocabulary intervention studies have yet been found that measure phonic reading outcomes, although researchers have posed the question.

Building upon the above evidence, Chapter 10 will now consider the principles of effective vocabulary instruction to create a strong pedagogy to serve as the basis for both vocabulary intervention groups.

Chapter 10: Literature Review Four - Principles of Effective Vocabulary Instruction

10.1 Chapter Overview

An intervention premised on strong pedagogical principles will act as the basis for further experimental enquiry to determine which vocabulary instruction approach (semantic, combined, control) delivers the best outcomes for vocabulary and other skills. Researchers in the field have identified key practices that characterise effective vocabulary programmes (Biemiller, 2003; Blachowitz et al., 2006; Graves, 2006; Hiebert & Kamil, 2005; Kameenui & Baumann, 2012; Nagy, 2005), based predominantly on semantic models. The current chapter will firstly review the impact of semantic instruction on vocabulary performance and then synthesise the main tenets of effective instruction with a view to informing the intervention design.

10.2 The Effect of Semantic Instruction on Vocabulary Outcomes

A number of reviews and meta-analyses have been published on the impact of (semantic) vocabulary training on reading comprehension (e.g. Elleman et al., 2009; Stahl & Fairbanks, 1986), however only a single meta-analysis was found relating directly to vocabulary outcomes. Marulis and Neuman (2010) examined 67 studies and 216 effect sizes in the 5-6 year old age range. A large effect size (d=.88) was calculated for interventions focussed on receptive and expressive vocabulary outcomes (shown by overlapping confidence intervals). Whilst this meta-analysis incorporates the age group and outcome of interest to the current study, the inclusion of 20 unpublished papers dictates careful interpretation of the results. However, since effect sizes were significantly higher in peer-reviewed papers than the unpublished studies (p<.05), the averaged results can be considered conservative.

Based on a subset of papers (N=11), the meta-analysis discovered no significant difference (p=.58) between vocabulary performance immediately post-intervention and after a delay. The dearth of included studies with a delayed maintenance test further confirms the need to incorporate this feature in future studies to enhance vocabulary intervention research, as suggested by Hairrell et al. (2011).

Several useful moderator analyses were performed in the meta-analysis. The largest effect sizes were linked to interventions delivered by trained adults. For example, in studies with effect sizes above 1.0 (N=79), 43% were delivered by research teams and 42% by teachers. Equivalent gains were made regardless of group size (p=.75), with large effect sizes for whole-class (g=1.04), small group (g=.88) and individual (g=.98) models of delivery. Vocabulary gains for low-SES pupils were not significantly different to an amalgamated group of mid- and high-SES pupils (p=.10). This result represents a dual-edged sword: as a group, low-SES children may be able to make equivalent vocabulary progress, but lower starting points even at this early stage of schooling do not appear to change their learning trajectory sufficiently to close the word gap.

10.3 Explicit Teaching of Word Meanings

Higher performance is observed when vocabulary meanings are explicitly taught compared to implicit or incidental word exposures. The previously discussed meta-analysis by Marulis and Neuman (2010) discovered higher effect sizes for vocabulary outcomes using explicit instructional approaches (g=1.11) or mixed explicit-implicit methods (g=1.21), compared to incidental methods (g=.62), such as defining a word when it occurs in a text. Further confirmation is provided by a meta-analysis of 21 vocabulary interventions by Fukkink and de Glopper (1998), which reported that explicit methods outperformed incidental approaches with a medium effect size (g=.43). Many subsequent investigations have likewise documented significantly better outcomes when explicit instruction is included (Apthorp et al., 2012; Beck & McKeown, 2007; Brabham & Lynch-Brown, 2002). The literature largely suggests the use of child-friendly definitions (Beck & McKeown, 2007) for explicit teaching of word meanings.

Evidence shows that younger pupils are significantly less adept than older children at deducing word meanings from text (Biemiller, 2001; Nagy et al., 1985). Incidental vocabulary opportunities are not strong enough to boost pupils with low starting vocabulary, whilst explicit instruction produces significant vocabulary gains in children with both high and low vocabulary levels compared to controls (Coyne et al., 2004; Damhuis et al., 2014; Gallagher et al., 2019; Justice et al., 2005), thus signifying an inclusive pedagogy suited to a wide range of learners.

10.4 Storybook Reading

The literature presents convincing evidence that high-quality storybooks are an effective vehicle for explicit vocabulary teaching in the primary age range. A pretest post-test design by Biemiller and Boote (2006) with 112 pupils aged 5-8 directly tested the effect of storybook reading with and without definitions. Significantly better gains with a medium effect size were observed when instruction included word definitions (p<.001, d=.53). Confirmation of this finding can be found in an investigation by Justice et al. (2005) with 57 pupils aged 5-6, showing a significant improvement with large effect size when word meanings were given (p<.001, d=1.22), compared to no significant vocabulary improvement in the implicit condition (p=.26, d=.53). Results by Elley (1989) and Penno et al. (2002) further validate this outcome.

On the other hand, variable results have arisen from investigations of implicit vocabulary exposure linked to storybooks. A meta-analysis of implicit approaches by Mol et al. (2009) reported a medium effect size for receptive vocabulary (d=.62) and expressive vocabulary (d=.45) based on 31 studies. However, a meta-analysis of 54 implicit storybook investigations by Noble et al. (2019) reports a more cautious overall effect size for receptive vocabulary outcomes (g=.19), with the suggestion that studies comparing intervention to a passive control group tend to overinflate results.

10.5 Robust Vocabulary Instruction

Considerable research points to the benefits of wider instructional activities, termed *rich* or *robust vocabulary instruction* (Beck et al., 2013). The strategy involves teachers in selecting words for instruction, introducing their meanings through a variety of contexts and engaging children in learning activities to promote word comprehension and use.

A set of quasi-experiments by Beck and McKeown (2007) investigated the effect of a broad range of rich (robust) vocabulary strategies on groups of children compared to a control group receiving usual classroom storybook reading. In the first experiment, 54 pupils aged 5-7 (kindergarten and first grade in the US) learned target vocabulary linked to storybooks using definitions, examples, sentence contexts, tally charts and classroom application opportunities throughout the week. Rich instruction produced significantly higher gains in target receptive vocabulary than usual practice with large effect sizes (kindergarten $p<.001$; $d=1.17$; first grade $p=.01$, $d=1.71$).

A second investigation by the authors (ibid) sought to determine the additive effect of a higher dosage of rich (robust) instruction over the school week, termed *more rich instruction*. Using a within-subjects design with 76 pupils aged 5-7, three words were taught using rich instruction, and three words were taught using more rich instruction. The higher dosage produced significantly greater gains with a large effect size ($p<.001$, $d=2.71$), however an important confound is introduced by the additional amount of time allocated to more rich instruction. These two studies underscore the effectiveness of rich instruction compared to business as usual, although rigour is diminished by the lack of an active comparison group.

Two experimental studies compared the effect of three strategies within the context of storybook reading: (1) incidental word exposure, (2) explicit teaching of (pupil friendly) definitions and (3) rich instruction. Coyne et al. (2009) used a within-groups design with a sample of 5-6 year olds ($N=42$) to teach three target words in each condition. Extended instruction (analogous to rich instruction) incorporated multiple contexts, questions, examples and use in a sentence. Target definitions improved significantly more with large effect size when using extended (rich) instruction ($p<.01$) compared to explicit definitions ($d=1.34$) and an even larger effect size compared to implicit encounters ($d=2.57$). Explicit definitions produced significantly higher performance than incidental encounters with a large effect size ($p<.01$; $d=.87$). A comparative investigation conducted by Brabham and Lynch-Brown (2002), this time in a larger sample of 246 participants aged 5-8, reflected the results of Coyne et al. (2009), with rich instruction producing significantly greater improvement than storybook definitions (*p*=.01) and both showing better performance than an incidental approach (*p*<.001); no effect sizes were reported in this study.

Whilst robust vocabulary instruction benefits vocabulary learning in typically developing pupils (Coyne et al., 2009; Maynard et al., 2010), including young children from low-SES backgrounds (Beck & McKeown, 2007), the results of several experimental studies suggest that it may be less effective for those with DLD, (Levlin et al., 2021; McGregor et al., 2020), potentially due to the high cognitive load and language levels involved in the approach. Levlin et al. (2021) found greater vocabulary gains for teenage pupils using a decontextualised method of *retrieval practice*, in which words or definitions are regularly recalled from memory during word acquisition, thus strengthening the lexical representation each time. This approach also has a strong evidence base of effectiveness for primary aged children (see McDermott, 2021 for a review).

A useful evidence-based model of explicit vocabulary instruction can be found in the four stages of the *STAR* approach (Blachowitz & Fisher, 2015; Parsons & Branaghan, 2014): (1) *SELECT* appropriate high utility words, (2) *TEACH* the words through structured multi-sensory instruction with pupil-friendly definitions linked to a text, (3) *APPLY* or activate the words through various contexts and activities and (4) *REVIEW* the words to consolidate and retrieve words from long-term memory.

10.6 Vocabulary Selection – Tier Two Words

Since not all words can or indeed should be taught explicitly (Biemiller, 2015), decisions must be reached regarding which words to select. Beck et al. (2002) suggest a hierarchy to support word choice based on the frequency and specificity of words. *Tier one* contains high-frequency words that occur in everyday oral language and do not typically require teaching at school age. *Tier two* medium-frequency words occur regularly in literary texts and span a wide range of curriculum subjects; the authors argue that these domain-general and often abstract words should be the focus of instruction. *Tier three* words are lower in frequency and limited to a specific subject or topic, making them potentially less useful for direct instruction. Boosting tier two vocabulary is a prudent instructional aim to influence literacy and wider school attainment. The previous section established that storybooks provide an optimal context for acquiring tier two vocabulary.

10.7 Visual Support for Vocabulary Learning

Lexical representation can be enhanced by a range of multimodal information. Dual-coding theory (Clark & Paivio, 1991) posits that visual information contained in illustrations, images, symbols and other media can augment lexical representation.

Despite the common use of visuals in vocabulary intervention studies (e.g. Coyne et al, 2004; Dickinson et al., 2019; Hadley et al., 2016; St John & Vance, 2014), scant attention has been paid to their impact. Lawson-Adams and Dickinson (2020) carried out a systematic review of 40 papers to investigate the value of visual scaffolds during vocabulary instruction. The authors concede that it is difficult to disentangle the impact of visuals from other verbally mediated components of intervention but argue that images may support vocabulary learning in two principal ways, i.e. through conveying semantic content and by directing attention to relevant aspects of word meaning. Since books do not consistently display the target vocabulary in the illustrations, and since children cannot always independently extract the relevant semantic information from pictures (Flack et al., 2018), the use of specific images and symbols may provide relevant support for word learning.

In addition to supporting semantic understanding and attention, visual images can also act as an aid to memory. Unlike oral words whose trace fades quickly in memory, picture images are enduring, thus reducing the load on working memory and assisting the encoding process. This position was validated in an experiment by Catling et al. (2021), which demonstrated that receptive recall of a vocabulary list improved significantly more with a high effect size using line drawings than with no image ($p<.001$, $\eta_p^2=.25$).

Abstract tier two words are less imageable and therefore harder to define and learn. Evidence for this can be seen in a storybook vocabulary intervention by Hadley et al. (2016) with 240 young children aged 4-5. Results indicated that concrete nouns were easiest to learn ($p<.001$ comparing nouns to each word type), and concrete verbs were significantly easier than abstract nouns and adjectives (p<.001 in each case). Visual support may thus play an important role in learning abstract vocabulary by supplementing less concrete conceptual information (Lawson-Adams & Dickinson, 2020).

Few experimental studies to date have specifically investigated the impact of visual images on vocabulary learning in a school-age sample, and no studies were found in the primary school age range. There is also minimal evidence regarding the optimal type of image to support vocabulary instruction. A preschool study by Callaghan (2000) found that 3 year olds learned vocabulary more readily with detailed images as opposed to simple line drawings. School-age pupils would possess more highly developed perceptual skills, so this outcome may be less pertinent to at primary level. Results of a preschool study by Pollard-Durodola et al. (2018) indicated significantly better performance with the use of *meaning-congruent images* that overtly depict the intended semantic meaning.

Based on the limited research available, the use of visual symbols when teaching tier two vocabulary appears a promising strategy to support new vocabulary learning.

10.8 Frequency, Dosage and Exposure

In contrast to the strength of evidence regarding explicit teaching principles, there are few clear messages on the optimal *quantity* of instruction. Experimental vocabulary studies with primary aged children do not confirm that a longer duration of intervention necessarily leads to higher gains for target vocabulary (e.g. Fricke et al., 2017). The meta-analysis by Marulis and Neuman (2010) discovered that interventions lasting up to the median duration of 42 days had no less effect than those lasting longer (p=.51). In terms of frequency, studies with fewer than the median 18 sessions had significantly higher effect sizes (p<.05) than those with 18 sessions or more, although it should be borne in mind that results for longer intervention duration and higher frequency are generally associated with pupils with more substantial language needs (ibid). The authors point out that shorter durations may be more appropriate to the targeted goals of intervention, as opposed to the longer-term educational aim of vocabulary development. The length of session did not moderate the effect of intervention, with no significant difference between sessions lasting less than the median of 20 minutes compared to longer sessions (*p*=.74). Given the short average attention span of young pupils (Cooper et al., 1978), it is predictable that longer sessions would not further enhance learning.

The number of word exposures is certainly a factor in lexical acquisition (Beck & McKeown, 2007), although experimental research does not precisely specify the optimal number of exposures. Beck et al. (2013) recommend that pupils experience 4–10 word exposures to facilitate deep understanding and use. Special populations are known to require extra encounters to facilitate oral word learning, including pupils with DLD, (Best, 2005; Dollaghan, 1987; MacGregor et al., 2021; Storkel et al., 2017) and possibly also EAL (Schmitt, 2010; Webb, 2007). The number of required encounters is likely to depend upon the lexical proficiency of the learner (Schmitt, 2010).

10.9 Active Engagement

Pupil engagement is a central factor impacting the outcomes of vocabulary instruction. Active engagement entails high levels of attention and involvement, which can result in increased retention and application of vocabulary knowledge (Laufer & Hulstijn, 2001). Nation and Webb (2011) created a checklist to assess the required level of cognitive engagement or *involvement load* of pupil activities. Examples of activities with high involvement load include higher-level questioning and interactive games (Beck & McKeown, 2007; Coyne et al., 2007). Routines may also support vocabulary learning, by enabling greater focus on the learning targets rather than peripheral procedures (Kamil, 2004). A related factor affecting engagement is the level of word knowledge required by a task (Blachowicz et al., 2006), which can fall along a continuum ranging from implicit to explicit knowledge (previously discussed in Chapter 2, section 2.5.1).

10.10 Review Cycle

Research highlights the benefits of oral retrieval practice of words and definitions to boost vocabulary learning and retention, including children of primary school age (Fazio & Marsh, 2019; Karpicke et al., 2016; McDermott, 2021).

Vocabulary interventions that incorporate a systematic review schedule can further enhance performance (Biemiller & Boote, 2006; Coyne et al., 2009; Damhuis et al., 2016), although there is some controversy in the literature over the optimum schedule. *Distributed practice* over time supports learning better than consecutive *massed practice* of the same word (Bahrick & Hall, 2005; Carpenter et al., 2012; Kang, 2016; Pavlik & Anderson, 2005). There is mixed evidence regarding the optimal time interval between review sessions, however research indicates that both expanding and fixed intervals lead to higher retention than massed schedules (Carpenter & DeLosh, 2005). Some studies favour an expanding retrieval schedule (Baddeley, 1990; Bahrick & Hall, 2005), with the added suggestion that first testing should take place quickly, when an item can still be successfully recalled (Gathercole & Baddeley, 1990; Carpenter et al., 2012). Evidence reviewed in Bahrick and Hall (2005) stresses the advantage of retrieval failures in providing an opportunity to re-evaluate one's learning strategy and to encode more correctly.

Interleaved practice involves interspersing the target word with other previously learned words. A review by Kang (2016) suggests this as an effective strategy to support long-term retention, explaining that during interleaved retrieval practice, executive function is tapped by switching between different items.

10.11 The Effect of Initial Vocabulary Size on Vocabulary Outcomes

The reviewed studies provide useful information on outcomes for pupils with high and low vocabulary levels. Many papers report at least equivalent vocabulary gains for pupils with low starting points when explicit instruction is used (Coyne et al., 2004, 2009; Damhuis et al., 2016; Elley, 1989; Justice et al., 2005; Marulis & Neuman, 2010). An exception is seen in a study by Penno et al. (2002), which evidenced less progress in the low-vocabulary group. Contradictory findings can perhaps be explained by several factors. The type of teaching input was described by the authors as an 'in flight' explanation of the target word. A more explicit approach, particularly with a child-friendly definition, would have potential for higher gains. Secondly, the assessment battery included a receptive picture pointing task, rather than the more common expressive definitions measure, creating difficulty for comparison to other studies. The definitions probe included ten words from the story and five unexposed items, therefore not a fully valid measure of target word learning.

10.12 Chapter Summary

There is wide agreement in the vocabulary instruction literature that explicit teaching of child-friendly definitions in the context of storybook reading is an effective vehicle for vocabulary learning in younger pupils. Rich vocabulary instruction offers further potential for vocabulary improvement, including an emphasis on tier two words, diverse sentence contexts, relevant examples, active engagement, multi-sensory learning and a dual focus on word comprehension and expression. Whilst the impact of frequency and dosage is unclear, there is agreement regarding the need to start instruction early and to provide multiple exposures. A short daily vocabulary input linked to the class literacy text is likely to offer optimal support, as these are a valuable source of tier two vocabulary. Systematic review of target words using principles of distributed (expanded) retrieval practice and interleaved learning can aid long-term retention. The STAR approach provides a useful direct instruction model incorporating evidence-based principles for word learning.

A vocabulary teaching programme that incorporates pedagogical principles based on research will offer a solid basis for both instructional groups, allowing experimental control to be exerted on the main variable of interest, i.e. the type of facilitation cue (semantic, combined or usual practice) to observe which group achieves the highest performance on vocabulary, phonemic awareness and phonic reading outcomes.

The current review, combined with the literature on combined instruction reviewed in Chapter 9 will inform the research questions and hypotheses for the intervention study in the Chapter 11.

Chapter 11: Aims, Research Questions and Hypotheses for The Vocabulary Intervention Study

11.1 Aims of the Experimental Intervention Study

The current study will investigate the causal impact of vocabulary instruction on outcomes of vocabulary, phonemic awareness and decoding. The overarching question for the experimental research phase is whether a whole-class vocabulary teaching programme (with and without an additional focus on phonological form) leads to improved outcomes for vocabulary, phonemic awareness and phonic reading in the Year One age group. The general aims of the experimental research phase are linked to the following identified gaps in the literature.

11.1.1 Experimental Study Aim 1

To Investigate a Sample of Mainstream Pupils

The investigation will test the combined sound-meaning approach with the full ability range of pupils to extend previous literature, which has focussed largely on language-impaired youngsters.

11.1.2 Experimental Study Aim 2

To Evaluate a Whole-Class Intervention Delivered by the Teacher

If the combined approach is being used in mainstream vocabulary teaching (e.g. Parsons & Branaghan, 2014), then it should be functionally evaluated in the classroom setting under typical conditions, using a whole-class style of delivery by the teacher. This would help to address the concern of the National Reading Panel (2000) regarding the lack of vocabulary research conducted under normal classroom circumstances.

Year One (5-6 year olds) provides the optimal cohort for testing combined sound-meaning instruction, since the cross-sectional study confirmed that vocabulary size yields its greatest influence in the early primary years when vocabulary still predicts both phonemic awareness and word reading (Wagner et al., 1997). Studies need to be conducted in the UK context, as well as in the primary age range to rival existing experimental designs in the secondary sector.

11.1.3 Experimental Study Aim 3

To Include a Comparison and Control Group

A design that includes a comparative semantic condition *and* a waiting control group is an essential next step for the evidence base, since existing designs generally include one but not both of these conditions. The control group is essential to indicate the effect of usual teaching and maturation on research outcomes. The semantic comparison group can provide information regarding the causal role of increases in vocabulary size on phonemic awareness and decoding gains, ostensibly via lexical restructuring. The combined group in turn tests the additional effect of explicit sound-based input on these outcomes. Both teaching groups should be premised on the same strong pedagogical principles to aid fair comparison.

11.1.4 Experimental Study Aim 4

To Add Homogenous Phonemic Awareness and Phonic Literacy Measures

The use of homogenous measures will provide a clearer picture of how vocabulary teaching can impact on a range of outcomes. Several additional test measures will be instrumental to extending the empirical literature. Vocabulary studies rarely include assessments of phonemic awareness and nonword reading, which are now needed to extend the emerging evidence on the impact of vocabulary instruction on distal outcomes. An index of phonic reading would provide a useful step in the evidence hierarchy for sound-meaning vocabulary instruction, as suggested by Duff et al.: "Such clear-cut evidence is not available regarding the relationship between vocabulary and reading accuracy; however, the present results provide important evidence at least for the plausibility of this causal relationship, and ought to give rise to more training studies which probe this connection" (2015: 853). Assessments for spelling and expressive vocabulary will also support gaps in the literature.

11.1.5 Experimental Study Aim 5

To Achieve Adequate Sample Size and Experimental Control Features

The existing literature would benefit from an investigation with a high level of methodological rigour, including sufficient power to allow extrapolation of results to the full range of mainstream pupils. A study with adequate participant numbers would provide convincing evidence to build upon existing small-scale designs. Robust causal evidence can be delivered by an experiment with a high level of control over confounding variables, as well as incorporating random allocation to groups, strict inclusion/exclusion criteria, blinding of testers and teachers and a standardised instructional programme.

11.1.6 Experimental Study Aim 6

<u>*To Assess Long-Term Gains Through a Follow-Up Testing Point*</u>

Many existing studies in the field of combined vocabulary instruction comprise pretest post-test designs. More investigations are therefore needed with a delayed maintenance test to ensure that long-term effects are documented (Hairrell et al., 2011).

11.1.7 Summary of Research Aims

An investigation including the above additional design features will provide an important contribution to the evidence base to support educational practice. More precise research questions and hypotheses for the experimental study can now be posed.

11.2 Research Questions and Hypotheses for the Experimental Intervention Study

Questions 1 and 2 relate to vocabulary outcomes, both for taught items and standardised vocabulary measures. Questions 3-5 ask which teaching approach will yield the highest impact on phonological awareness in terms of rhyme and varying levels of phonemic awareness skill. Questions 6 and 7 consider the novel topic of whether vocabulary teaching can also impact nonword reading and spelling.

11.2.1 Vocabulary Outcomes

11.2.1.1 Taught Vocabulary Outcomes

<u>*Experimental RQ 1*</u>*: Which teaching approach (combined, semantic, control) improves target vocabulary most in 5-6 year olds?*

Hypothesis 1: It is feasible that both teaching methods will produce similar and significantly better improvement on taught items compared to the control group. Conclusive evidence of better expressive outcomes with combined instruction exists for language-impaired populations, both in clinical and mainstream settings. Results for receptive vocabulary gains are less well-researched. Few studies have examined the impact of the combined sound-meaning approach on vocabulary learning in the mainstream pupil population. A meta-analysis of semantic-based instruction in the Year One age group suggests significantly higher gains compared to controls with medium to large effect size (Marulis & Neuman, 2010).

11.2.1.2 Standardised Receptive and Expressive Vocabulary Outcomes

<u>*Experimental RQ 2*</u>*: Which teaching approach (combined, semantic, control) improves standardised vocabulary most in 5-6 year olds?*

Hypothesis 2: No significant group differences are expected on standardised measures of receptive and expressive vocabulary, since few intervention studies have demonstrated wider gains beyond taught items.

11.2.2 Phonological Awareness Outcomes

11.2.2.1 Rhyme Outcomes

<u>*Experimental RQ 3*</u>*: Which teaching approach (combined, semantic, control) improves rhyme most in 5-6 year olds?*

Hypothesis 3: It is expected that lexical restructuring effects due to vocabulary gains will cause both experimental groups to significantly outperform controls. The combined group should improve rhyme significantly more than the semantic group given the combined teaching input on large-segment awareness, including rhyme. However, there is little experimental evidence to support this proposition. Just one preliminary study by Munro et al. (2008) demonstrated significantly better improvement on a rhyme task over controls after combined instruction.

11.2.2.2 Alliteration Outcomes

Experimental RQ 4: Which teaching approach (combined, semantic, control) improves alliteration most in 5-6 year olds?

Hypothesis 4: It is predicted that the combined group with additional sound-based cues will improve on phonemic awareness measures more than both other groups, since the intervention specifically teaches this skill. The indirect process of lexical restructuring (Metsala & Walley, 1998) could theoretically affect the intervention groups more than controls due to enhanced vocabulary learning. The control group is projected to display the least improvement, since they are not receiving the explicit phonological input nor acquiring additional target vocabulary. Phonemic awareness gains resulting from vocabulary teaching is secondary to the main outcome of enhanced vocabulary, so small gains are expected in line with regression studies showing that vocabulary predicts a small amount of the variance in phonemic awareness in the intended age group (Wagner et al., 1997). Preliminary experimental evidence points to the possibility of phonemic gains as a result of instruction that includes a focus on phonological form for pupils in the current age group (Janssen et al., 2018; Munro et al., 2008). A significant improvement in phonemic awareness, even with a small effect size, can be considered a supplementary and therefore valuable benefit.

11.2.2.3 Elision Outcomes

Experimental RQ 5: Which teaching approach (combined, semantic, control) improves elision most in 5-6 year olds?

Hypothesis 5: For reasons stated above, the predicted outcome is that the combined group will improve elision skill more than the semantic group, and both experimental groups will outperform controls. Since elision is one of the latest phonemic awareness skills to develop, ceiling effects should be minimised by using this measure, although some floor effects may be present.

11.2.3 Nonword Literacy Outcomes

11.2.3.1 Nonword Reading Outcomes

Experimental RQ 6: *Which teaching approach (combined, semantic, control) improves nonword reading most in 5-6 year olds?*

Hypothesis 6: To the researcher's knowledge, no experimental research has yet tested whether vocabulary instruction impacts phonic reading. It is anticipated that the combined group will have superior nonword reading results to both other groups due to the explicit phonemic awareness content of instruction. Both teaching groups are expected to experience significantly higher gains compared to controls, as a result of restructuring linked to increased vocabulary size. This supposition is based on two threads of evidence: cohort studies showing that vocabulary predicts a small amount of variance in word reading (e.g. Wagner et al., 1997) and experimental research demonstrating that oral phonemic awareness training, if linked to print, has a significant effect on reading and spelling with a medium effect size (Ehri et al., 2001). However, since the current approach provides no direct phoneme-grapheme teaching, the influence of the oral phonemic input on decoding outcomes is less certain.

11.2.3.2 Nonword Spelling Outcomes

Experimental RQ 7: *Which teaching approach (combined, semantic, control) improves nonword spelling most in 5-6 year olds?*

Hypothesis 7: It is not anticipated that vocabulary intervention, even with additional phonemic awareness facilitation, will produce significant improvement in nonword spelling compared to controls. Prediction studies have not identified vocabulary as a significant predictor of spelling in younger pupils (Caravolas et al., 2001; Kim et al., 2013), and no experimental studies have been published on the impact of vocabulary instruction on spelling outcomes. More explicit instruction in grapheme-phoneme correspondence may be needed to support spelling, which tends to develop after children learn to read (Stackhouse & Wells, 1997). Based on extremely limited evidence, vocabulary is unlikely to improve spelling outcomes over time. The variable has been included to begin to develop an evidence base for this area.

11.3 Chapter Summary

The current chapter outlined the aims, research questions and hypotheses for the vocabulary intervention study, which will be operationalised in Chapter 12 detailing the method for the investigation.

Chapter 12: Method of the Vocabulary Intervention Study

12.1 Chapter Introduction

A vocabulary intervention programme was designed using evidence-based principles to provide a strong underpinning pedagogy for both teaching groups. This enabled experimental manipulation of the instruction type (combined, semantic or control) to examine improvements in taught vocabulary, as well as distal phonemic awareness and phonic reading. To measure these effects, an assessment battery was developed, based largely upon the standardised assessments and researcher-developed probes trialled in the cross-sectional phase. A vocabulary teaching programme was carried out with 273 Year One pupils for 26 weeks during the 2018-19 academic year.

The current chapter describes the research design, participants and sampling process for the vocabulary intervention study. Thereafter, it will explain the materials and procedures for the assessment stage and then for the vocabulary intervention itself.

12.2 Ethical Approval

The experimental study was approved by the ethics review panel of the University of Sheffield Human Communication Sciences department (see Appendix A).

12.3 Research Design

12.3.1 Experimental Design

The current study used a randomised quasi-experimental methodology to determine the impact of vocabulary teaching with and without phonemic awareness cues on vocabulary, phonemic awareness and phonic literacy, compared to an age-matched control group. According to Slavin (2008), this type of research design offers many of the control features of an RCT, albeit with statistical analysis at the level of pupil outcomes rather than at the school-level of randomisation. The research design aimed to fill important gaps in the experimental literature (see Chapter 11, section 11.1).

12.3.2 Three Experimental Groups

Classes were randomly assigned to one of three groups to answer the research questions concerning which vocabulary teaching method made the greatest impact on vocabulary, phonemic awareness and phonic literacy outcomes.

Treatment group 1 (semantic) received meaning-based prompts to support vocabulary learning. Semantic instruction is known to be effective in boosting children's vocabulary skills (Beck & McKeown, 2007; Biemiller & Boote, 2006), however little is known about its effect on phonemic awareness or phonic literacy, since these measures have not traditionally been included in vocabulary studies.

Treatment group 2 (combined) received additional phonemic awareness facilitation cues as part of their vocabulary instruction, alongside the semantic cues as in group 1. Results for the combined instructional group are intended to indicate improvement in the outcome variables resulting from the additional phonemic awareness input during vocabulary teaching.

Treatment group 3 (waiting control) received usual vocabulary instruction prior to the intervention, which according to the teacher questionnaire (see Appendix S) consisted mainly of incidental word discussion in stories and subject lessons (equivalent to the intervention groups). Additionally, preteaching of vocabulary took place in one class. The control condition provided essential comparative information on the impact of treatment as usual (mainly incidental) vocabulary instruction and lexical development on the outcome measures.

In line with experimental protocol, the treatments were devised to be equivalent, apart from the respective type of facilitation cue (semantic, combined, control). Every effort was made to standardise the intervention activities, dosage, timings and vocabulary teaching environment to control for these influential variables.

12.3.3 Three Timepoints for Testing

The longitudinal design facilitated measurement of skill development over the period of a year. Data collection took place at three time points, each lasting approximately four weeks. In general, each child received the test battery over the course of one week.

Time 1 (T1): the *pretest* was administered over four weeks from mid-September 2018, prior to the start of the teaching intervention. The programme was then implemented from 8th October 2018 to 14th June 2019, equating to 26 weeks of intervention, excluding two weeks before Christmas due to school festivities, school holiday periods and two planned catch-up weeks for classes that may have fallen behind.

Time 2 (T2): the immediate *post-test* was administered between mid-June and mid-July 2019, directly after completion of the teaching intervention.

Time 3 (T3): the follow-up *maintenance test* took place the following academic year in November 2019 after a delay of four months once the children had moved across to Year Two. The maintenance test examines whether lasting gains were made.

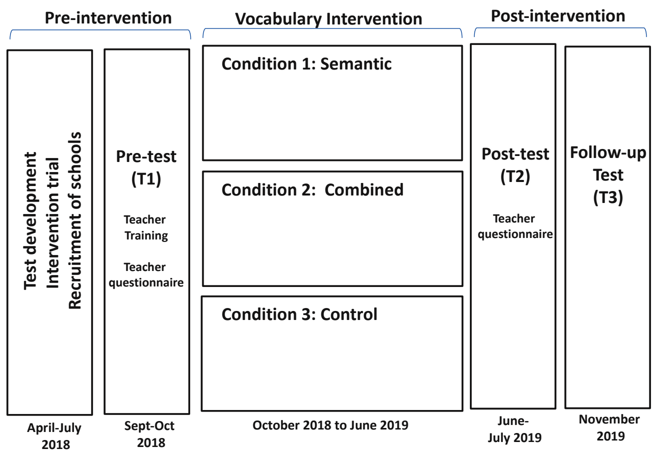
It should also be noted that during the 26 week intervention period, all groups experienced daily synthetic phonics instruction (blending phonemes to produce a word) in line with government policy (DFES, 2007). Year One additionally takes part in the national Phonics Screening Check in the summer term to assess nonword reading proficiency.

12.3.4 Overview of the Experimental Phase

The intervention groups and timescale of the vocabulary study are depicted in Figure 12.1.

Figure 12.1

Design of the Vocabulary Intervention Study



After all testing was completed, the waiting control group schools were informed that they would be offered the more effective vocabulary intervention from November 2019.

12.4 Participants

12.4.1 Sample Size and Power

An a priori power analysis was computed using G*Power 3.1 (Faul et al., 2007) yielding a required sample size of 261 pupils. The calculation was based on three testing points using repeated measures and between-factor analyses. A moderate correlation (r=.50) was assumed between the outcome variables based on previous research (e.g. Wagner et al., 1997). Effect sizes reported in the literature differed according to the dependent variable in question, i.e. vocabulary, phonemic awareness, phonic reading. A medium effect size for vocabulary outcomes was postulated by Hattie (2012), and a medium effect size for phonemic awareness by Ehri et al. (2007). No existing studies are available for phonic reading outcomes resulting from vocabulary interventions. Given the paucity of literature regarding distal outcomes, a small overall effect size was estimated (d=.20). Using an alpha of .05, the current study therefore has 95% power (two-tailed) to detect a significant difference between the three groups.

12.4.2 Participant Characteristics

12.4.2.1 Pupil Characteristics

273 Year One pupils (124 girls and 149 boys) with a mean age of 5 years 6 months (age range 5;1-6;0; *SD*=3.31) were included in the testing sample at Time 1. The vast majority of the sample (96%) were monolingual English speakers (*N*=262). The 4% of EAL learners is well below the 21% national average for primary schools in England (DFE, 2018). The EAL speakers were spread across six classes and additionally spoke Arabic (*N*=4), Polish (*N*=4) and Mayalam (*N*=2).

In total, 21.2% of pupils had identified special educational needs (*N*=58), higher than the national average of 14.6% (DFE, 2018). There were 56 pupils at School Action and two pupils with an EHCP – one issued for SLCN and the other for Cognition and Learning. Hearing was reported as normal for 96.7% of pupils (*N*=264) on the parental consent form. A further 3.3% (*N*=9) had a diagnosed hearing loss, spread across nine classes. Table 12.1 outlines the identified needs in the sample.

Table 12.1

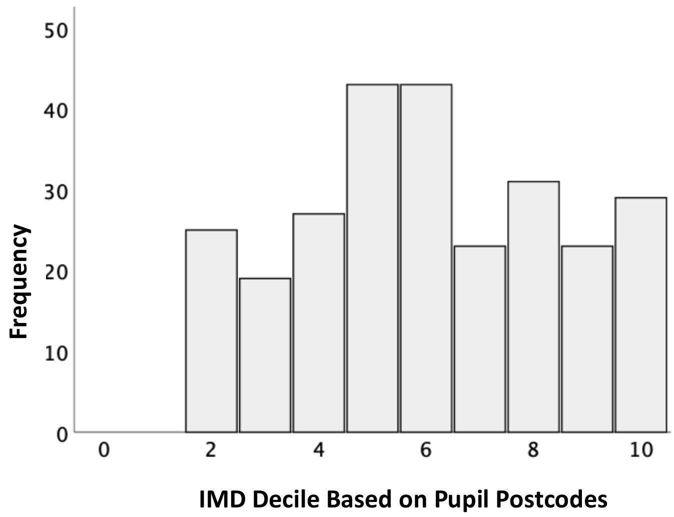
SEN Frequency and Types

Type	*Frequency*
None	215
Speech, Language and Communication Needs	36
Social, Emotional and Mental Health (SEMH)	5
Autism	3
Cognition	8
Sensory - Hearing/Visual Impairment	1
Medical	4
Other	1

In terms of SES, the Pupil Premium Grant payable to schools for low-income pupils was awarded to 13.2% of the Year One sample (*N*=36), rather lower than the national average of 19.1% (Education and Skills Funding Agency, 2018). After gaining parent/carer consent, the IMD deciles were gathered for available pupil postcodes, revealing a mean decile of 6 across the cohort (*N*=263, *SD*=2.424), slightly higher than the average decile 5 but still representing a broad spectrum of socioeconomic backgrounds ranging from decile 2 to 10 (see Figure 12.2).

Figure 12.2

Histogram of IMD Deciles Based on Pupil Postcodes



12.4.2.2 School Characteristics

Pupils attended 11 primary schools in Devon. Participating schools were located across urban (*N*=4), suburban (*N*=2) and rural (*N*=3) areas. They also varied in size, as indicated by the number on roll in Table 12.2. An adequate spread of socioeconomic levels is suggested by the IMD range based on the school postcode and the percentage of Free School Meals, which can be compared to the national average of 24.3% (GOV.UK, 2019). Schools were accepted into the research study if they received an OFSTED rating of 'good' or above, to reduce the likelihood of other school initiatives taking priority over the vocabulary project. In March 2019, halfway through the intervention, School G was reassessed by OFSTED as 'requires improvement'. Despite this status change, the class teacher completed the vocabulary intervention programme in full. Characteristics of participating schools are presented in Table 12.2.

Table 12.2

School Characteristics

	Setting	*Number of Y1 Classes*	*Number on Roll*	*School IMD Decile*	*FSM %*	*OFSTED Rating*
School A	suburban	two	412	8	7.5	good
School B	rural	one	198	10	10.1	good
School C	suburban	two	420	9	9.9	good
School D	rural	one	189	4	23.3	good
School E	urban	one	235	3	43.4	good
School F	urban	one	211	2	47.8	good
School G	urban	one	377	4	37.6	good***
School H	rural	two*	438	10	19.9	good
School I	urban	two**	335	2	38.6	good

Note. Number on roll (excluding nursery) and OFSTED rating taken from the school's most recent OFSTED report; FSM=Free School Meals information taken from *www.compare-school-performance.service.gov.uk/school* (accessed 27/4/19); IMD=Index of Multiple Deprivation, ranging from 1 (lowest) to 10 (highest)
* Equivalent to two classes; actually five smaller mixed Y1/Y2 classes
** A Year 1 class and a mixed Foundation/Y1 class
*** OFSTED rating changed to 'requires improvement' in March 2019.

12.4.2.3 Teacher Characteristics

Information regarding participating teachers was gleaned via a *Pre-intervention Teacher Questionnaire* in Appendix S. Teachers responded to questions using the Likert scale (1=low, 5=high) and open field questions. Responses for the three experimental groups are summarised in Appendix S.

Vocabulary teaching was rated as very important (scores 4-5) across the board. Confidence in vocabulary teaching displayed an equivalent range of scores: the intervention groups scored 2-4 in each case and control classes scored 3-4. Existing vocabulary approaches were quite restricted across all groups, consisting mainly of incidental discussion of new words during stories and subject lessons. Additional strategies used by the semantic group teachers included visual imagery, word banks, talk time, word of the week, acting out words – a greater range than the combined group (incidental teaching plus word walls) or the control classes (incidental teaching plus pre-teaching for some groups). In terms of years of experience, the combined group teachers had taught for longer (5-14 years) than the control class teachers (2-10 years) and considerably more than the semantic teachers (1-4 years). Teachers in all groups had experience in Year One. Some combined and control teachers had experience across all the primary age groups, whereas the semantic teachers displayed a narrower range from Year 1-4. Vocabulary training was severely limited across all groups: combined (part of a school INSET, tier 2 vocabulary), semantic (oracy training, Talk for Writing) and controls (school INSET session).

12.4.3 Sampling

12.4.3.1 Recruitment of Schools

All eligible schools in the designated area received a standard *School Invitation Letter* (in Appendix H) via the school administrator's mailbox. Since the researcher was an Advisory Teacher for SLCN in the area, it was appropriate to offer the opportunity to schools in a fair and unpressurised manner, so schools would not feel obligated to take part. As a further measure, the *Research Information for Schools – Vocabulary Intervention Study* (in Appendix I) states that participation is voluntary and does not affect school or pupil support in any way. Invitation letters were sent out to Devon Head teachers for schools that met the following eligibility criteria:

(1) School location within an hour's travel for the researcher and testers

(2) Pupil Admission Number (PAN) of at least 25.

The PAN is the maximum number of pupils admitted to each year group. A PAN of 25 suggests that the school is likely to have at least one full Year One class, rather than mixed age groups. The criterion was aimed at recruiting single-age Year One classes to support planning and implementation of the whole-class teaching programme.

12.4.3.2 High- and Low-SES Blocks

As a result of the literature confirming that vocabulary size differs significantly according to socioeconomic status (Hart & Risley, 1995; Spencer et al., 2017), schools were placed into stratified high-SES and low-SES blocks when joining the research programme. The purpose of the blocks was not to provide dichotomous group comparisons, but rather to recruit across a wide socioeconomic range and to achieve parity across experimental conditions on this influential variable. Another purpose of blocking was to reduce potential variance arising from the different levels of analysis in the research design, i.e. school assignment versus pupil analysis (Slavin, 2008).

The IMD based on the school postcode was used initially as the proxy SES measure, as per the cross-sectional study. Upon recruitment, classes from schools in the IMD range 1-4 were included in the low-SES block, and IMD range 5-10 joined the high-SES block.

12.4.3.3 School Randomisation

The design resulted in a total of six cells of around 45 participants, equating to approximately two classes of children needed for each cell.

Twelve schools responded to the invitation. Each was assigned a letter of the alphabet from A-L to signify joining order and immediately placed in the appropriate SES block. The first nine recruited classes were allocated to the intervention arm of the study. Afterwards, the next recruits were randomly allocated to the control group slots. Ideally in an RCT design, there would be random allocation to both the intervention and control groups, however, the Head teachers who responded first were not prepared to wait until the next academic year for the teaching intervention, despite this being explained in the invitation letter.

Using cluster randomisation, the classes in the intervention arm were allocated to the intervention (semantic and combined) cells. Graphpad (2018) was used to shuffle the order of classes within the SES block, which were then allocated sequentially to slots 1-8 as shown in Table 12.3. When a class from a larger school was randomly allocated to a cell, then the partner class was also placed in the same cell to avoid exposure to the alternate teaching strategy. The randomisation of intervention classes resulted in nine classes taking part in the intervention groups.

Table 12.3

Randomisation of Schools

	Low-SES			High-SES		
	Slot	*Class*	*School*	*Slot*	*Class*	*School*
Semantic	**1**	6	D	**5**	4	C
Intervention	**2**	8	F	**6**	5	C
					3	B
Combined	**3**	7	E	**7**	1	A
Intervention	**4**	9	G	**8**	2	A
Control	**9**	10	I	**11**	12	H
Group	**10**	11	I	**12**	13	H

Note. Slot 6 was filled with two smaller classes to achieve the correct cell size.

Schools H and I came forward towards the end of the recruitment process and became the control schools. Control school H contained five parallel Year 1/2 classes, each with small numbers of Year One pupils, equating to two classes. School I had one straight Year 1 class and a mixed Y1/2 class, comparable to one and a half classes. The criterion for single-age classes was relaxed for the control group, since they would not be receiving the programme during the intervention year. The control classes filled slots 9-12, however each control school was treated as a single entity for analysis purposes. Once all slots were filled, remaining applicant schools J, K and L were informed that recruitment was complete and thanked for their interest.

While the IMD based on school postcode (school IMD) was initially used as a proxy for SES for the purposes of school randomisation, once parent/carer consent was gained the more accurate IMD based on pupil postcode (pupil IMD) was collected and used in all further analyses. A one sample t-test was conducted for each school to determine whether the school IMD differed from the pupils' IMD. A Bonferroni correction was applied to each comparison. Results in Table 12.4 indicate that the school IMD was significantly different to the pupil IMD in seven out of the nine schools. Nonetheless, the high-SES schools had all been randomised correctly, since they were still above average on the pupil IMD measure. The low-SES schools were also correctly categorised, except for school D, whose mean pupil IMD would place it in the high-SES category. This issue will be discussed further in section 12.4.4.7 on group equivalence.

Table 12.4

IMD Based on School vs. Pupil Postcode

School/ class	*School IMD*	*Pupil IMD M*	*Differences M*	*t*	*df*	*Sig. (2 tailed)*
School A	8.00	7.00	-1.000	3.620	45	.001*
School B	10.00	9.11	-0.895	2.445	18	.025
School C	9.00	8.03	-0.966	4.197	28	<.001*
School D	4.00	5.95	1.952	-5.968	20	<.001*
School E	3.00	4.17	1.174	-3.761	22	.001*
School F	2.00	3.54	1.542	-4.156	23	<.001*
School G	4.00	4.85	0.846	-2.335	25	.028
School H	10.00	7.00	-3.000	8.755	38	<.001*
School I	2.00	4.47	2.472	-7.554	35	<.001*

Note. * Significant at p=0.0056 (.05 divided by nine comparisons).

12.4.3.4 Pupil Recruitment and Inclusion Criteria

After recruitment, a researcher visit was made to confirm arrangements and to gain written consent from the Head teacher and Year One teachers. The *Parent/Carer Consent Form* and the *Research Information for Parents and Carers* (available in Appendix J and Appendix K) were then sent home with all pupils. Parent/carer consent was required for the test battery but not for the teaching programme, which comes under the auspices of the Head teacher. During the visit, the researcher took the opportunity to explain the project to the children using a standard protocol that covered the study rationale, pupil decision whether to take part, activity formats and an opportunity for questions.

No specific pupil eligibility criteria were implemented at the start of testing, because the full range of ability was required to test the effectiveness of the programme for mainstream teaching. It was not clear at the outset which children would be unable to manage the test battery, however several exclusions were implemented once testing began (see section 12.4.4.6).

12.4.3.5 Gaining Parent/Carer Consent

In addition to the main consent form, the following measures were implemented to maximise participant numbers and to reduce sample bias:

1. *Supplemental Information for Parents/Carers* to provide pictorial support for low levels of literacy (see Appendix L)
2. A short Youtube video describing the project and proposed assessments: https://www.youtube.com/watch?v=Q8aOmUpj4oE&feature=em-share_video_user
3. *Newsletter Blog* for parents for use in the school newsletter or in an introductory letter from the Head teacher (in Appendix M)
4. Discussions with parents/carers alongside the Year One teacher on the playground before and after school
5. Duplicate consent forms sent out if the first form was not returned
6. A short introductory meeting for parents/carers about the project and assessments, with an opportunity to ask questions.

Table 12.5 summarises the consent form returns by class.

Table 12.5

Signed Consent Forms by Class

Class	*Number of Pupils*	*Number of Consents*	*Permission Not Granted*	*Reasons if known*
Class 1	25	23	2	Anxious child, holiday
Class 2	26	25	1	Low attainment, holiday
Class 3	26	21	5	2 declined, 2 late, 1 no reply
Class 4	18	13	5	5 no reply; forms not resent
Class 5	19	17	2	2 no reply; forms not resent
Class 6	24	23	1	1 declined
Class 7	29	24	5	5 no reply after second form sent
Class 8	26	24	2	1 declined, 1 chicken pox
Class 9	26	26	0	All consented
Class 10 *	50	42	8	12 no reply; forms not re-sent
Class 11 *	42	40	2	1 ill health, 1 SEMH on reduced timetable
Totals	*311*	*278*	*33*	

Note. * Classes were combined for control Schools H and I.

Group 1 (semantic intervention) had 113 eligible participants. Since there were more potential candidates in high-SES classes 1 and 2, it was decided to restrict numbers in low-SES classes 3, 4 and 5 to aim for balanced numbers. This was achieved through careful adherence to the consent form deadline, and teachers were instructed not to send home duplicates once the required number was reached. This culminated in consent for 98 pupils. Group 2 (combined intervention) had 106 participants available for recruitment, and 98 parental consents were returned, so the two intervention groups had equal pupil numbers. Group 3 (waiting control) had a potential recruitment sample of 92 pupils. Since potential high-SES participants (Class 10) again outweighed low-SES candidates (Class 11), it was decided to adhere to a strict deadline for form submission in the low-SES school. The resulting 82 consents received meant that the control group was slightly smaller than the two intervention groups.

In summary, parent/carer consent forms were sent to the initial recruitment pool of 311 pupils. Overall, 89% of consent forms were signed and returned for 278 pupils to take part in the study; consent forms were not returned for 33 pupils.

12.4.3.6 Flow of Participants Through the Experimental Study

Details of the recruitment, allocation, and flow of participants through the study are presented in the *Flow of Participants Through the Intervention Study* in Appendix N. Table 12.6 details the exclusions that were implemented once testing started.

Table 12.6

Exclusions at Each Testing Point

Timepoint	*Number of Pupils*	*Reasons*
Time 1	4	Severe SEN, could not engage
	1	Significant behaviour needs, could not engage
Time 2	5	Moved house
	1	Refused testing
Time 3	1	Moved house
	1	Home schooled
	1	Excluded from class (SEMH)
Total	*14*	

A sample of adequate power was reached for Time 1 testing prior to the teaching intervention (N=273). Attrition from T1 to T3 amounted to nine pupils, culminating in a final sample for T3 analysis of 264 pupils (Semantic N=89, Combined N=95, Control N=81).

12.4.4 Group Equivalence

Since school D should have been in the high-SES block on the basis of pupil postcodes, it is important to establish whether randomisation groups are similar on this potentially influential variable in order to fairly compare their outcomes.

Table 12.7 displays the descriptive information by randomisation group for the pupils whose IMD was available for their postcode (*N*=263).

Table 12.7

Pupil IMD by Intervention Group

	N	*M*	*SD*	*Min*	*Max*
Group 1 - Semantic Intervention	93	6.62	2.600	2	10
Group 2 - Combined Intervention	95	5.73	2.171	2	10
Group 3 - Waiting Control	75	5.79	2.407	2	10

The mean pupil IMD was higher in the semantic group, leading to the decision to perform a one-way ANOVA to check for any significant SES differences between the three conditions. Assumptions were met for normal distribution of residuals and equality of variances on the Levene's test, $F(2,260)=2.191$, $p=.11$. The ANOVA revealed a significant difference between groups based on pupil IMD, $F(2,260)=3.981$, $p=.02$, $\eta_p^2=.03$. Post hoc analysis with Bonferroni correction indicated that the semantic condition had significantly higher SES than the combined condition ($p=.03$). However, no significant difference was found between the semantic condition and controls ($p=.99$) or between the combined condition and controls ($p=.07$).

Since the groups differed on the SES variable, it is important to determine whether differences exist on other measures. To assess this possibility, a one-way ANOVA was conducted on each T1 dependent variable. Residuals were normally distributed in each case, and homogeneity of variance was confirmed for all measures except nonword spelling. ANOVA results are displayed in Table 12.8.

Table 12.8

ANOVA Results - Group Differences on Dependent Variables

Variable	*F*	*df 1*	*df 2*	*Sig.*
PhAB2 Rhyme	2.365	2	270	.10
PhAB2 Alliteration	1.561	2	270	.21
CTOPP2 Elision	1.218	2	270	.30
PhAB2 Nonword Reading	2.185	2	267	.12
BPVS3 Receptive Vocabulary	4.254	2	270	.02*
CELF4 Expressive Vocabulary	2.772	2	270	.06
Taught Vocabulary Definitions	2.281	2	268	.10

Note. * Significant at p=.05.

Significant group differences were discovered only on the BPVS3, $F(2,270)=4.254$, $p=.02$; $\eta_p^2=.03$, equal to 3% of the variance in performance. As suspected, post hoc comparisons with Bonferroni adjustment revealed that the semantic condition had significantly higher BPVS3 results than the combined condition (p=.01), but there were no significant differences between the semantic condition and controls (p=.41) or between the combined condition and controls (p=.59). Raw scores results are available for all variables in Chapter 13. Due to unequal variances on the nonword spelling test, a nonparametric Welch's ANOVA test was conducted, after checking that other assumptions were met. Results revealed no significant group differences on nonword spelling performance, $F(2,270)=.206$, $p=.84$. Baseline equivalence between randomisation groups was established for all remaining measures.

Given the significant pretest difference on the T1 BPVS3 score, future analyses will control for receptive vocabulary wherever possible, in line with a range of vocabulary intervention studies that have similarly included a receptive vocabulary covariate due to pretest differences (Damhuis et al., 2016, Ehri et al., 2020, Janssen et al., 2018). Adjustment for covariate imbalance is a design feature that strengthens the reliability of results by reducing the variance caused by the group discrepancy (Altman, 1998; Slavin, 2008).

A further potential group difference worth exploring is the variability in teaching experience. As seen in the responses to the initial questionnaire in Appendix S, three out of five of the semantic group teachers had considerably less experience in terms of years in teaching and breadth of primary experience compared to the combined and control group teachers. On the other hand, the priority given to vocabulary instruction and confidence scores for these newer teachers were similar to the other groups, and two of these teachers had some previous vocabulary training. Notwithstanding the value of experience, the standardised intervention protocol and resources were intended to ease this differential.

12.5 Materials: The Experimental Test Battery

The cross-sectional study confirmed that the assessment battery could sensitively measure the performance of Year One mainstream pupils on the majority of chosen constructs. The battery likewise achieved its aim to provide short, interactive activities that were enjoyable for the children. The trialled assessments were therefore utilised in the intervention study, subject to the modifications described in Chapter 8, section 8.3.1. Two additional assessments were added to the experimental test battery, i.e. the CTOPP2 Elision task and a bespoke tool to measure gains in taught vocabulary (described in section 12.5.1 and 12.5.2 below). Table 12.9 lists the assessments used in the experimental study. Reliability and validity data are provided in Appendix B.

Table 12.9

Assessment Battery for the Vocabulary Intervention Study

Variables	*Test*	*Age Norms*
Receptive Vocabulary	BPVS3 British Picture Vocabulary Test	3-16
Expressive Vocabulary	CELF4 Expressive Vocabulary	5-9
Taught Vocabulary	Taught Vocabulary Definitions	none
Phonemic Awareness	PhAB2 Alliteration	5-11
Phonemic Awareness	CTOPP2 Elision	4-24
Rhyme Awareness	PhAB2 Rhyme	5-6
Nonword Reading	PhAB2 Nonword Reading	5-11
Nonword Spelling	Nonword Spelling	none

12.5.1 Replaced Phonemic Awareness Assessment: CTOPP2 Elision Test

The CTOPP2 (Wagner et al., 2013) assesses reading-related phonological processing skills. The Elision subtest measures the ability to transpose phonological segments from a spoken word to form another word. In this oral task, words of increasing length are presented, and the child is asked to say the word without a given syllable or phoneme, e.g. Say *sunshine* without saying *sun*; Say *farm* without saying /*f*/. Each correct item receives a score of one point. The Elision task contains 34 items. There are no unmarked practice items, however pupils receive feedback on the first 14 items. Items 1- 9 tap the ability to delete a word part or syllable, items 10-15 measure elision of the first phoneme, and items 16-34 require deletion of a phoneme in a non-initial position. Testing is discontinued when the child makes three errors or if the child is unable to provide a correct answer to at least one of the first three test items.

12.5.2 Additional Taught Vocabulary Definitions Task

12.5.2.1 Overview of the Taught Vocabulary Definitions Task

An assessment was devised to index pupils' developing knowledge of the taught vocabulary at each testing point. In line with the extant literature, a definitions task was developed with the aim of capturing increases in vocabulary size as well as depth of understanding. As a late-developing and particularly rigorous test of expressive vocabulary skill (Kameenui et al., 1987), a definitions task is useful to prevent ceiling effects in the Year One age group. A variety of standardised vocabulary definitions tests were consulted to glean best practice for instructions, scoring and prompts, leading to the protocols adopted here.

Responses were written verbatim onto the test record in *Taught Vocabulary Definitions Test* in Appendix O. In an earlier trial, pupil responses were audio-recorded to support transcription, however this was deemed unnecessary, since pupil responses could be written quickly and accurately by the testers.

12.5.2.2 Randomisation of the Probe Vocabulary

A subset of 21 words was randomly selected for the probe from the full set of 119 items using Graphpad (2018; accessed 26/7/18). The full word set can be seen in Appendix P (*Vocabulary Teaching Blocks Overview*), and the probe items are on the test record in Appendix O. Two words were randomly selected from each of the 11 teaching blocks, excluding the trial block at the beginning and the final teaching block.

After randomisation, several items were replaced to minimise confusion. In block 10, the word *camouflage* was too close in meaning to a previous item (*disguise*) and was therefore discarded. Also, *waist* could potentially be confused with its dual meaning of waste. In each case, the next consecutive item in the block was chosen. Once testing began, it became apparent that children were confused by the American term *headlamp* (headtorch in the UK), so this item was removed. The final randomised probe of 21 items amounted to 21% of the taught vocabulary.

12.5.2.3 Equivalence of the Word Sets

To ensure that the probe items (*N*=21) were a valid measure of the whole set (*N*=119), a comparison was made of the relevant word characteristics affecting word learning described in Chapter 2 (section 2.4.1). AoA ratings were taken from Kuperman et al. (2012), which offered a large word set. The online MRC Psycholinguistic Database (Wilson, 1988) was consulted for ratings of written word frequency, word length (number of syllables), imageability and word type. The online Phonotactic Probability Calculator (described in Vitevitch & Luce, 2004) provided information on phonotactic probability, more specifically the sum of biphone probabilities to indicate the prevalence of the sound patterns in a word. These variables are detailed in Appendix O for the taught vocabulary. Descriptive statistics including the mean, range and standard deviation for variables are available in Table 12.10, and the percentage of each word type is shown in Table 12.11.

Table 12.10

Comparison of Lexical Characteristics – Definitions Probe and Full Vocabulary Set

	N	*M*	*Min*	*Max*	*SD*
Age of Acquisition					
Probe	21	7.96	5.37	11.63	1.500
Whole Set	119	7.52	3.95	12.50	1.460
Written Word Frequency					
Probe	20	25.45	1	69	25.099
Whole Set	85	27.81	1	223	42.124
Word Length (number of syllables)					
Probe	21	1.86	1	5	1.108
Whole Set	119	1.87	1	5	0.882
Phonotactic (Biphone) Probability					
Probe	21	.0201	.0019	.0928	0.016
Whole Set	119	.0249	.0006	.0928	0.028

Table 12.11

Comparison of Word Type - Definitions Probe and Full Vocabulary Set

	Nouns	*Verbs*	*Adjectives*	*Adverbs*
Probe	38.1%	52.4%	9.5%	0%
Whole Set	35.3%	49.6%	13.4%	1.7%

An independent samples t test compared means of the full set and probe items. There were no significant differences in AoA, $t(20)=1.322$, $p=.20$, written frequency, $t(19)=-.324$, $p=.75$, word length, $t(19)=.326$, $p=.75$ or phonotactic probability $t(20)=.791$, $p=.44$.

Although AoA means were noticeably similar, the range of scores was more restricted in the probe set. Written frequency ratings were available for most but not all items, as noted in Appendix O. Although the means are similar, the probe had a frequency range that was lower than the main vocabulary list, indicating that some of the probe items may have been more difficult to learn. Descriptive statistics for word length and word type indicate a similar balance between the word sets. Even though biphone probabilities display similar means, the probe has a slightly more restricted range, with a higher minimum value, signifying that the sound patterns were more common and therefore potentially easier to acquire.

Research shows that imageability also impacts on vocabulary difficulty (Gilhooley & Logie, 1980). Unfortunately, these norms were only available for around half of the probe items (10 out of 21) and main vocabulary items (66 out of 119). This variable will therefore not be analysed here. The available imageability norms for the taught vocabulary are shown in Appendix O.

The above analyses demonstrate that the probe and full word sets were statistically similar on a range of important lexical variables. The assessment probe was therefore a valid representation of the full set of taught vocabulary.

To determine the reliability of the taught vocabulary measure, an item analysis was carried out on the T1 probe data. The Cronbach's alpha for the 21-item test showed a good level of reliability ($\alpha=.78$), suggesting that the test achieved good internal consistency.

12.5.2.4 Test Administration

The task was introduced by saying: '*I am going to say some words. Can you tell me what they mean? You can say if you don't know the word'.* The tester allowed up to 10 seconds for the child to consider a response before moving on to the next item. Two unscored demonstration items were practised first to help the child to understand the difference between a definition and a sentence. Repetitions of the stimulus word were given freely, and there was no requirement to note these on the record sheet. Prompts were used consistently to draw out the maximum knowledge from participants. One prompt was allowed per definition**,** indicated by a (P) on the record sheet:

1. After the practice items, if the child provided a sentence response, the tester said *'That is a good sentence. Can you tell me what x means?'* This prompt was only provided once in order not to annoy or dishearten the pupil
2. If the child mistakenly heard a different word, e.g. heard 'June' for /*dunes*/, the tester said *'Not June, it is dune',* emphasising the misheard word and then allowed the child to give a response
3. If the response was unclear or too vague for scoring, the tester asked, '*Can you tell me more about what x means?'*
4. If the child gave a nonverbal response, e.g. pointing or an action, the tester prompted with **'***Well done. Can you also tell me what x means'?*

Pupil responses were written verbatim onto the recording sheet for later scoring.

12.5.2.5 Scoring

A fine scoring matrix was developed to account for all pupil responses and to further enhance consistency of scoring. An iterative process was used until all new responses were included. The scoring protocol is found in Appendix O. No inter-rater reliability check was conducted, since the definitions task was jointly marked by a Speech and Language Therapist and the researcher to achieve optimal marking consistency. Any items of disagreement were discussed until a joint conclusion was reached. There were three possible scores for each item: 0 for no understanding demonstrated, 1 for a partial but imprecise response or 2 if clear understanding was demonstrated. More concise scoring guidelines are detailed in Appendix O.

12.6 Procedure: The Experimental Test Battery

12.6.1 Data Management and Confidentiality

Confidentiality of all pupil data was safeguarded through a data management plan. Each participant was assigned a unique Participant Identification Number (PIN) after written parent/carer consent was received. The PIN, rather than the child's name, was written on the test record. Audio recordings were similarly labelled with the PIN. All pupil data will be destroyed after a period of five years after completion of the PhD, in line with ethical approval.

Prior to the start of testing, teachers were asked to provide a range of pupil-level data, some of which was confidential in nature, i.e. (1) age, (2) gender, (3) prior attainment, (4) SEN, (5) ethnicity, (6) EAL designation, (7) pupil's home postcode and (8) Pupil Premium status. This information and pupils' test results were entered into an anonymised SPSS dataset. The master list of pupil-level data with children's names was securely stored on the researcher's encrypted laptop with digital password protection and backed up in the same way on a secure university drive. A hard copy was stored in a locked filing cabinet.

12.6.2 Pupil Testing Sessions

Pupils were tested individually in a quiet area of the school by either one or two testers. Two short testing sessions, each lasting around 20 minutes (set out in Table 12.12), were administered on separate days over the course of a week to minimise pupil fatigue. Trialling during the cross-sectional battery suggested this was manageable for pupils, however participants were told they could have a break or stop testing at any point. The PhAB2 Nonword Reading task was audio-recorded for later reliability checks.

Table 12.12

Pupil Testing Sessions for the Vocabulary Intervention Study

Session 1	*Session 2*
PhAB2 Rhyme	CELF4 Expressive Vocabulary
CTOPP2 Elision	BAS3 Matrices *
PhAB2 Alliteration	Nonword Spelling **
PhAB2 Nonword Reading	Taught Vocabulary Definitions
BPVS3 Receptive Vocabulary	

Note. * Only used at T1 testing to establish participant characteristics. ** The Nonword Spelling test was administered in small groups by the tester or the teacher.

12.6.3 Assessor Background and Training

Six testers were recruited to administer the test battery for the vocabulary intervention study. The researcher was an additional member of the tester team, acting as the team supervisor, as well as administering assessments if required. The testers were drawn from the fields of education, psychology and speech and language therapy and were chosen for their interest and experience in working with children.

All testers were known to the researcher and undertook a full DBS check, which was made available for schools to inspect upon arrival. A confidentiality agreement (provided by the contracts team at the Sheffield University Research Services department) was signed by the four testers who were not members of the Communication and Interaction team (this clause is already in the employment contract for the Communication and Interaction team).

Testers received two individual training sessions, each lasting approximately two hours. Each session included practice to (1) administer all items, including discontinuation rules, (2) listen carefully to pupil responses, (3) audio record pupil responses, (4) use encouraging language and (5) score tests accurately. Individual assessors completed an assessment after each training session, with follow-up practice until 100% accuracy was achieved. Considerably more practice was provided for the PhAB2 Nonword Reading test at T1 to improve accuracy and agreement, which was also formally monitored through an inter-rater reliability check described in Chapter 13 section 13.6. The researcher observed each tester on the first day of testing to monitor adherence to the instructions and protocol. Testers were observed on at least one further occasion at each time point with guidance provided as required.

Assessors visited the classroom beforehand to get acquainted with the children. Assessments took place in a public area or in a room with an open door as a safeguarding measure for both testers and pupils. Testers were blind to the group allocation of the participants they assessed.

12.7 Materials: The Vocabulary Intervention

12.7.1 Programme Overview

The aim was to devise an effective and inclusive model of classroom vocabulary teaching for Year One pupils that would be time-efficient for teachers and enjoyable for pupils. A strong instructional design was a vital basis upon which to compare the combined vs. semantic methods in terms of their impact on vocabulary, phonemic awareness and phonic reading. The materials and protocols embedded in this section reflect the evidence-based vocabulary teaching principles discussed in Chapter 10.

The programme involved daily whole-class teaching of tier two vocabulary (Beck, 2002) drawn from high-quality read-aloud storybooks. An explicit instruction model was employed to boost outcomes for target vocabulary: one target word was taught per day, followed by a pupil activity and a review session. A range of rich instruction activities were incorporated to further support word learning (Beck et al., 2013). A specific review schedule was included to optimise retention.

The teachers received the same set of intervention resources, with differences only in the type of vocabulary facilitation cues (semantic or combined). The programme included (1) a set of fiction books, (2) vocabulary linked to each text and (3) a manualised scheme of intervention materials. Resources for the intervention activities and record-keeping templates were supplied to teachers both in hard copy and electronically to support consistency and fidelity of delivery.

12.7.2 Storybook Selection

The selection of books was based on Year One fiction lists compiled by the Babcock Literacy Team and the Centre for Literacy in Primary Education (CLPE, 2018), as these are two popular sources of high-quality texts commonly used by Devon teachers. A visit was made to the main Schools Library Service to peruse books individually to locate texts that met certain criteria for inclusion in the programme, including (1) at least 10 tier two words, (2) engaging pictures and format and (3) appropriate language for a mainstream Year One class.

This process identified relevant texts, which were then discussed with three Year One teachers outside the study. The final selection is displayed in Table 12.13 in order of introduction over the school year.

Table 12.13

Storybooks Chosen for the Vocabulary Intervention

Block	*Title*	*Author*
Autumn 2018		
Trial	Naughty Bus	Jan and Jerry Oke
1	Augustus and his Smile	Catherine Rayner
2	How to Babysit a Grandad	Jean Reagan
3	How to Catch Santa	Jean Reagan
Spring 2019		
4	Don’t Spill the Milk	Stephen Davies; Christopher Corr
5	How to Hide a Lion at School	Helen Stephens
6	Could a Penguin Ride a Bike?	Camilla de la Bedoyere
7	The Day Louis Got Eaten	John Fardell
8	Previously	Alan Ahlberg
9	Wanted the Perfect Pet	Fiona Robertson
Summer 2019		
10	Traction Man is Here	Mini Grey
11	Mrs Armitage on Wheels	Quentin Blake
12	The Sand Horse	Ann Turnbull

Each school received two free books: the trial book was kindly donated by the author – Jan Oke, and Scholastic Publishing donated *How to Hide a Lion at School.* Five schools were members of the Schools Library Service and received the rest of the book set without cost for the academic year. Two schools who did not buy into this service decided to purchase their own sets. The researcher topped up any books not procured from other sources to facilitate smooth intervention delivery. Teachers were requested to keep the books in the intervention box to minimise (and standardise) exposure outside of lessons.

12.7.3 Target Vocabulary Selection

A set of nine vocabulary items was chosen from each storybook that met the criteria for medium frequency tier two words (Beck et al., 2013). The selected vocabulary had AoA norms from age 5-10 to cater for the wide ability range in a mainstream Year One class. The following word types were excluded:

- phrases, e.g. bundle up. These are not generally included in vocabulary databases, so norms would be hard to acquire
- words unfamiliar to the UK context, e.g. lumber, sprinkler
- words already selected as a target in a previous book
- low frequency (tier 3) words, e.g. armour

12.7.4 The Manualised Programme

A *Teaching Manual for the Vocabulary Programme* in Appendix P was provided to each participating teacher with the content listed below:

1. Vocabulary teaching protocol
2. Teacher planning and review format
3. Teaching cue card with suggested wording
4. Symbolised vocabulary cards
5. Vocabulary block lists
6. Pupil self-rating scale
7. Pupil-friendly definitions
8. Menu of activities
9. Word spy record
10. Word wall resources

12.7.4.1 Vocabulary Teaching Protocol

A concise two-sided sheet explained the standard teaching protocol, including classroom setup, daily teaching format and use of the resources.

12.7.4.2 Teacher Planning and Recording Sheet

Weekly planning sheets were prepopulated with the target vocabulary, definitions and review items. The review schedule was based on principles of interleaved learning to support retrieval (Kang, 2016), including review of the target word for that day, the previous day and the previous week. Teachers were requested to tick when they completed each lesson and to note down feedback at the end of each week.

12.7.4.3 Teaching Cue Card with Suggested Wording

The vocabulary instruction card was produced in a visual format using WidgitOnline (Widgit, 2007). Widgit provided complimentary use of the symbol processing software for use in the project schools. Colours adopted from Parsons and Branaghan (2014) were used to indicate cue type: red for what is said, green for what we mean.

The teaching card contained six cues. The semantic group learned vocabulary using three prompts (meaning, use in a sentence, acting it out), linked to rich instruction by Beck et al. (2013). Each prompt was used twice in the lesson. The combined group had the same three semantic cues supplemented with three sound-based strategies gleaned from the combined instruction literature (large segments of syllables or rhyme, sound out the word, say the word clearly). Both teaching groups thus had six vocabulary learning prompts.

The cue card was provided on laminated card in both A4 and A3 size, as well as in digital format for the interactive whiteboard. Suggested wording was provided to encourage teachers to say the target word 8-10 times during the lesson to boost word learning in line with research suggesting that extra exposure can support those with vocabulary learning difficulties (Best, 2005; Dollaghan, 1987).

12.7.4.4 Symbolised Vocabulary Cards

WidgitOnline was used to create visual symbol cards for all vocabulary items. Images were created on a grid of nine per page in colour with the text underneath. Teachers were requested not to draw attention to the written word in order to minimise the effect of orthography on new word learning. If Widgit did not contain the target symbol, or if trialling suggested it was ambiguous for Year One children, then a copyright-free image was substituted with appropriate acknowledgment. The vocabulary card sets were printed in A4 and A3 format for teaching and for the word wall. The card sets and individual PDFs were also provided in digital format for the interactive whiteboard.

12.7.4.5 Vocabulary Lists

The vocabulary list for each teaching block was given in the manual. See-through sticky notes were provided so teachers could highlight words in the storybooks. This task was carried out during the training session to familiarise teachers with the target words.

12.7.4.6 Pupil Self-rating Scale

The vocabulary items were typed on to a simple self-evaluation format for pupils to complete at the start and end of the teaching block. The pupils ticked binary yes / no answers to indicate whether they *knew* the word, which was also read out by the teacher. The teacher explained that *knowing* a word entails understanding the meaning. A master copy was provided for teachers to photocopy for each pupil on a weekly basis.

12.7.4.7 Pupil-Friendly Definitions

A definition was created for each vocabulary item using the Wordsmyth (2014) website. The aim was to write a simple definition using everyday tier one vocabulary (Beck, 2002) that would be straightforward for children to understand. Trialling showed that several definitions were still too difficult, so the Cobuild Primary Learner's Dictionary (Collins, 2014) was consulted for simpler wording, supplemented with suggestions by the trial teachers. The definitions were printed out on a double-sided A4 sheet; they were also written on to the appropriate weekly planning and recording sheet to aid consistency and save teacher time.

12.7.4.8 Menu of Pupil Games

Five multi-sensory games were created by the researcher, some of which were adapted from Parsons and Branaghan (2014) with kind permission from the authors. The initial five games were the Beetle Game, Dice Game, Spinner Game, Vocabulary Swat and Modified Pairs. The game formats for the two groups varied only in terms of the cue types and corresponding red/green colours. A set of 10 ready-made games was provided for each class - two of each type, in separate clear zipped wallets with all resources and instructions included. A total of 10 games was considered sufficient for pupils to play in small groups of three, although group size was not a specified variable. The games were designed to be interesting to Year One pupils and to require minimal adult support. A further consideration was to ensure that games could be played quietly to enable children to hear the words clearly. Specific resources were provided to achieve this aim, including foam dice, feathers and a traffic light prompt for voice levels.

In February 2019, half-way through the teaching programme, the researcher refreshed the resources in the game wallets after consultation with teachers and replaced two games that teachers and pupils found onerous. The Modified Pairs game and Vocabulary Swat required multiple sets of symbol cards, therefore creating extra workload for staff. In addition, the Modified Pairs game was difficult for many children to play independently, demanding more adult time. Two games were therefore substituted that were easier to administer and play: the Fortune Teller and an adapted version of Vocabulary Swat. The final set of games is available in Appendix P.

An analysis was performed of the engagement demand of the games according to the construct of *involvement load* by Nation and Webb (2011). *Checklists for Technique Feature Analysis* in Appendix Q were completed for the five original games in the combined and semantic conditions, which resulted in similar scores, as summarised in Table 12.14. Analysis of the replacement Fortune Teller game (score=13) and the adapted Vocabulary Swat game (score=14) yielded the same scores in both conditions. This higher difficulty level was considered acceptable for the second half of the academic year.

Table 12.14

Involvement Load Scores of Five Games

	N	*M*	*Min*	*Max*
Semantic Intervention	5	10.6	9	12
Combined Intervention	5	10.3	9	14

12.7.4.9 Word Spy Record

The *Word Spy Record* with symbolised vocabulary was housed on the word wall to enable staff and pupils to place a tick next to the target word each time it was used outside the lesson. The aim was to encourage application of taught vocabulary. Symbols were included so pupils could complete this task independently if desired.

12.7.4.10 Word Wall Resources

A standard suite of *Word Wall Resources* was provided for teachers to create a class display. Items included (1) a *Word of the Day* template, (2) a vocabulary chart to display previously taught words, (3) a traffic light image to remind pupils to use quiet voices during game times, (4) an extra instructional cue card, (5) the symbolised set of nine vocabulary for the teaching block and (6) the word spy sheet. Teachers were also encouraged to display pupils' written work demonstrating use of the target words.

12.7.4.11 Storage of Resources

A concertina folder was provided for weekly storage of the planning sheets, vocabulary cards, words spy sheets and self-rating scales. Teachers generally kept the teaching items in a plastic storage crate, which was brought to the lesson each day.

12.8 Procedure: The Vocabulary Intervention

12.8.1 Piloting the Intervention

The intervention materials and procedure were trialled in three Year One classrooms in July 2018 for two weeks to further improve their validity and effectiveness before use in the Autumn term. The researcher also piloted a training session, lesson observation and all intervention materials. The story text used for the trial teaching block was Augustus and his Smile.

In addition to trialling the teaching protocol and materials, a consistency check was made of the researcher’s tier two vocabulary choices compared to those selected by the Year One trial teachers. Using the book Augustus and his Smile, the researcher selected the following tier two words: crept, cluster, searched, scaled, swirled, frost, patterns, paraded, realised, shoal. Teacher 1 selected the same 10 words, teacher 2 chose 8/10 of the words, and teacher 3 chose 8/10 of the words, culminating in 87% agreement, which was considered an acceptable level to take the researcher’s vocabulary choices forward.

The intervention trial demonstrated that the protocol and resources were useful to teachers and enjoyable for pupils. Adaptations made to the intervention following the trial feedback are included in Table 12.15.

Table 12.15

Adaptations Resulting from the Intervention Pilot

Intervention Feature	*Adaptations*
Vocabulary teaching protocol	A pre-teaching session was introduced on day one
Teacher planning and recording sheet	No changes recommended
Instructional cue card	Several wording changes were made to the combined sound-meaning cue card
Symbolised vocabulary sets	These were made available in digital format; 10% of symbols were changed to improve comprehension
Vocabulary lists for teaching blocks	The number of items per block was lowered to 9 to accommodate the preteaching session on day 1
Vocabulary self-rating scale	Three levels of word knowledge (high/a bit/not at all) were reduced to two judgements (yes/no)
Definitions sheet	30% of the definitions were simplified even further
Menu of activities	Treasure chest game was discarded due to poor durability
Word spy record	No adaptations made; it was decided to retain this feature to see if it enhanced learning
Word wall resources	A traffic light prompt was included to reduce noise levels during the pupil games

The intervention trial also generated a range of practical suggestions for introducing the games to young children, which formed part of the advice for the substantive intervention programme. Trial teachers made the following recommendations: (1) familiarising children with one game for a whole week, (2) initially splitting the class into two large groups led by the teacher and TA to learn the game formats and (3) forming mixed-ability groups so that an able pupil could support others, thus making the group more independent and reducing the demand on adults.

12.8.2 The Vocabulary Teaching Protocol

A whole-class programme of vocabulary instruction was delivered for a short session each day. The vocabulary was linked to high-quality storybooks, which were read to the pupils outside the vocabulary session. Schools organised the vocabulary programme in two alternative ways: the majority used the set storybooks as a basis for their literacy curriculum, whilst two classes used the books during a separate story time. In either structure, the vocabulary lesson was delivered at a different time to story reading, and schools were asked not to directly teach any other story vocabulary. Prior to intervention, a word wall was set up in each classroom.

The session followed the STAR approach suggested in Parsons and Branaghan (2014). The two intervention conditions shared the same structure and were designed to run over 13 fortnightly teaching blocks, including the trial block. The STAR components are described below.

12.8.2.1 Select

Nine tier two words were prepopulated onto the teacher planning and recording sheet. Symbolised vocabulary cards were created to accompany each target word.

12.8.2.2 Teach

The relevant teaching cue card was displayed alongside the picture symbol for the word of the day, either on the digital whiteboard or in A3 card format. The teacher considered each cue on the card and asked for pupil responses, using a question-answer format or talk partners. Standardised wording was provided, so the pupils would hear the word spoken at least eight times. A standard pupil-friendly definition was provided, but this was not a mandatory part of the protocol, as teachers were allowed to adapt their language according to the pupils' needs. The word of the day was displayed prominently on the word wall to facilitate application throughout the day. For the semantic group, three meaning-based cues were discussed (each twice): word meaning, use in a sentence and act out/describe the word. The combined group also discussed these three meaning prompts (once each) and additionally worked on three sound-based facilitation cues: say it clearly, break it into syllables/rhyme and sound out the phonemes.

Day 1 of the teaching block took the form of a preteaching session to familiarise pupils with the target vocabulary for the two-week block. Days 2-10 used the teaching protocol above. In the case of teacher absence or special events, teachers were asked to consider the following options: (1) facilitate the Teaching Assistant to deliver the programme, (2) double up on the target word the following day or (3) use the planned catch-up weeks allocated in the programme timetable. Teachers were asked to note any gaps in teaching on the planning and recording sheet.

12.8.2.3 Apply/Activate

Pupils received an additional opportunity to verbally practise target words and the relevant learning cues through a menu of five games. After an introductory period to learn the game formats, pupils generally worked in groups of 2-4 pupils. Introduction to the games took a variety of formats, but by the end of the second teaching block, 7 out of 8 classes were playing all of the games on a rotating basis as planned. The single school not doing this at the first consistency check was provided with further support to adhere to the protocol. On day 10, the game was omitted to provide time for pupils to complete the self-rating scale at the end of the teaching block.

12.8.2.4 Review

In the final part of the session, the teacher gathered the pupils together for a brief review of three target words based on the principles of interleaved learning and distributed practice (expanding intervals) – the word of the day, the previous day and the previous week. Teachers were allowed to decide on the format for this review activity, however it was suggested that they provide the word and ask for a definition or a sentence to encourage expressive vocabulary usage. All symbol cards for previously taught words were available on the word chart displayed on the word wall. Children were encouraged to tick the Word Spy Record if they heard or used a target word.

Teachers were asked not to send the target vocabulary home to minimise differences on this variable.

12.8.3 Amount of Instruction

The daily lesson was intended to last for 10-12 minutes per day to fit into the busy school day. Suggested timings were as follows:

- *4-5 minutes*: *teach* the vocabulary using the cue card and symbol
- *4-5 minutes*: *activate* and apply the word of the day through a game, including setup and clearing away
- *2 minutes*: review target word of the day, previous day and previous week.

On the basis of 12 minutes per day, children received 60 minutes of direct vocabulary instruction per week. Over the course of the 13 fortnightly teaching blocks, the total vocabulary training time was 26 hours. Observations carried out for the consistency checks confirmed that the observed lessons generally lasted between 10 and 15 minutes, close to the suggested average of 12 minutes. Most teachers read the book in full at the start of the week, and then read excerpts daily to point out target vocabulary.

12.8.4 Intervention Training for Teachers

Schools and teachers were not made aware of their intervention condition. The Year One teachers learned the same intervention structure, apart from the teaching cues, i.e. combined or semantic. This does not present an ethical dilemma, as both types of vocabulary instruction have evidence of efficacy, and both interventions were premised on strong pedagogical principles.

A two-hour training session was delivered in each school. Schools with two Year One classes (two cases) received the training together. Otherwise, teachers received individual training (five cases). As part of the training session, each teacher received a storage crate with the teacher's manual, laminated programme resources, storybooks, word wall display, vocabulary symbol cards and a concertina storage folder. The structure of the training session is outlined in Table 12.16, and the *Vocabulary Training Session Presentation* can be viewed in Appendix R.

Table 12.16

Content of Training Session for Year One Teachers

1.	Pre-intervention *Teacher Questionnaire* (available in Appendix S)
2.	Project overview (alternative approach not discussed)
3.	Familiarisation with the story books and opportunity to highlight the target vocabulary
4.	Classroom organisation and setting up the word wall
5.	Demonstration of the STAR approach protocol
6.	Time to explore the vocabulary resources
7.	Play each of the practice games and discuss organisation
8.	Consistency checks, administrative issues and diary dates

Teachers had the opportunity to ask questions at the session, and the researcher's phone number and email address were provided in case any queries arose. Researcher visits were made four times during the intervention (see section 12.8.5), providing further opportunities for feedback and questions. The researcher sent email updates once per month throughout the intervention period.

12.8.5 Fidelity Checks

Adherence to the intervention protocol was measured through three lesson observations using a *Consistency Checklist for Lesson Observations* (in Appendix T). An informal supportive observation was made during the trial teaching block in October 2018 without a score sheet. A further three scored observations were made to observe adherence to the protocol. These took place once per term - in November 2018, February 2019 and May 2019. The timings and procedure were undertaken in accordance with union guidelines for teacher observation (NASUWT, 2018).

Ten intervention components were observed and scored on a binary scale of 0 or 1. Components for development were discussed with the teacher as well as positive points from the observation. A score of 9 or 10 was considered excellent, a score of 8 was good, and scores of 7 and below were deemed inadequate, requiring a more in-depth discussion and review of the protocol. Teachers' viewpoints and suggestions regarding the intervention were also gathered during these visits. The 10 consistency check items are set out below.

1. The targeted word is taught daily, with notations made on planning sheet if lessons are missed

2. Cues on the vocabulary teaching card are included in the lesson

3. Approximate timings are followed: 5 minutes teaching, 5 minutes game, 2 minutes review

4. Simple definitions are used, although these can be expanded

5. Games are updated fortnightly with the new vocabulary cards

6. Noise levels are monitored during activity time

7. Pupils have access to the full range of games over the fortnightly teaching block

8. Words are reviewed according to the planning sheet

9. Pupil self-rating scales are completed at the start and end of each fortnightly teaching block

10. The word wall is updated daily with the new target word and fortnightly for the new block

Fidelity check results showing the number of teachers who achieved each item are depicted in Table 12.17 below.

Table 12.17

Intervention Fidelity Check (nine classes)

Item	First observation		Second observation		Third observation	
	No. of teachers	*% fidelity*	*No. of teachers*	*% fidelity*	*No. of teachers*	*% fidelity*
1	9	100	9	100	9	100
2	8	89	9	100	9	100
3	8	89	9	100	8	89
4	9	100	9	100	9	100
5	9	100	9	100	*	*
6	9	100	9	100	7	78
7	6	67	8	89	9	100
8	9	100	9	100	8	89
9	9	100	9	100	9	100
10	8	89	9	100	9	100
Mean score	9.3		9.9		8.6	

Note. * Item 5 was no longer applicable at T3, since the games requiring vocabulary cards to be changed were removed midway through the programme.

A high level of fidelity was observed and documented at each observation point. After the first check, one classroom received further support on several components, i.e. to implement the full range of games, to stick to the cues on the vocabulary teaching card and to monitor noise levels during the game. These aspects were rectified after discussion, with full marks achieved at subsequent observation points. On the last observation, two teachers received a lower score on item 6 due to pupil noise levels, one teacher received a lower score on item 3 reflecting a lesson longer than the designated 12 minutes, and a teacher needed support on item 8 regarding review of the correct words. On average, across all three lesson observations, teachers achieved a high mean consistency rating of 9.3 out of 10 possible points, indicating that the vocabulary lesson was delivered as intended.

The daily planning and recording format acted as a further measure of fidelity. Information gleaned from the teachers' record-keeping provided information about the number of lessons delivered.

12.9 The Teacher Questionnaire

Two *Teacher Questionnaires* (available in Appendix S) were devised to gather information from participating teachers. Two item types were included. Teachers responded to statements on a five-point Likert scale (1=low, 5=high). Additional free text fields were provided to enable further comments. The questionnaires were piloted with a small sample of teachers not in the intervention study.

For the pre-intervention questionnaire, teachers responded to six items. Two Likert scale questions were included on the importance of vocabulary teaching and confidence in vocabulary teaching. The rest were free field questions to describe existing vocabulary practice, perceived barriers to vocabulary instruction, vocabulary training and range of teaching experience. The questionnaire was administered to intervention teachers prior to the training session and took around 10 minutes to fill in. In the control schools, the headteachers completed the questionnaire on behalf of their Year One teachers to minimise researcher involvement at this stage.

The post-intervention questionnaire was administered to the intervention teachers. The questionnaire consisted of six Likert scale items with a field for extra comments. The question regarding teacher confidence was included to enable comparison to initial ratings. Subsequent items focussed on programme evaluation. Teachers rated the effectiveness of the whole programme and each element separately in terms of perceived impact on vocabulary attainment, before moving onto the feasibility of implementing each programme element in future. Finally, teachers had the opportunity to describe challenges and to suggest modifications to the programme. The second questionnaire was completed at the exit interview in 20-30 minutes.

The teacher questionnaire was analysed in two ways. Mean Likert scale ratings were calculated to establish perceived impact on teachers and students. Additionally, the free field data was subject to thematic analysis to draw out the salient qualitative themes.

12.10 Data Analysis

The results of the vocabulary intervention study in the next chapter are based on a range of analyses. Descriptive raw score information was collated for the three randomisation groups. ANOVA and ANCOVA analyses were used as the main statistical analyses to compare raw score means on the outcome variables for the three groups. Since there was a group difference on the BPVS3 receptive vocabulary measure, this was included as a covariate for most analyses. Data from the teacher questionnaires was intended mainly for descriptive analysis due to small sample size.

12.11 Chapter Summary

The current chapter has described the method for an experimental intervention study comparing the effectiveness of two vocabulary teaching approaches to an age-matched control group. An intervention programme was designed using evidenced principles of effective vocabulary instruction to underpin both approaches, including the same robust teaching structure, materials and protocol, with differences only in the instructional cue types used in vocabulary teaching.

The current chapter set out the research design, participants and sampling process needed to answer the research questions. The assessment battery and procedure were outlined to assess pupils at three time points on outcomes of vocabulary, phonemic awareness and phonic literacy. The design of a Year One whole-class daily vocabulary teaching programme was described in detail. The results of the intervention study for all groups and timepoints will be reported next in Chapter 13.

Chapter 13: Results of the Vocabulary Intervention Study

13.1 Chapter Overview

Chapter 13 presents the results of the experimental vocabulary intervention assessing the effectiveness of classroom vocabulary instruction for Year One mainstream pupils. The overarching aim was to examine whether semantic, combined or usual vocabulary instruction would produce the highest gains in vocabulary, phonological awareness and nonword literacy performance. Assessment outcomes are reported according to the following structure. The first section sets out the data analyses and procedures used for checking assumptions. Subsequent sections report outcomes linked to the research questions in three main areas: (1) vocabulary (Taught Vocabulary Definitions, BPVS3 and CELF4 Expressive Vocabulary), (2) phonological awareness (PhAB2 Rhyme, PhAB2 Alliteration and CTOPP2 Elision) and (3) nonword literacy (PhAB2 Nonword Reading and the bespoke Nonword Spelling task). The performance of pupils with high and low starting points was also evaluated. Finally, the results of the teacher questionnaire provide supplementary data on the perceived impacts of the programme.

13.2 Statistical Analyses

All statistical analyses in this chapter were conducted on raw scores using SPSS version 25 and 26. The study employed ANOVA and ANCOVA to answer the experimental research questions posed in Chapter 11. Using mixed model analyses, the within-subjects variable was *time*, with three levels (pretest, post-test, maintenance test), and the between-subjects variable was *group*, also incorporating three levels (semantic group, combined group, waiting controls). The main result of interest was the interaction between *time* and *group*, demonstrating potential differences between instructional conditions over the three testing points. Main effects of time and group are also reported. Any significant effects were followed up with post hoc pairwise comparisons at each timepoint, adjusted with a Bonferroni correction.

The pretest BPVS3 raw score was used as a covariate in analyses where applicable, as a result of a previously reported oneway ANOVA in Chapter 12 section 12.4.4.7, highlighting that the semantic group had significantly higher initial receptive vocabulary than the combined group. The T1 BPVS3 covariate was therefore used to partial out the variance attributed to this pre-existing difference and to consequently improve power to detect any significant differences in outcomes between the three intervention groups at each timepoint (Huitema, 2011; Leppink, 2018).

Effect sizes showing the magnitude of impact are included to aid interpretation of the educational importance of the findings. Partial eta squared (η_p^2) is reported for the interaction term using parameters by Cohen (1988) of small=.01, medium=.06 and large=.14 effect. Cohen's *d* effect size compared group outcomes at a specific timepoint using conventions of small=0.2, medium=0.5 and large=0.8, signifying gains of 0.2, 0.5 and 0.8 SD respectively. Cohen's *d* calculations were based on the following formula: the difference in group means divided by the pooled standard deviation (ibid).

Prior to all analyses, procedures were followed for data screening and normality checks. Normal distributions were assessed through visual inspection of histograms and Q-Q plots, as well as through application of stringent skewness values between -1 and +1 and absolute kurtosis values below 7 (Kim, 2013). Formal tests were not used to assess departure from normality, since measures such as Kolmogorov-Smironov and Shapiro-Wilk are known to be unreliable with large sample size (Field, 2013; Yazici & Yolacan, 2007), defined by some researchers as 30 or more cases (Field, 2013; Pallant, 2013). The dataset was inspected for extreme outliers using a deliberately wide threshold of three times the inter-quartile range (Field, 2013) to incorporate the full range of pupil ability experiencing the inclusive teaching programme. The decision whether to remove extreme outliers was made on a case-by-case basis.

ANOVA assumptions were tested for each variable. The Greenhouse-Geisser correction was applied if Mauchly's test indicated a violation of the sphericity assumption. The Levene's statistic was initially consulted for homogeneity of variance at all timepoints, although analysis could not rely on this sole indicator, since the statistic is often inflated in large samples (Field, 2013). Equally, ANOVA is robust to this assumption if group sizes are similar (Field, 2013; Zimmerman, 2004) as in the current experiment. If the Levene's test was significant, this was followed by a calculation of the *variance ratio*, dividing the largest group variance by the smallest to ensure that the result was less than three (Jaccard, 1998).

ANCOVA analysis carries the additional requirement to examine covariate assumptions. Linearity was determined through visual inspection of scatterplots between the T1 BPVS3 covariate and the outcome variable at each timepoint. Homogeneity of regression slopes (HRS) between timepoints was established through scrutiny of an ANCOVA comparing the outcome x covariate interaction for all groups, followed by visual inspection of grouped scatterplots. In the case of a significant interaction term (indicating unequal influence of the covariate across groups), separate ANCOVAs were conducted on each pair of slopes at the timepoint in question (Huitema, 2011; Johnson, 2016). The interaction term for slope pairs and the original scatterplot was consulted to check HRS between the pairs before deciding whether to retain the covariate in the model.

13.3 Vocabulary Outcomes

Three variables provide relevant information on vocabulary outcomes: the Taught Definitions task, BPVS3 and CELF4.

13.3.1 Taught Definitions Results

<u>Experimental RQ 1</u>: *Which teaching approach (combined, semantic, control) improves target vocabulary most in 5-6 year olds?*

13.3.1.1 Taught Definitions Results by Experimental Group

Means (*M*), means adjusted for the T1 BPVS3 covariate (M_{adj}), standard deviations (*SD*), minimum scores (*Min*) and maximum scores (*Max*) for the taught definitions task are presented in Table 13.1 for each teaching condition.

Table 13.1

Taught Vocabulary Definitions by Group (out of 42 Points)

		N	*M*	M_{adj}	*SD*	*Min*	*Max*
Semantic Intervention	T1	89	6.88	6.24	4.819	0	26
	T2	90	18.90	17.72	8.276	0	33
	T3	88	21.75	20.45	7.947	2	36
Combined Intervention	T1	95	6.18	6.63	3.989	0	18
	T2	95	16.52	17.56	7.277	0	31
	T3	94	23.53	24.56	7.243	4	37
Waiting Control	T1	80	5.21	5.32	4.741	0	18
	T2	81	10.22	10.39	6.099	0	25
	T3	81	11.52	11.68	7.267	0	30

The standard deviation across all groups is lower at T1 than at the other timepoints, potentially influenced by pretest floor effects for some pupils - a commonly identified reason for lower variance (Lewis-Beck et al., 2004). Vocabulary items were selected to have a higher AoA rating than the chronological age of the sample to prevent ceiling effects at the end of intervention, causing many words to be unknown prior to intervention.

A mixed ANCOVA adjusting for pretest BPVS3 scores was conducted to examine the effects of the three groups over time on the definitions task. The data followed a normal distribution, with no extreme outliers. Levene's test indicated homogeneity of variance at all timepoints - at T1, $F(2,259)=.191$, $p=.83$; at T2, $F(2,259)=2.847$, $p=.06$; at T3, $F(2,259)=.174$, $p=.84$.

In terms of the covariate assumptions, a linear relationship was found between the T1 BPVS3 covariate and taught definitions at each timepoint. The HRS assumption was not met at T1, $F(2,258)=3.058$, $p=.05$ or at T2, $F(2,260)=3.451$, $p=.03$, but it was fully met at T3, $F(2,257)=.716$, $p=.49$. To clarify which regression slopes were unequal, ANCOVAs were performed on each pair of groups at T1 and at T2, as reported in Table 13.2.

Table 13.2

T1 and T2 Homogeneity of Regression Slopes Paired Results - Taught Definitions Task

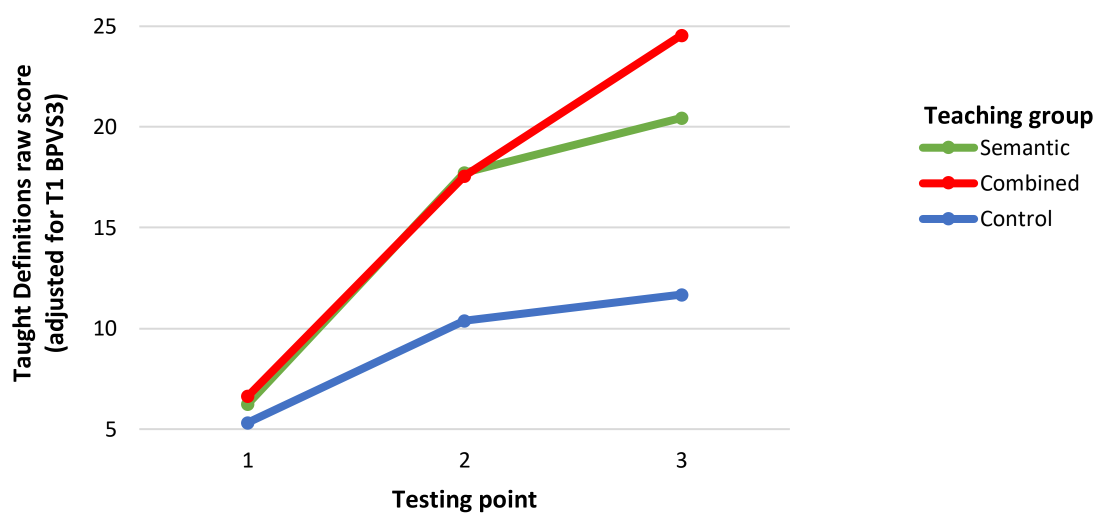
Intervention Groups	*T1 Result*	*T2 Result*
Semantic-Control	$F(1,165)=1.198$, $p=.28$, $\eta_p^2=.007$	F $(1,167)=6.735$, $p=.01$, $\eta_p^2=.04$*
Combined-Control	$F(1,171)=1.858$, $p=.18$, $\eta_p^2=.01$	$F(1,172)=.273$, $p=.60$, $\eta_p^2=.002$
Combined-Semantic	$F(1,180)=6.360$, $p=.01$, $\eta_p^2=.03$ *	$F(1,181)=3.535$, $p=.06$, $\eta_p^2=.02$

Note. * Significant at $p=.05$.

The assumption was met in two out of three pairs at each timepoint. Scatterplots indicated a positive relationship between BPVS3 and definitions in all groups with some differences in slope. Given that four out of six pairs met the assumption, together with the superior ability of ANCOVA to detect group differences (Vickers & Altman, 2001), the covariate was retained in the model. Cautious interpretation will be undertaken to account for the unequal regression slopes at T1 and T2, although confidence can be placed in the T3 results. Definitions results are illustrated in Figure 13.1.

Figure 13.1

Taught Definitions ANCOVA Outcomes by Group



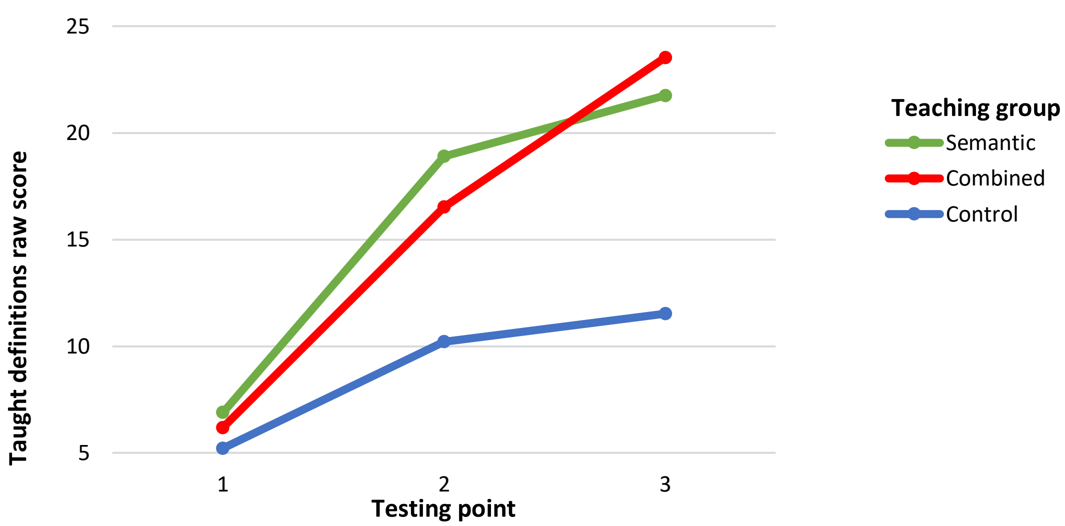
The ANCOVA detected a statistically significant time x group interaction, $F(4,516)=60.032$, $p<.001$, $\eta_p^2=.32$, large effect size (sphericity assumed in all ANCOVA results in this section). Post hoc pairwise comparisons with a Bonferroni correction indicated no significant group differences in definitions performance at T1. At T2, the combined and semantic interventions displayed comparable impact, with both groups significantly higher than controls: combined intervention (p<.001, d=1.06, large effect) and semantic intervention (p<.001, d=1.01, large effect). Cautious interpretation is warranted at these timepoints due to the HRS violation. Greater trust can be placed in the T3 result, when the HRS assumption was fully met. At T3, the combined teaching group knew significantly more taught vocabulary than the control group (p<.001, d=1.78, very large effect), and the semantic group also knew significantly more than controls (p<.001, d=1.15, large effect). Comparing the impact of the two intervention groups at T3, the combined group resulted in significantly higher performance than the semantic group (p<.001, d=.54, medium effect).

The main effect of time was not significant, $F(2,516)=1.384$, $p=.25$; $\eta_p^2=.005$. There was a significant main effect of group, $F(2,258)=67.615$, $p<.001$; $\eta_p^2=.34$. Post hoc pairwise comparisons with Bonferroni correction indicated that the combined group and semantic group performed significantly higher than controls (p<.001 in each case), but the two intervention groups did not differ significantly from each other (p=.07).

As a precaution and for comparison purposes, the analysis was also run as an ANOVA without the covariate (results depicted in Figure 13.2).

Figure 13.2

Taught Definitions ANOVA Outcomes by Group



Homogeneity of variance was met for T1, $F(2,259)=1.022$, $p=.36$ and for T3, $F(2,259)=.471$, $p=.63$ but not at T2, $F(2,259)=3.196$, $p=.04$. However, the variance ratio of 1.3 was well within tolerance, so the assumption was deemed to be met.

Results at T1 and T2 were equivalent to the previously reported ANCOVA. There was a significant time x group interaction, $F(3.766, 487.736)=52.571$, $p<.001$; $\eta_p^2=.29$, large effect size (sphericity not assumed). As in the ANCOVA model, post hoc pairwise comparisons with Bonferroni correction revealed (1) no significant group differences at T1 ($p<.001$), (2) significantly better performance of both intervention groups over controls at T2 ($p<.001$ in each case) and (3) no significant difference between intervention groups ($p=.09$). The results of the two models diverged at T3 however, when both intervention groups were significantly better than controls ($p<.001$ in each case), but the ANOVA did *not* find a significant difference between the combined and semantic groups ($p=.33$), which had previously been detected by the ANCOVA ($p<.001$).

The fact that the two models resonate at T1 and T2 signifies that the HRS violation did not greatly affect the ANCOVA results at these timepoints. The T3 ANCOVA result, which met all assumptions, showed that the combined group achieved significantly higher results than the other two groups.

13.3.1.2 Taught Definitions Results by Class

Analysis of class performance on the taught definitions task can illuminate the extent to which class-level results reflect the ANCOVA outcomes for the intervention groups. Table 13.3 depicts the class means (M), means adjusted for the T1 BPVS3 covariate (M_{adj}) and standard deviations (SD) for the taught definitions probe.

Table 13.3

Taught Vocabulary Definitions (out of 42 Points) by Class

Class+ School		*Group*	*SES*	*N*	T1			T2			T3		
					M	M_{adj}	*SD*	*M*	M_{adj}	*SD*	*M*	M_{adj}	*SD*
1	A	Comb	High	23	6.91	6.69	4.327	18.13	17.68	6.455	25.04	24.56	4.857
2	A	Comb	High	24	6.67	6.20	3.460	20.71	19.79	6.161	28.12	27.13	7.055
3	B	Sem	High	20	6.85	5.87	4.511	22.90	20.98	7.276	24.80	22.73	6.971
4	C	Sem	High	13	3.85	3.05	3.105	17.69	16.14	8.798	19.00	17.33	9.301
5	C	Sem	High	17	7.00	6.17	3.889	19.41	17.79	4.287	21.29	19.55	4.714
6	D	Sem	High*	19	10.47	8.87	5.399	21.00	17.85	8.353	24.63	21.25	7.826
7	E	Comb	Low	22	6.55	7.75	4.688	14.23	16.58	7.124	20.64	23.17	8.209
8	F	Sem	Low	19	5.16	6.44	4.400	13.16	15.68	7.596	17.95	20.66	8.676
9	G	Comb	Low	25	4.60	5.89	3.266	13.36	15.89	7.082	20.28	22.99	5.748
10	H	Con	High	40	7.25	6.64	4.913	11.73	10.52	5.905	14.48	13.18	7.802
11	I	Con	Low	40	3.17	3.99	3.587	8.63	10.23	6.011	8.43	10.15	6.214

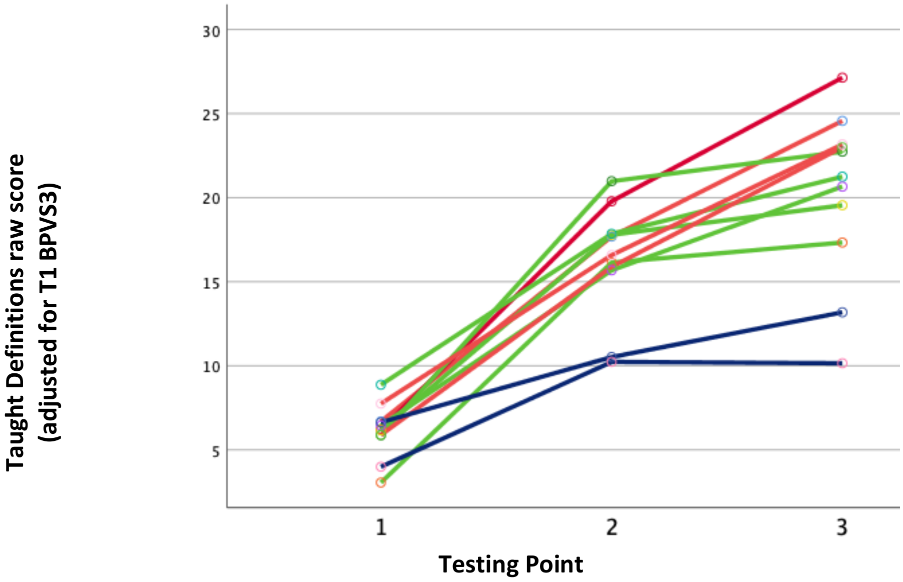
Note. * School D was randomised as a low-SES school based on the IDACI index; later investigation of pupil postcodes determined that this was a high-SES school (see Chapter 6). Comb=combined, Sem=semantic, Con=control.

The low T1 mean score for Class 4 (*M*=3.85) prompted a further accuracy check, since scores in the partner class 5 were nearly twice as high, and not in line with expectations for a low-SES school. It appears that the low mean in Class 4 was influenced by four pupils scoring either zero or one point. No reasons were apparent for the high T1 performance in Class 5. Further enquiry confirmed that testing was performed in line with protocol, so the scores were considered valid.

After checking assumptions, a mixed ANCOVA was conducted for the definitions task with 11 classes at each timepoint. Data were normally distributed, with no extreme outliers. Levene's test confirmed homogeneity of variance at T1, *F*(10,251)=1.051, *p*=.40 and at T3, *F*(10,251)=.595, *p*=.82, but not at T2, *F*(210,251)=2.021, *p*=.03. This was accepted, however, as the variance ratio of 2 was within limits. In terms of the covariate assumptions, linearity was achieved at each timepoint. Homogeneity of regression slopes was met between classes at all timepoints - at T1, *F*(10,242)=1.225, *p*=.28, η_p^2=.05; at T2, *F*(10,244)=.893, *p*=.54, η_p^2=.04; at T3, *F*(10,241)=1.788, *p*=.06, η_p^2=.07. Figure 13.3 depicting class-level outcomes has been colour-coded to show the three teaching conditions.

Figure 13.3

Taught Definitions ANCOVA Outcomes by Class



Note. Red=classes in combined group, green=classes in semantic group, blue=classes in control group.

The ANCOVA detected a significant time x class interaction, $F(20,500)=15.163$, $p<.001$, $\eta_p^2=.38$, large effect size (sphericity assumed in all results in this section). Post hoc comparisons with Bonferroni correction largely confirmed results of the previous ANCOVA in Figure 13.1. At T2, the intervention classes performed significantly better than control classes ($p<.001$ to $p=.03$), but there were no significant differences *between* intervention classes ($p=.17$ to $p=1.00$). The exception was class 4 (discussed earlier in this section), which had the lowest pretest mean and did not outperform control classes ($p=.06$ and $p=.09$). At T3, all intervention classes surpassed controls ($p<.001$ in each case). Descriptively, at T3 the four classes in the combined group had adjusted raw scores ($M_{adj}=22.99$ to 27.13) higher than the four classes in the semantic group ($M_{adj}=17.33$ to 22.73).

There was no significant main effect of time, $F(2,500)=.802$, $p=.45$, $\eta_p^2=.003$. A significant main effect of class was found, $F(10,250)=16.129$, $p<.001$, $\eta_p^2=.39$. Post hoc comparisons with Bonferroni correction showed that the main difference lay between the intervention classes and controls ($p<.001$ to $p=.03$), although Class 1 (combined) also significantly outperformed class 4 (semantic), $p=.007$.

13.3.2 Standardised Vocabulary Outcomes

Experimental RQ 2: *Which teaching approach (combined, semantic, control) improves standardised vocabulary most in 5-6 year olds?*

13.3.2.1 BPVS3 (receptive vocabulary) Results by Experimental Group

Means (*M*), standard deviations (*SD*), minimum scores (*Min*) and maximum scores (*Max*) for the BPVS3 variable are presented in Table 13.4.

Table 13.4

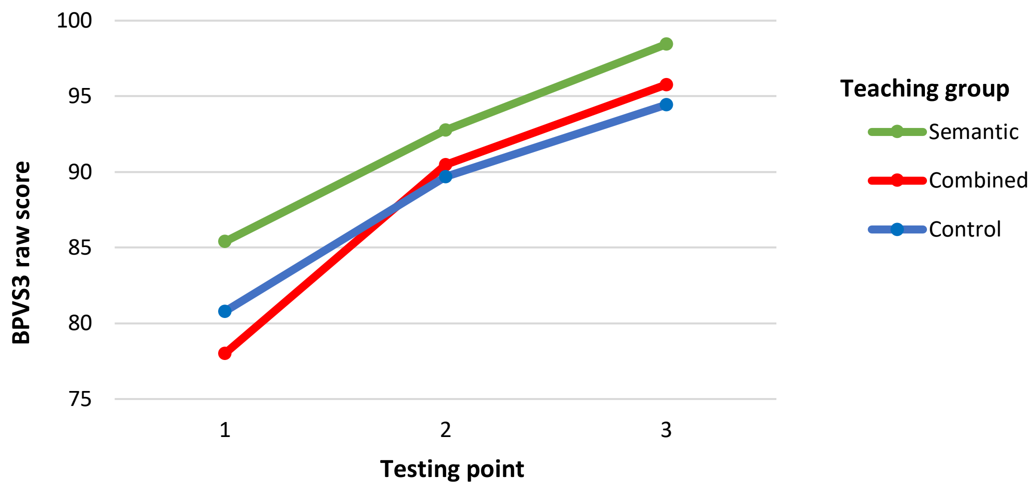
BPVS3 Outcomes by Group

		N	*M*	*SD*	*Min*	*Max*
Semantic Intervention	T1	90	85.40	14.865	46	119
	T2	90	92.77	12.965	56	120
	T3	89	98.44	14.771	57	132
Combined Intervention	T1	95	78.01	15.031	28	106
	T2	95	90.49	10.053	47	108
	T3	94	95.77	12.278	51	134
Waiting Control	T1	81	80.79	16.046	28	108
	T2	80	89.70	14.136	36	115
	T3	81	94.44	15.364	23	133

Table 13.4 reveals the significant group difference on the BPVS3 at T1 that was highlighted in Chapter 12. Since the covariate was also the outcome variable, it was not possible to adjust for the T1 BPVS3 variable. A mixed ANOVA was therefore performed, with results depicted in Figure 13.4.

Figure 13.4

BPVS3 ANOVA Outcomes by Group



BPVS3 data were normally distributed at all timepoints. One extreme outlier was evident at all testing points. This control group pupil had severe special needs but was not a statistical outlier on all measures. To check the impact of the outlier, the analysis was also run excluding this case, with negligible difference in results. The outlier was therefore retained in the analysis to capture the full ability range. The Levene's statistic suggested homogeneity of error variances at T1, $F(2,260)=.320$, $p=.73$ and at T3, $F(2,260)=2.330$, $p=.10$ but not at T2, $F(2,260)=4.407$, $p=.02$, although the variance ratio of 1.4 was well within acceptable levels.

The mixed ANOVA resulted in a significant time x group interaction, $F(3.858,501.553)=3.850$, $p=.005$, $\eta_p^2=.03$, small effect size (sphericity not assumed), although post hoc comparisons with Bonferroni correction revealed only the aforementioned significant pretest difference on the BPVS3. A significant main effect of time was detected, $F(2,520)=300.146$, $p<.001$, $\eta_p^2=.54$. Post hoc comparisons with Bonferroni correction disclosed significant improvement at each timepoint ($p<.001$ for each). The effect of group was not significant, $F(2,260)=2.675$, $p=.07$, $\eta_p^2=.02$.

13.3.2.2 CELF4 (Expressive Vocabulary) Results by Experimental Group

Means (*M*), means adjusted for T1 BPVS3 (M_{adj}), standard deviations (*SD*), minimum scores (*Min*) and maximum scores (*Max*) for the CELF4 variable are presented in Table 13.5.

Table 13.5

CELF4 Expressive Vocabulary Outcomes (out of 54 Points) by Group

		N	*M*	M_{adj}	*SD*	*Min*	*Max*
Semantic Intervention	T1	90	23.48	21.57	9.400	0	45
	T2	90	28.34	26.47	9.170	4	47
	T3	89	32.49	30.70	9.692	4	50
Combined Intervention	T1	95	21.11	21.25	8.495	2	37
	T2	94	28.41	28.51	7.467	11	47
	T3	94	31.03	31.07	7.413	7	48
Waiting Control	T1	81	20.11	20.11	8.073	2	39
	T2	81	25.21	25.21	8.699	3	50
	T3	81	28.89	28.89	8.485	2	48

Using T1 BPVS3 scores as the covariate, a mixed ANCOVA was conducted to examine the effects of the three different teaching approaches over time on the CELF4. Data were normally distributed at all timepoints with no outliers. The Levene's statistic demonstrated homogeneity of variance at all timepoints - at T1, $F(2,260)=.310$, $p=.73$; at T2, $F(2,260)=1.553$, $p=.21$; at T3, $F(2,260)=.097$, $p=.91$.

Turning to the covariate assumptions, there was a linear relationship between the T1 BPVS3 covariate and CELF4 Expressive Vocabulary at each testing point. The homogeneity of regression slopes assumption was not met for T1, $F(2,260)=5.067$, $p=.007$ or for T3, $F(2,258)=6.166$, $p=.002$; it was met for T2, $F(2,259)=2.491$, $p=.09$. Separate exploratory ANCOVAs were therefore conducted on pairs of groups at each timepoint revealing the results in Table 13.6 below.

Table 13.6

T1 and T3 Homogeneity of Regression Slopes Paired Results - CELF4

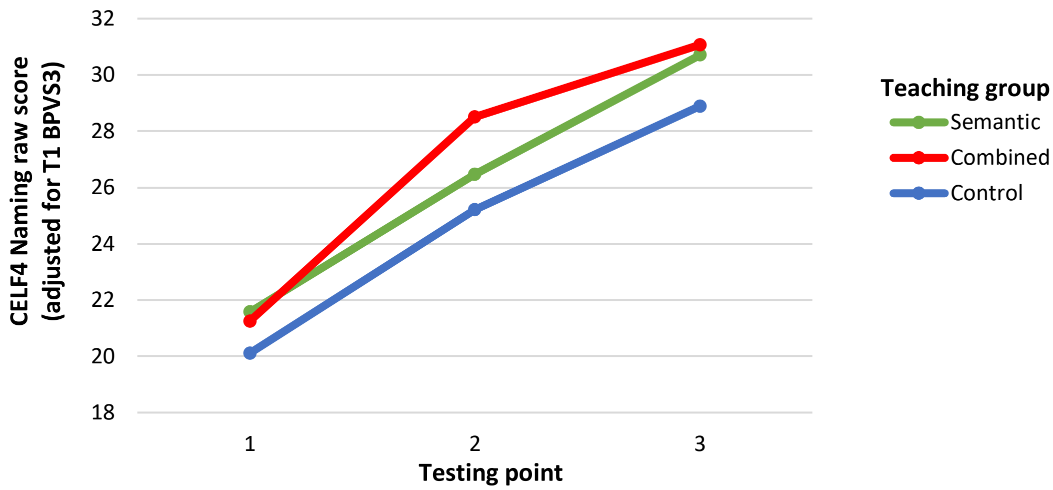
Intervention Groups	*T1 Result*	*T3 Result*
Semantic-Control	$F(1,167)=9.587$, $p=.002$, $\eta_p^2=.05$*	$F\ (1,166)=5.336$, $p=.02$, $\eta_p^2=.03$*
Combined-Control	$F(1,172)=.115$, $p=.74$, $\eta_p^2=.001$	$F(1,171)=1.278$, $p=.26$, $\eta_p^2=.007$
Combined-Semantic	$F(1,181)=6.634$, $p=.01$, $\eta_p^2=.=.04$*	$F(1,179)=12.064$, $p=.001$, $\eta_p^2=.06$*

Note. * Significant at $p=.05$.

A review of the scatterplots suggested that despite four pairs violating the HRS assumption, each pair had a positive relationship with some differences in slope. The ability of ANCOVA to account for the T1 BPVS3 difference on the CELF4 results favoured retention of the covariate. Figure 13.5 displays adjusted CELF4 results over time in each teaching group.

Figure 13.5

CELF4 ANCOVA Outcomes by Group



No significant time x group interaction was found, $F(4,518) = 2.034$, $p=.09$; $\eta_p^2=.02$ (sphericity assumed in all results in this section). There was a significant effect of time, $F(2,518)= 17.633$, $p<.001$, $\eta_p^2=.06$. Post hoc tests with Bonferroni correction revealed a significant increase at all timepoints ($p<.001$ in each case). There was also a significant effect of group, $F(2,257)=6.763$, $p=.001$, $\eta_p^2=.05$. According to post hoc pairwise comparisons, the combined group had significantly higher adjusted scores compared to controls ($p<.001$). The semantic group did not significantly differ from the combined group ($p=.09$) or controls ($p=.15$).

13.4 Phonological Awareness Outcomes

Three assessments are relevant to these outcomes: PhAB2 Rhyme, PhAB2 Alliteration and CTOPP2 Elision.

13.4.1 PhAB2 Rhyme Results by Experimental Group

<u>*Experimental RQ 3*</u>*: Which teaching approach (combined, semantic, control) improves rhyme most in 5-6 year olds?*

Means (*M*), standard deviations (*SD*), minimum scores (*Min*) and maximum scores (*Max*) of the original rhyme variable are presented in Table 13.7, followed by transformed means (M^T) and transformed means adjusted for the T1 BPVS3 covariate (M^T_{adj}).

Table 13.7

PhAB2 Rhyme Outcomes (out of 10 Points) by Group

		N	*M*	*SD*	*Min*	*Max*	M^T	M^T_{adj}
Semantic	T1	90	7.10	2.922	0	10	1.80	1.66
Intervention	T2	90	8.51	2.040	1	10	2.35	2.23
	T3	89	8.89	2.145	0	10	2.60	2.49
Combined	T1	95	6.06	3.222	0	10	1.42	1.54
Intervention	T2	95	7.77	2.623	1	10	2.07	2.18
	T3	94	8.67	1.774	1	10	2.36	2.44
Waiting	T1	81	6.42	2.765	0	10	1.49	1.51
Control	T2	81	7.41	2.919	0	10	1.92	1.94
	T3	81	7.78	2.559	1	10	2.06	2.08

Note. M^T = transformed mean (square root transformation), M^T_{adj} = transformed mean adjusted for the T1 BPVS3 covariate.

Data for the rhyme task were normally distributed at T1 but displayed a negative skew greater than one at timepoints T2 and T3, due to anticipated ceiling effects. The rhyme variable was therefore transformed at all timepoints to improve normality of the distribution. Square root transformation was chosen over Log10 as it had a superior effect on reducing the skew as well as improving homogeneity of variance. Square root transformation was used consistently across skewed variables (rhyme and alliteration) at all three time points to enable comparison using the same unit scale.

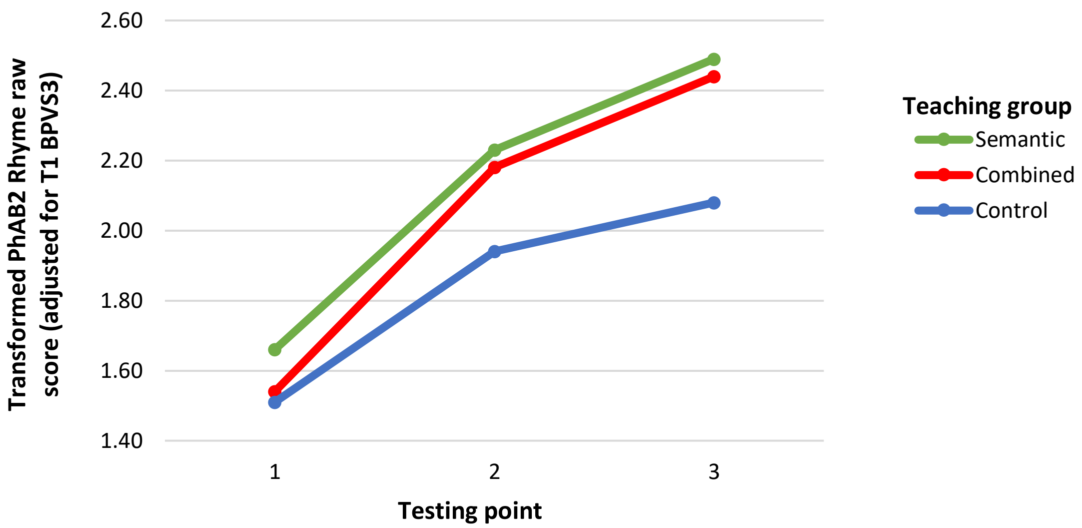
A mixed ANCOVA adjusted for pretest BPVS3 scores was carried out to examine the impact of the three teaching groups on rhyme over time. Normality for the transformed variables was within parameters at all timepoints, although visual inspection of histograms suggested continued skew. There were two extreme outliers (based on untransformed means) at the final testing point, scoring 0 and 1. These pupils with significant special needs were considered to provide important data to inform programme efficacy, and since excluding these datapoints made negligible difference to results, their scores were retained. Levene's test indicated homogeneity of variance at T1,

F(2,261)=1.014, p=.36 and at T2, F(2,261)=1.427, p=.24 but not at T3, F(2,261)=4.185, p=.02. Since the variance ratio of 1.2 was well within tolerance, the assumption was considered to be met.

In terms of the covariate assumptions, there was a linear relationship between the T1 BPVS3 covariate and the transformed rhyme variable at each timepoint. The HRS assumption was met at all timepoints - at T1, F(2,260)=1.793, p=.17; at T2, F(2,260)=.015, p=.99; at T3, F(2,258)=.772, p=.46. Rhyme results are shown in Figure 13.6.

Figure 13.6

Transformed PhAB2 Rhyme ANCOVA Outcomes by Group



The ANCOVA did not display a significant time x group interaction, F(3.737,485.875)=1.837, p=.13, η_p^2=.01 (sphericity not assumed for all results in this section). The main effect of time was significant, F(1.869, 485.875)=11.490, p<.001; η_p^2= .04; post hoc comparisons with Bonferroni correction indicated significant differences at each timepoint (p<.001 in each case). There was also a significant effect of group, F(2,260)=3.992, p=.02; η_p^2=.03; post hoc pairwise comparisons with Bonferroni correction showed a significant advantage of the semantic group over controls (p=.02) but no significant differences between the combined group and controls (p=.12), nor between the intervention groups (p=1.00).

13.4.2 PhAB2 Alliteration Results by Experimental Group

Experimental RQ 4: *Which teaching approach (combined, semantic, control) improves alliteration most in 5-6 year olds?*

Means (*M*), standard deviations (*SD*), minimum scores (*Min*) and maximum scores (*Max*) of the original alliteration variable are presented in Table 13.8, followed by transformed means (M^T) and transformed means adjusted for the T1 BPVS3 covariate (M^T_{adj}).

Table 13.8

PhAB2 Alliteration Outcomes (out of 10 Points) by Group

		N	*M*	*SD*	*Min*	*Max*	M^T	M^T_{adj}
Semantic	T1	90	5.99	3.143	0	10	1.40	1.27
Intervention	T2	90	8.27	2.217	0	10	2.19	2.07
	T3	89	8.74	2.135	1	10	2.32	2.21
Combined	T1	95	5.16	3.190	0	10	1.11	1.21
Intervention	T2	95	8.03	2.195	0	10	2.06	2.16
	T3	94	8.76	1.683	2	10	2.24	2.33
Waiting	T1	81	5.83	2.919	0	10	1.34	1.35
Control	T2	81	7.44	2.612	0	10	1.87	1.88
	T3	81	7.81	2.545	1	10	1.89	1.90

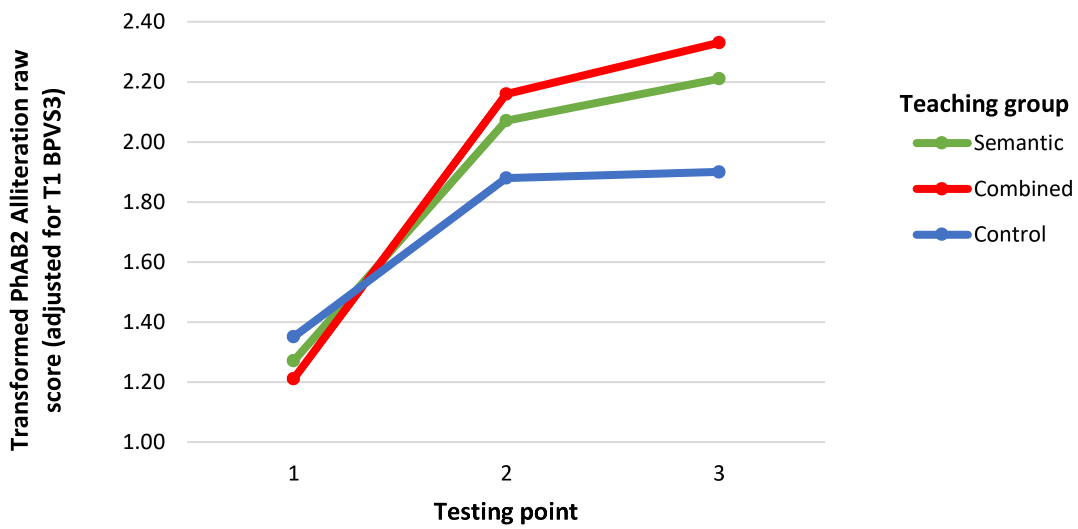
Note. M^T = transformed mean (square root transformation), M^T_{adj} = transformed mean adjusted for the T1 BPVS3 covariate.

Alliteration scores were normally distributed at T1 but not at T2 or T3, due to ceiling effects. The alliteration variable was transformed at all timepoints using a square root transformation. A mixed ANCOVA was conducted on the transformed alliteration variable at the three testing points, adjusted for pretest BPVS3 scores. After transformation, skewness values fell within tolerance, although histograms indicated a continued skew, particularly at T3. There was just one extreme outlier (based on untransformed means) - the same outlier as the rhyme variable, and once again the data for this participant was deemed appropriate to maintain in the dataset. Homogeneity of variance was achieved at all timepoints - T1, $F(2,261)=1.407$, $p=.25$; T2, $F(2,261)=1.725$, $p=.18$; at T3, $F(2,261)=2.383$, $p=.09$.

In terms of the covariate assumptions, there was a linear relationship between the T1 BPVS3 covariate and the transformed alliteration variable at each timepoint. The HRS assumption was met at all timepoints – at T1, $F(2,260)=2.330$, $p=.10$; at T2, $F(2,260)=.813$, $p=.45$; at T3, $F(2,258)=.242$, $p=.79$. Figure 13.7 depicts results of the ANCOVA for the transformed alliteration variable.

Figure 13.7

Transformed PhAB2 Alliteration ANCOVA Outcomes by Group



The ANCOVA detected a significant time x group interaction, $F(4,520)=5.449$, $p<.001$, $\eta_p^2=.04$, small effect size (sphericity assumed for all results in this section). Post hoc analysis with Bonferroni corrections found no group differences at T1 or T2. At T3, a significant difference emerged between the semantic and control groups ($p=.04$, $d=0.3$, small effect size) and between the combined and control groups ($p=.001$, $d=0.5$, medium effect size); however, no difference was observed between intervention groups ($p=.93$).

There was a significant main effect of time, $F(2,520)=9.212$, $p<.001$, $\eta_p^2=.03$. Post hoc comparisons with a Bonferroni correction indicated a significant difference from T1 to T2 ($p<.001$) and from T1 to T3 ($p<.001$) but not from T2 to T3 ($p=.10$). The effect of group was not significant, $F(2,260)=2.102$, $p=.12$, $\eta_p^2=.02$.

13.4.3 CTOPP2 Elision Results by Experimental Group

<u>Experimental RQ 5</u>: *Which teaching approach (combined, semantic, control) improves elision most in 5-6 year olds?*

Means (*M*), means adjusted for T1 BPVS3 (M_{adj}), standard deviations (*SD*), minimum scores (*Min*) and maximum scores (*Max*) for the elision variable are presented in Table 13.9.

Table 13.9

CTOPP2 Elision Outcomes (out of 34 Points) by Group

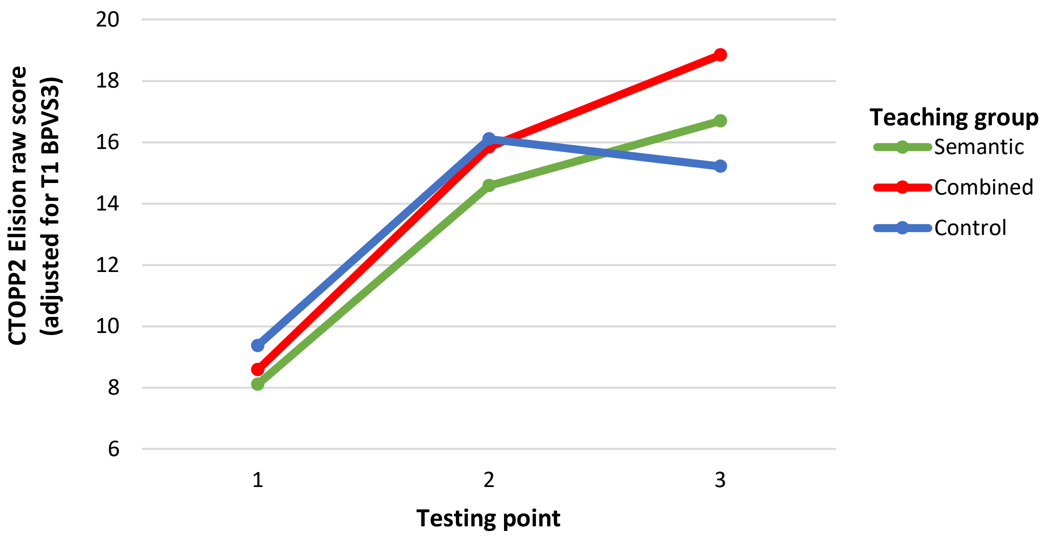
		N	*M*	M_{adj}	*SD*	*Min*	*Max*
Semantic Intervention	T1	90	8.94	8.10	6.150	0	32
	T2	90	15.31	14.57	7.135	0	33
	T3	89	17.45	16.69	7.750	0	33
Combined Intervention	T1	95	7.92	8.57	5.748	0	28
	T2	95	15.16	15.83	7.197	0	29
	T3	94	18.22	18.84	6.386	0	33
Waiting Control	T1	81	9.26	9.37	6.121	0	24
	T2	81	15.99	16.10	7.955	0	32
	T3	81	15.09	15.20	8.543	0	32

A mixed ANCOVA was conducted on the elision variable at T1, T2 and T3, adjusted for pretest BPVS3 scores. Data were normally distributed at all timepoints and no extreme outliers were detected. Homogeneity of variance was achieved at all timepoints – T1, $F(2,261)=.808$, $p=.45$; T2, $F(2,261)=1.024$, $p=.36$; T3, $F(2,261)=2.542$, $p=.08$.

In terms of the covariate assumptions, there was a linear relationship between the T1 BPVS3 covariate and the transformed alliteration variable at each timepoint. The HRS assumption was also met at all timepoints: at T1, $F(2,260)=1.401$, $p=.25$; at T2, $F(2,260)=.416$, $p=.66$; and at T3, $F(2,258)=1.082$, $p=.34$. Figure 13.8 depicts results of the ANCOVA for the elision task.

Figure 13.8

CTOPP2 Elision ANCOVA Outcomes by Group



The ANCOVA detected a significant time x group interaction, $F(3.695, 480.337)=6.951$, $p<.001$, $\eta_p^2=.05$, medium effect size (sphericity not assumed for all results in this section). Post hoc analysis with Bonferroni correction found no group differences at T1 or T2. At T3, a significant difference emerged between the combined group and controls ($p=.002$, $d=0.5$, medium effect size); no significant difference was detected between the semantic group and controls ($p=.51$) or between the two intervention groups ($p=.12$).

There was a significant main effect of time, $F(1.847, 480.337)=7.867$, $p<.001$, $\eta_p^2=.03$. Post hoc comparisons with Bonferroni correction indicated significant differences between all timepoints ($p<.001$ for each pair). The effect of group was not statistically significant, $F(2,260)=1.374$, $p=.26$, $\eta_p^2=.01$.

13.5 Nonword Literacy Outcomes

Two assessment tasks are relevant to nonword literacy outcomes - the PhAB2 Nonword Reading task and a bespoke Nonword Spelling test.

13.5.1 Nonword Reading Results by Experimental Group

Experimental RQ 6: *Which teaching approach (combined, semantic, control) improves nonword reading most in 5-6 year olds?*

Means (*M*), means adjusted for the T1 BPVS3 covariate (M_{adj}), standard deviations (*SD*), minimum scores (*Min*) and maximum scores (*Max*) for PhAB2 Nonword Reading are presented in Table 13.10.

Table 13.10

PhAB2 Nonword Reading Outcomes (out of 24 Points) by Group

		N	*M*	M_{adj}	*SD*	*Min*	*Max*
Semantic Intervention	T1	87	6.23	5.76	4.574	0	19
	T2	90	13.12	12.53	5.382	0	24
	T3	89	13.47	13.00	6.533	0	24
Combined Intervention	T1	95	5.13	5.51	4.273	0	16
	T2	95	13.06	13.44	4.957	0	24
	T3	94	16.46	16.81	4.830	1	24
Waiting Control	T1	81	6.52	6.57	5.109	0	23
	T2	81	12.81	12.87	6.333	0	24
	T3	81	13.94	14.00	6.731	1	24

A mixed ANCOVA adjusted for pretest BPVS3 scores investigated the effects of the three teaching interventions over time on the nonword reading task. The data displayed a normal distribution, with no extreme outliers. Levene's test indicated homogeneity of variance at T1, $F(2,258)=1.117$, $p=.33$ but not at T2, $F(2,258)=4.152$, $p=.02$ or at T3, $F(2,258)=.3.653$, $p=.03$. The variance ratios of 1.3 at T2 and 1.4 at T3 were well within tolerance, suggesting homogeneity of variance.

In terms of the covariate assumptions, there was a linear relationship between the T1 BPVS3 covariate and the nonword reading variable at each timepoint. The HRS assumption was met at T1, $F(2,257)=.308$ $p=.74$ and at T2, $F(2,260)=.017$, $p=.98$ but not at T3, $F(2,258)=3.514$, $p=.03$. To further understand the unequal regression slopes at T3, ANCOVAs were performed on each pair of groups. Results for the interactions are provided in Table 13.11.

Table 13.11

T3 Homogeneity of Regression Slopes Paired Results - PhAB2 Nonword Reading

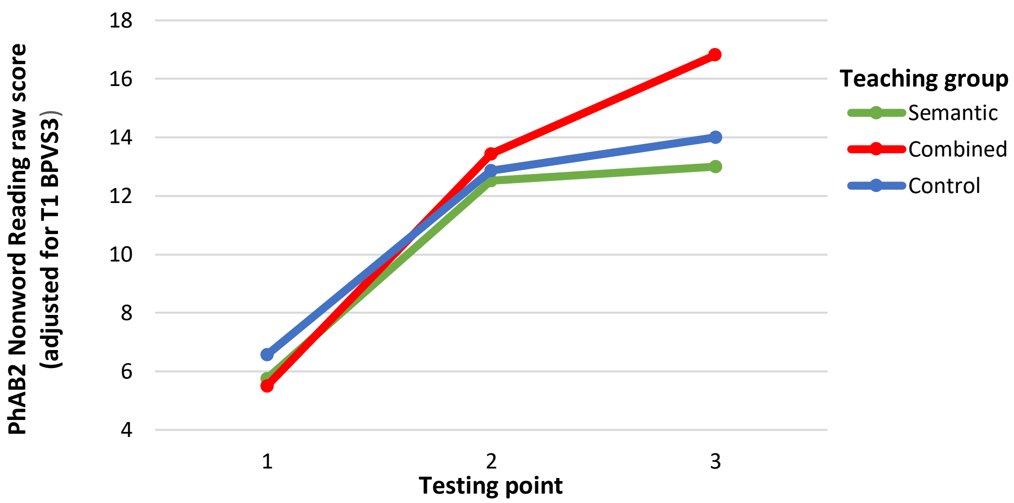
Intervention Groups	*T3 Result*
Semantic-Control	$F(1,166)=.479$, $p=.49$, $\eta_p^2=.003$
Combined-Control	$F(1,171)=3.553$, $p=.06$, $\eta_p^2=.02$
Combined-Semantic	$F(1,179)=7.366$, $p=.007$, $\eta_p^2=.04$*

Note. * Significant at $p=.05$.

At T3, the HRS assumption was met in two out of three pairs. Given the HRS violation in just one pair of slopes at T3, and the benefit of ANCOVA to adjust for the T1 BPVS3 variable, the covariate was retained in the model. Cautious interpretation will be undertaken at T3 to account for the unequal regression slopes between the semantic and combined groups. Figure 13.9 displays the ANCOVA results for nonword reading.

Figure 13.9

PhAB2 Nonword Reading ANCOVA Outcomes by Group



The ANCOVA detected a statistically significant time x group interaction, $F(3.762, 483.387)=11.593$, $p<.001$, $\eta_p^2=.08$, medium effect size (sphericity not assumed in all ANCOVA results in this section). Post hoc pairwise comparisons with a Bonferroni correction indicated no significant group differences at T1 or T2. At T3, the combined intervention performed significantly better than the semantic group ($p<.001$, $d=.67$, medium effect size), although this result should be interpreted cautiously due to the HRS violation at T3. The combined group also scored significantly higher than controls (p=.005, d=.49, medium effect size). There was no significant difference between the semantic and control groups (p=.81).

The main effect of time was significant, F(1.881, 483.387)=17.634, p<.001; η_p^2=.06. Post hoc pairwise comparisons with Bonferroni correction indicated a significant effect at all timepoints (p<.001 in each case). A significant main effect of group was not found, F(2,257)=2.453, p=.09; η_p^2=.02.

13.5.2 Nonword Spelling Results by Experimental Group

<u>Experimental RQ 7</u>: *Which teaching approach (combined, semantic, control) improves nonword spelling most in 5-6 year olds?*

Means (M), means adjusted for the T1 BPVS3 covariate (M_{adj}), standard deviations (SD), minimum scores (Min) and maximum scores (Max) for the nonword spelling task are presented in Table 13.12.

Table 13.12

Nonword Spelling Outcomes (out of 15 Points) by Group

		N	*M*	M_{adj}	*SD*	*Min*	*Max*
Semantic Intervention	T1	90	5.74	5.31	3.283	0	13
	T2	90	10.30	9.97	3.210	0	15
	T3	89	10.40	10.12	2.972	1	15
Combined Intervention	T1	95	5.61	5.95	2.940	0	13
	T2	95	10.33	10.61	2.923	0	15
	T3	94	10.27	10.50	2.988	1	15
Waiting Control	T1	81	5.49	5.55	4.090	0	13
	T2	81	9.30	9.34	3.822	0	15
	T3	81	9.63	9.67	3.265	0	15

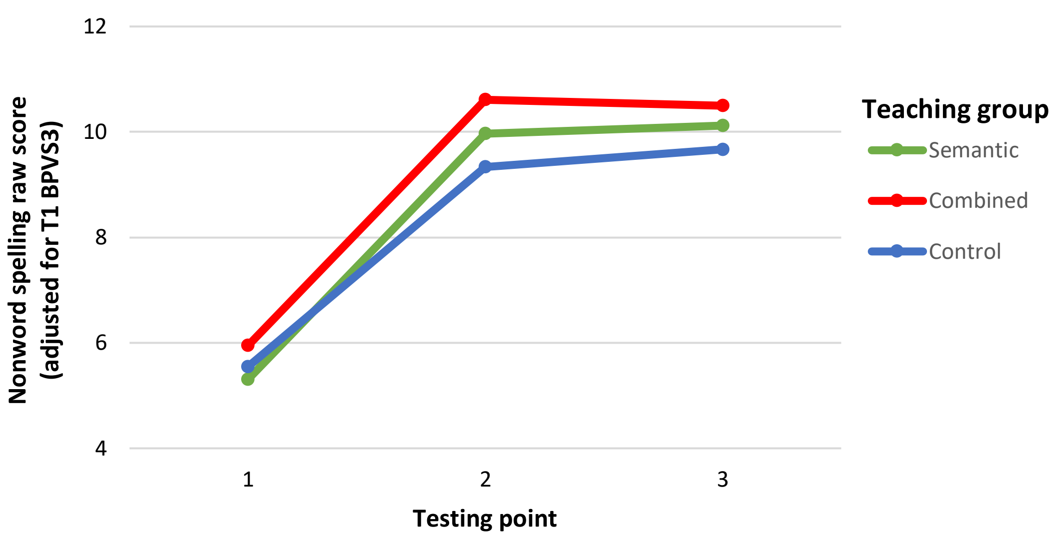
A mixed ANCOVA adjusted for pretest BPVS3 scores was conducted to investigate the effects of the three teaching interventions over time on the nonword spelling task. The data displayed a normal distribution, apart from at T2 when there was a skewness value of -1.025, slightly outside the stated parameters and therefore assumed normal for the purpose of analysis. No extreme outliers were detected at any timepoint. Levene's test indicated a lack of homogeneity of variances at T1,

F(2,261)=5.684, p=.004 and at T2, F(2,261)=3.255, p=.04, but the assumption was met at T3, F(2,261)=.588, p=.56. Homogeneity of variance was assumed, given the variance ratios of 1.4 at T2 and 1.3 at T3, which were well within prescribed limits.

In terms of the covariate assumptions, there was a linear relationship between the T1 BPVS3 covariate and the nonword spelling variable at each timepoint. The HRS assumption was met at all timepoints - at T1, F(2,260)=1.268 p=.28; at T2, F(2,260)=.119, p=.89; at T3, F(2,258)=.071, p=.93. ANCOVA results for nonword spelling are depicted in Figure 13.10.

Figure 13.10

Nonword Spelling ANCOVA Outcomes by Group



The ANCOVA did not result in a statistically significant time x group interaction, F(3.990, 506.998)=1.271, p=.28, η_p^2=.01 (sphericity not assumed in all ANCOVA results in this section). The main effect of time was significant, F(1.950, 506.998)=22.991, p<.001; η_p^2=.08. Post hoc pairwise comparisons with Bonferroni correction indicated a significant effect of time from T1 to T2 and from T1 to T3 (p<.001 in each case) but not from T2 to T3 (p=1.00). The main effect of group was not significant, F(2,260)=2.594, p=.08; η_p^2=.02.

13.6 Inter-Rater Reliability Checks

The Cohen's kappa statistic was used to report the extent to which two testers assigned the same score to items on the PhAB2 Nonword Reading test, since it required fine speech discrimination rather than a clear-cut binary response. The tests were audio-recorded for the purpose of second marking by a different tester. Recording quality was improved on the basis of lessons learned from the cross-sectional study, this time taking place in a quiet space with the handheld device near the child's mouth.

The IRR check was designed such that testers each listened to the audio recordings for eight pupils from another tester. Cases were first randomised for each tester using Random.org (2019) The first eight pupils on the randomised list were selected for the IRR check. One child with severe speech impairment was excluded. Tester 1 listened to tester 2, tester 2 listened to tester 3, tester 3 listened to tester 4, tester 4 listened to tester 6 (tester 5 was not included due to an acknowledged difficulty in perceiving fine phonemic differences) and tester 6 listened to tester 1 to complete the circle. This method provided inter-rater checks for 40 pupils, equating to 15% of the sample.

Part A of the PhAB2 Nonword Reading subtest (14 items) was included in the IRR check but not Part B, since very few children were able to attempt this harder level at T1. Thus, for each pupil, a maximum of 112 items was available for second marking. The number of actual scored responses varied considerably due to the PhAB2 discontinuation rule after four consecutive errors.

Each tester marked the subtests blind to randomisation group by listening to an MP3 file of the recordings through earphones on a laptop in a quiet setting. They were allowed to listen as many times as needed before making a judgement. Table 13.13 shows the inter-rater agreement scores, which fall in the *substantial* to *almost perfect* range.

Table 13.13

Inter-Rater Agreement for PhAB2 Nonword Reading

Number of Scored Responses	*Original Tester*	*IRR Second Marker*	*Cohen's Kappa κ*	*Agreement Rating*	*Percentage Agreement*
97	Tester 2	Tester 1	.78	substantial	89.7%
75	Tester 3	Tester 2	.84	almost perfect	92.0%
86	Tester 4	Tester 3	.79	substantial	89.5%
100	Tester 6	Tester 4	.67	substantial	86.0%
90	Tester 1	Tester 6	.85	almost perfect	92.0%

13.7 Further Analysis: Effects of Vocabulary Teaching on Pupils with Differing Starting Points

13.7.1 Section Overview

A post hoc analysis considered the question of whether response to intervention differed by pupils' starting points. The combined approach showed significantly better results on taught definitions, CTOPP2 Elision and PhAB2 Nonword Reading, so it is useful to know whether this result was uniform across the ability range. Partial correlations adjusting for the effect of the T1 BPVS3 covariate were computed to determine the strength and direction of the relationship between the T1 score and the progress score (T1 to T3) for the three variables in question. A low coefficient would suggest little relationship between starting ability and subsequent progress, whereas a high correlation indicates that not all pupils benefit equally from combined instruction.

Assumptions of linearity, normality and absence of outliers have already been described in the main analyses.

13.7.2 Taught Definitions

The combined group performed significantly higher than both other groups on this variable. A significant negative partial correlation indicated that pupils' initial definitions ability was related to their progress on the taught definitions task, $r_{\text{partial}}(91) = -.297$, $p=.004$, albeit to a weak extent. Pupils with lower starting points made slightly more gains than those with higher initial scores in the combined condition.

13.7.3 Elision

The combined group performed significantly better than controls on the elision task. A significant negative partial correlation of a moderate magnitude was found between T1 elision scores and progress on this variable, $r_{\text{partial}}(91) = -.475$, $p=.001$, indicating that initial elision ability was related to progress, with lower start points making moderately more progress and higher initial ability making moderately less progress.

13.7.4 Nonword Reading

The combined group performed significantly better than both the semantic and control groups on this variable. A significant negative partial correlation of moderate magnitude was discovered between T1 nonword reading and the level of progress, $r_{\text{partial}}(91) = -.407$, $p<.001$. Pupils' initial nonword reading ability was moderately related to their progress, i.e. children in the combined group with lower nonword reading scores made moderately higher gains than better phonic readers.

13.8 Teacher Questionnaire

13.8.1 Overview of the Questionnaire

The results below of the post-intervention questionnaire are based on nine teacher questionnaires - one per intervention class. If two teachers shared a class, the questionnaire was completed by the main contact teacher. Descriptive analysis will be presented, alongside actual comments where numbers are low, or through coding of themes for larger numbers of comments. Coding was completed by the main researcher only. Comparative analyses were performed on relevant variables to monitor any differences attributable to teaching group. Nonparametric tests were chosen due to the small sample size of teachers (combined N=4; semantic N=5). Assumptions were confirmed for each analysis, including visual inspection of histograms to ensure similar distributions between the two teaching groups.

13.8.2 Impact of the Overall Programme

13.8.2.1 Teacher Impact

Teachers rated their confidence to teach vocabulary before and after the intervention. Results depicted in Table 13.14 demonstrate a 55% increase in confidence between the two timepoints. A Wilcoxon signed rank test determined that the increase between pre- and post-intervention ratings was significant, z=-2.549, p=.01.

Table 13.14

How Confident Do You Feel About Teaching New Vocabulary? (Out of 5 Points: 1=Not Confident, 5=Very Confident)

	M	*SD*	*Min*	*Max*
Pre-Intervention (T1)	3.00	.866	2	4
Post-Intervention (T2)	4.67	.500	4	5

Mann-Whitney *U* tests were used to ascertain whether teacher confidence ratings differed between the two intervention groups. At T1, the median confidence scores for the two groups (combined=3.50, semantic=3.00) were not significantly different, U=7.00, z=-.775, p=.56, using an exact sampling distribution for *U* due to small sample size in all analyses in this section (Dineen & Blakesley, 1973). At T2, the medians (combined=4.50, semantic=5.00) were also not significantly different, U=7.00, z=-.894, p=.56. Results thus indicate that confidence in vocabulary teaching ability increased equally in both conditions.

At T2 post-intervention, teachers rated their enjoyment in delivering the vocabulary programme (see Table 13.15), resulting in a high score with little variation between individuals.

Table 13.15

To What Extent Did You Enjoy Delivering the Year One Vocabulary Programme? (Out of 5 Points: 1=Not Enjoyed, 5=Enjoyed Very Much)

M	*SD*	*Min*	*Max*
4.67	.500	4	5

The additional verbatim comments in this section are shown below. All were linked to the theme of children's enjoyment or engagement:

- *Great for engaging the children and developing multi-sensory learning*
- *Fun and inventive way to teach vocabulary daily*
- *Got quite repetitive – we need other methods besides the cue card*
- *The children were engaged in the activities throughout*

Once again, a Mann-Whitney *U* test was run to test whether enjoyment differed by intervention group. The median confidence scores for the two groups (combined=4.00, semantic=5.00) were not significantly different, U=2.500, z=-2.236, p=.06, displaying similar high levels of enjoyment across the two teaching groups.

At T2, teachers were asked to assess the change in their vocabulary teaching practice as a result of participation in the programme. The mean score in Table 13.16 indicates strong improvement with little variation in responses.

Table 13.16

To What Extent Has Participation in the Year One Vocabulary Programme Changed Your Vocabulary Teaching Practice? (Out of 5 Points: 1=Not Changed, 5=Changed Very Much)

M	*SD*	*Min*	*Max*
4.78	.441	4	5

All additional verbatim comments in the section are set out below, each referring to the theme of improved teacher knowledge.

- *I have gained an understanding of how new vocab can be embedded into a child's use of language*
- *I would not have taught in such detail before, I would have explained the word and moved on*
- *I focus on words a lot more, so pupils are really good at spotting the words we have done*

The median change in practice scores (combined=5.00; semantic=5.00) were equal, indicating a similar high degree of perceived change by both groups.

13.8.2.2 Pupil Impact

Teachers rated the effectiveness of the programme to improve target vocabulary in terms of: (1) understanding (2) expressive use and (3) application to written work. Results in Table 13.17 display a high level of effectiveness (score of 4 and above) for all three areas.

Table 13.17

How Effective Was the Programme for Improving These Aspects of the Taught Vocabulary? (Out of 5 Points: 1=Not Effective, 5=Very Effective)

	M	*SD*	*Min*	*Max*
Understanding Words	4.66	.500	4	5
Verbal Use (in talk)	4.33	.500	4	5
Use in Writing	4.00	.866	3	5

All written comments from this section are shown verbatim below. Within the theme of pupil application of target words, the comments raise two further points, i.e. pupils' difficulty in applying words to their writing and differing levels of application according to ability.

- *During the vocabulary session, the children were very expressive and were able to implement the vocabulary*
- *The children found the words independently in reading*
- *The children really enjoyed spotting vocab words in texts or having opportunities to use words in their writing*
- *More attention is needed to encourage children to use the vocabulary in their writing*
- *Some of the words were challenging for the children to apply in their writing*
- *High achievers challenge themselves and apply the vocabulary; lower achievers showed less application after sessions*

Mann-Whitney *U* tests ascertained whether the teachers' perceptions of pupil progress differed between the combined and semantic groups. Median scores for understanding did not significantly differ between the two groups on any aspect: understanding (combined=4.50, semantic=5.00), U=7.00, z=-.,894, p=.56, expressive use (combined=4.00, semantic=5.00), U=4.00,

z=-1.789, p=.19, or writing (combined=4.00, semantic=5.00), U=7.00, z=-.775, p=.56. In summary, teachers rated all three pupil outcomes highly with no significant differences attributed to intervention type.

13.8.4 Impact of Each Programme Element

Teachers were asked their views on the impact of the programme elements on pupils' learning of target vocabulary, producing the results tabulated in Table 13.18.

Table 13.18

How Effective Were These Elements for Improving Children's Taught Vocabulary? (Out of 5 points: 1=Not Effective, 5=Very Effective)

Programme Element	*M*	*SD*	*Min*	*Max*
Fiction Story Books	4.67	.707	3	5
Daily Lesson - STAR Approach	4.88	.333	4	5
Picture Symbols	4.44	.726	3	5
Teaching Cue Card	4.44	.726	3	5
Pupil-Friendly Definitions	4.56	.527	4	5
Teacher Planning Sheet	3.89	.928	2	5
Pupil Games	3.78	1.201	2	5
Pupil Self-Rating Scale	2.33	1.000	1	4
Word Spy Record	3.33	.866	2	5
Word Wall	3.56	1.013	2	5

On average, teachers rated certain programme elements as more effective than others for enhancing pupils' target word learning. Scores of 4-5 denote effective elements of the programme to support learning, i.e. fiction story books, the daily lesson (STAR), picture symbols, teaching cue card and pupil-friendly definitions. These elements displayed little variation in teachers' ratings (3-5). Scores between 3-4 indicated elements that *some* teachers perceived to be effective, i.e. teacher planning sheets, pupil games, word spy record and word wall, with much higher variation in ratings (2-5). Finally, a mean rating below 3 was given for the pupil self-rating scale, suggesting that this was considered least effective, with concomitant high variability between teacher ratings (1-4). Qualitative comments coded into themes are summarised in Table 13.19, lending further weight to the numerical findings. Six teachers stated that the pupil self-rating scale task was too difficult for most pupils.

Table 13.19

Programme Elements, Themes and Frequency – Effect on Improving Target Vocabulary

Programme Element	*Theme*	*Frequency*
Daily Lesson - STAR	The teaching input was the most valuable aspect	2
	The review was important	2
Picture Symbols	Symbols worked well for less able pupils / nonreaders	2
Teaching Cue Card	The sound cues helped with phonics	2
Pupil Self-Rating Scale	Children could not say whether or not they understood a word; could not do independently	6
Pupil Games	Games help to consolidate	1
	Pupils became disengaged with the games towards the end	1
Word Spy Record	Teacher praised the use of the word orally instead	2
Word Wall	Children referred to the word wall for previous vocabulary	1

To gauge whether teachers would wish to continue implementing aspects of the teaching programme in future, they were asked to rate each programme element again for feasibility of delivery. Results are shown in Table 13.20.

Table 13.20

How Likely is it That You Will Use These Strategies Again in Classroom Vocabulary Teaching? (Out of 5 points: 1=Not Likely, 5=Very Likely)

Programme Element	*M*	*SD*	*Min*	*Max*
Fiction Story Books	5.00	.000	5	5
Daily Lesson - STAR Approach	4.78	.441	4	5
Picture Symbols	4.56	.726	3	5
Teaching Cue Card	4.67	.500	4	5
Pupil-Friendly Definitions	4.78	.441	4	5
Teacher Planning Sheet	3.56	1.130	2	5
Pupil Games	3.44	1.510	1	5
Pupil Self-Rating Scale	2.22	.833	1	3
Word Spy Record	2.67	1.66	1	5
Word Wall	3.89	1.17	2	5

Teachers' willingness to continue using aspects of the programme mirrors the results of the previous section on pupil effectiveness. The same five elements were rated as very/quite likely to be used again (Means 4-5) accompanied by little variation in response (3-5). The moderately rated aspects (Means 3-4) were similar to the previous section, except the word spy record joined the pupil self-rating scale as less likely to be implemented (Means 2-3) with the widest range of ratings (1-5). In addition to the numerical data above, a few additional written comments were made, which did not illuminate any new themes. The programme elements, themes and frequency of comments are presented in Table 13.21 below.

Table 13.21

Teacher Intention to Use Programme Elements in Future

Programme Element	*Themes*	*Frequency*
Daily Lesson - STAR	Will teach word of the day	1
Teaching Cue Card	Will use this teaching process	1
Teacher Planning Sheet	These were useful as a starting point	1
	Good for communicating with job share	1
Pupil Games	Supported independence	1

The challenges described by teachers fell into three main areas: timing considerations, pupil games and the session structure. Teacher comments have been amalgamated into themes in Table 13.22.

Table 13.22

Themes, Challenges and Suggested Programme Changes

Themes	*Challenges*	*Frequency*
Timing	Lessons took more than 15 minutes	1
	Hard to fit in 5 sessions per week	2
	English lesson already too full – story time instead	1
Pupil Games	Need to change more frequently to keep them interesting	5
	Squabbling and behaviour issues in the beginning	2
Lesson Structure	Whole class teacher input only, not other sections	1
	Split the lesson – AM input, PM games	1

The main challenges related to the need for more frequent game changes and the difficulty of fitting a daily vocabulary lesson into an already full curriculum.

13.9 Chapter Summary

In terms of vocabulary outcomes**,** group differences were evident on the taught vocabulary. Definitions were learned equally well in the two intervention groups at the end of the teaching programme (T2), both significantly better than controls. The combined group significantly outperformed both the semantic group and controls at T3. Standardised BPVS3 and CELF4 vocabulary assessments measuring untaught words showed no differences in group performance over time.

Varying results were seen on phonological awareness measures. Ceiling effects on the rhyme and alliteration variables at T2 and T3 necessitated transformation. No group differences were found on rhyme at any timepoint, and alliteration differed only at T3 between the two similarly performing intervention groups and controls. The higher-level elision task demonstrated a significant advantage for the combined group compared to controls at T3.

Nonword reading showed a significant advantage at T3 for the combined group over the semantic group and controls with medium effect size. The semantic group was not significantly better than controls. There were no significant group differences for the nonword spelling variable at any timepoint.

Correlational analyses addressed the question of whether children in the combined intervention group benefited equally or differently according to their starting abilities. The weak partial correlation for taught definitions suggests slightly higher impact for pupils with lower initial vocabulary scores. The moderate bivariate correlations for elision and nonword reading indicate that, on average, children's results would show a moderate degree of difference, benefitting those with lower starting ability more strongly on these measures.

The teacher questionnaire indicated a significant 55% improvement in confidence to teach vocabulary. Ratings indicated that teacher enjoyment of the programme was very high, as well as their perception of improved vocabulary teaching ability. Teachers stated that the programme had a strong impact on target vocabulary in terms of children's understanding and verbal use, but slightly less for writing. Turning to the perceived impact of each programme element on taught vocabulary outcomes, nearly all teachers rated five main elements very highly, including storybooks, the daily lesson approach, symbols, the teaching cue card and pupil-friendly definitions. The other four features received a moderate rating, including the teacher planning sheet, pupil games, word spy record and the word wall. The self-rating scale was considered a relatively weaker method to foster improvement. Ratings of feasibility to implement programme elements were largely similar to those just discussed. Lower value was placed on the word spy record and the self-rating scale. Challenges presented by the programme were linked to time pressures, the pupil games and the session structure. Qualitative comments aligned well with the ratings given. Results were statistically equivalent across the two intervention groups.

Next, the results for the intervention study will be discussed in Chapter 14 in terms of their relationship to the theoretical and empirical literature.

Chapter 14: Discussion of the Vocabulary Intervention Study

14.1 Chapter Introduction

Chapter 14 will consider the results of the main intervention study in relation to the theoretical and empirical literature. Interpretation is based on (1) pupil performance data for the semantic, combined and control groups at three testing points and (2) responses to a teacher questionnaire before and after the teaching intervention. Firstly, outcomes of vocabulary, phonemic awareness and nonword literacy will be considered. Next, the impact of vocabulary teaching for pupils of differing abilities will be explored. The discussion then centres on the effectiveness and feasibility of the intervention from the vantage point of participating teachers. Finally, the contributions and limitations of the experimental vocabulary intervention study will be explored.

14.2 Impact of the Teaching Intervention on Vocabulary Outcomes

14.2.1 Taught Vocabulary

<u>Experimental RQ 1</u>: *Which teaching approach (combined, semantic, control) improves target vocabulary most in 5-6 year olds?*

Hypothesis 1 anticipated that both intervention groups would demonstrate similar and significantly superior performance to controls. This was upheld at T2 when both the combined and semantic approaches produced significantly higher levels of taught vocabulary compared to controls with large effect sizes. The result concurs with the existing literature demonstrating that both semantic and combined training methods are effective for increasing taught vocabulary (Wisenburn & Mahoney, 2009). The large gains may have been facilitated by the extended 26-week timespan of the intervention, which is considerably longer than the average duration of vocabulary intervention of 8.4 weeks reported in Marulis and Neuman (2010).

However, the hypothesis was not supported at T3 four months after vocabulary teaching finished, at which time the combined group significantly outperformed both other groups. The combined group improved on a steady trajectory throughout the intervention, with sustained gains from T2 to T3 despite no further teaching of target words, a situation that also occurred in the form-based group in Janssen et al. (2018). In contrast, semantic group performance mirrors the control group from T2 to T3 (see Figure 13.1 in Chapter 13). This result was surprising, given that teachers in the combined condition divided instructional time between the phonological form and meaning of words, allowing just one opportunity to define target vocabulary per session, whereas the semantic group focussed solely on meaning and practised definitions twice per session. The additional practice time should logically favour higher definitions performance in the semantic condition. However, this was not the case. A single word definition in each lesson, alongside phonological input, led to better outcomes.

The unpredicted T3 result sits well with predictions of the lexical quality hypothesis regarding the benefit of multimodal input for lexical representation (Perfetti & Hart, 2002). Current results indicate that the additional phonological learning provided through combined instruction may have enhanced lexical representation quality for taught words more than a semantic focus alone. In theoretical terms, a phonological representation that is well-specified in terms of distinctness (Elbro, 1996) and segmental quality (Metsala & Walley, 1998) has the ability to compete with similar-sounding entries during auditory word recognition (Gaskell & Marslen-Wilson, 1997) and may thus optimally support vocabulary encoding and retrieval (Storkel & Morrisette, 2002). This enhancement of phonological representation may account for the observed advantage for the combined group.

Another potential explanation lies in the mutual facilitation of phonology and semantics in the combined pedagogy that could have strengthened the associative link needed for successful performance on the definitions task. To define a word, the child must recognise a phonological form and attach it to meaning to retrieve the stored lexical representation. Based on predictions by Storkel and Morrisette (2002), the combined emphasis enables superior recall of taught words over the longer term, as seen at T3 in the current study, seemingly by creating a more robust and enduring lexical representation. The definitions result also concurs with work by Dumay and Gaskell (2007) suggesting that the extended process of vocabulary consolidation is more successful when based on a well-specified multimodal word form. This effect may underlie the sustained trajectory of definitions learning in the combined condition even after teaching ended.

The resulting stronger definitions performance after combined instruction accords with a wealth of research on language-impaired pupils (Wisenburn & Mahoney, 2009), as well as the more limited mainstream literature (Damhuis et al., 2016; Janssen et al., 2018; Moran & Moir, 2018; Silverman, 2007). The current design extends findings by Damhuis et al. (2016) and Moran and Moir (2018) by including an alternative (semantic) teaching approach, which demonstrated that combined instruction delivered better definitions outcomes not only in comparison to a control group but also compared to meaning-based instruction. The inclusion of a delayed maintenance test extends the study by Janssen et al. (2018) by enabling observation of important effects that were not present directly after intervention. Current results do not coincide exactly with Silverman (2007), which found significantly higher definitions performance at T3 for the combined condition compared to the contextual group but equal performance to semantic instruction. The current study built upon this finding by also comparing progress to a control group representing usual teaching and development.

14.2.2 Standardised Vocabulary Measures

<u>*Experimental RQ 2*</u>*: Which teaching approach (combined, semantic, control) improves standardised vocabulary most in 5-6 year olds?*

Hypotheses 2 stated that no significant group differences would be seen on the standardised measures of vocabulary. As expected, and in line with numerous studies (Marulis & Neuman, 2010), global vocabulary measures were less sensitive to gains from intervention, so group differences were not observed. It should be noted that it was not possible to accurately evaluate BPVS3 results using the mixed ANOVA model due to the pre-existing significant group difference on this variable.

14.3. Impact of the Teaching Intervention on Phonemic Awareness

It was posited that phonological awareness variables would improve significantly more in the instructional groups than the control group due to restructuring of the extra vocabulary acquired through intervention. The combined group was furthermore expected to significantly outperform the semantic group owing to the explicit focus on phonological form.

<u>Experimental RQ 3</u>: *Which teaching approach (combined, semantic, control) improves rhyme most in 5-6 year olds?*

Hypothesis 3 maintained that rhyme outcomes would be highest in the combined condition due to the specific input on large segments of syllable and rhyme. However, this was not upheld, since no significant interaction was discovered. Two influential factors are worthy of exploration. The presence of ceiling effects at T2 and T3 severely restricted speculation about whether group differences would arise with a harder measure (Zhu & Gonzales, 2017). Current results contradict those of Munro et al. (2008) that found significantly better rhyme improvement in the combined teaching condition. This can perhaps be best explained by the younger sample (4-6 year olds) in the Munro et al. study, which did not reach ceiling performance.

<u>Experimental RQ 4</u>: *Which teaching approach (combined, semantic, control) improves alliteration most in 5-6 year olds?*

Hypothesis 4 for alliteration also suggested that the best outcomes would be seen in the combined condition, and this too was not confirmed. No group differences were evident at T2, however at T3 both instructional groups significantly outperformed controls as hypothesised. Whilst the T3 result hints at a possible lexical restructuring influence, the presence of ceiling effects prevents further interpretation. Alliteration ceiling effects at T2 became even more prominent at T3, so it is not clear whether group differences on alliteration would be observed at T3 if maximum performance had not been reached.

The above alliteration results agree with those of Munro et al. (2008), who found significantly improved alliteration performance in the combined condition compared to controls. However, the lack of a semantic intervention prevented analysis of teaching outcomes in an alternative condition. Their pretest post-test design did not allow analysis of long-term gains, which was made possible in the current study through a maintenance phase. Current results coincide with studies by Janssen et al. (2018) and Droop et al. (2005), which both observed significantly greater phonemic awareness results for instruction adding a focus on phonological form. It is not possible to directly compare these results to the current study, however, due to their use of composite assessment measures comprising rhyme, phonemic awareness and letter names.

Contrary to expectation, the additional phonological awareness input did not confer any further rhyme or alliteration advantage at T3 for the combined instructional group. One possibility is that the extra phonological input did not additionally enhance lexical quality. A more likely explanation is that the severe ceiling effects at T3 potentially obscured analysis of group differences.

Experimental RQ 5: Which teaching approach (combined, semantic, control) improves elision most in 5-6 year olds?

Hypothesis 5 for elision that the combined group would show the most improvement was partly confirmed. The elision task had higher age norms, so the lack of ceiling effects at any timepoint offers a clearer picture of the causal influence of vocabulary intervention on phonemic awareness and the theoretical processes involved. The hypothesis was not confirmed at T2, since all groups again made sizeable but equal progress. A contrasting result was seen at T3, when only the combined group performed significantly better than controls.

The T3 elision result is theoretically consistent with greater lexical quality (Perfetti & Hart, 2002) arising from dual sound-meaning input in the combined condition. However, since the semantic group did not also outperform controls on elision, the role for lexical restructuring in the results is unclear. Given that the teaching groups had significantly higher taught vocabulary outcomes compared to controls, lexical restructuring theory would predict that increased vocabulary size should promote more segmental representations in the teaching groups, and this would be reflected in higher phonemic awareness gains than control pupils who did not show the same vocabulary increases. However, T3 vocabulary gains were significantly higher in the combined group, potentially leading to a greater degree of restructuring, so perhaps this led to the lesser result in the semantic group. The premise of lexical restructuring was thereby not fully supported in the current data. One possible interpretation is that lexical restructuring as a natural developmental process is not influenced by immediate vocabulary gains over the shorter term. However, this does not necessarily imply that enhanced lexical restructuring effects would not be seen as a result of long-standing vocabulary instruction.

The nature of the elision task should also be borne in mind when interpreting the above result. As shown in the factor analysis of phonemic awareness tasks by Yopp (1988), elision belongs to the higher-level segmentation factor, characterised by increased demand on executive function, particularly working memory. Unfortunately, it is not possible to find a suitably difficult phonemic awareness task without this confound. The high correlation between factors (r=.77) indicates that elision is nonetheless a valid measure of phoneme awareness skill. Since the task was administered fairly across all conditions and no ceiling effects were detected, the higher performance in the combined group over controls can be considered a reliable result. As it was a novel measure in the literature, no papers with the elision variable were available for comparison.

Another question to consider is why all groups made similar (high) progress from T1 to T2 on elision, although outcomes diverged at T3. One possibility is that this may be related to the nature of the daily phonics instruction in Year One, which may be more intensive than in Year Two in preparation for the national Phonics Screening Check in the summer term (Standards and Testing Agency, 2012). The strong effect of phonics training, also observed by Fricke et al. (2013), may overshadow any potential effect of lexical restructuring or higher lexical quality. It is more valid to evaluate the T3 results the following academic year once the phonics screen had been completed and usual levels of phonics teaching presumably resumed.

14.4 Impact of the Teaching Intervention on Phonic Literacy

<u>Experimental RQ 6</u>: *Which teaching approach (combined, semantic, control) improves nonword reading most in 5-6 year olds?*

Analogous to the phonemic awareness variables, hypothesis 6 anticipated that both instructional groups would demonstrate significantly better nonword reading performance than controls due to restructuring, with the best ultimate outcome for the combined group linked to the explicit phonemic awareness content of their vocabulary teaching. The hypothesis was not supported at T2, since all groups made equal gains in nonword reading. The hypothesis was partly confirmed at T3 when the pattern changed markedly. Significantly better results were found in the combined group compared to both the semantic group and controls. However, contrary to expectation, the semantic group did not differ significantly from controls, further discounting a role for lexical restructuring, since the semantic group gained more vocabulary. As hypothesised, the combined group outperformed both other groups, arguably by nature of the additional phonemic input supporting a more segmental and multifaceted quality of lexical representation (Perfetti & Hart, 2002). It follows that words taught in the semantic condition may be represented by weaker lexical quality, since performance was the same as controls. The combined group advantage was expected to be a modest one, given that vocabulary predicts a small amount of the variance in phonemic awareness in younger pupils (Wagner et al., 1997). The medium effect size of the difference between the combined and semantic conditions was higher than anticipated (d=.67).

The equal group results at T2 were unsurprising, since nonword reading was the focus of the Year One phonics screen for which schools often provide intensive teaching (NFER, 2013). Further confirmation of the effect of phonics training can be seen in the equal group performance across three measures at T2 (alliteration, elision and nonword reading) and the diverging group outcomes at T3.

The result of higher nonword reading performance in the combined condition reflects predictions of connectionist models of reading (e.g. Harm & Seidenberg, 1999), which suggest that reading accuracy is supported by both sound and meaning. Semantic input alone did not confer a significant nonword reading advantage compared to typical (control) instruction. However, in line with connectionist principles, the combined teaching input produced the greatest support for reading accuracy.

A further potential account of the higher nonword reading result in the combined condition is provided by the literature demonstrating that phonemic awareness training has a causal impact on phonic reading (Ehri et al., 2001). Although the current intervention was not explicitly linked to print, the combined group received direct teaching of phonemic segmentation linked to the target vocabulary. It is plausible that this input led to enhanced phonemic awareness, as evidenced in the superior elision result, which we know mediates improvement in phonic reading (Hulme et al., 2015). The semantic group did not demonstrate significantly better elision results compared to controls, potentially explaining their similar pattern of nonword reading results.

The important role of print for teaching phonic decoding is well-established (ibid). The effect of this variable was deliberately minimised to experimentally test the influence of oral vocabulary segmentation. Even though the overall role of orthography in the nonword reading results cannot be quantified in the current design, the fact that both instructional groups experienced the same print exposure implies that this variable did not influence the significant group difference in nonword reading outcomes.

<u>Experimental RQ 7</u>: Which teaching approach (combined, semantic, control) improves nonword spelling most in 5-6 year olds?

As expected, hypothesis 7 confirmed no significant interaction for nonword spelling outcomes, with groups demonstrating comparable gains at each timepoint. Likely factors include the oral nature of the vocabulary programme, accompanied by a lack of explicit grapheme instruction. This result aligns with prediction studies that generally have not identified vocabulary to be a significant predictor of spelling in the identified age group (Caravolas et al., 2001, Kim et al., 2013).

14.5 Effects of Different Starting Points on Performance

Post hoc partial correlations were performed as a preliminary analysis to ascertain whether pupils benefitted equally from intervention on variables of taught vocabulary, elision and nonword reading. All results indicated that pupils' starting ability was significantly related to their progress, albeit to different degrees.

Taught definition results demonstrated a largely homogenous effect, suggesting that pupils with lower starting points made only slightly more gains than those with higher initial scores. This confers a degree of confidence that the combined approach can be used to enhance vocabulary across the ability range. The weak magnitude of partial correlation does not imply that combined intervention would substantially narrow the gap for pupils with low starting vocabulary. Current results largely concur with other investigations showing that children with low initial vocabulary derive similar benefit from vocabulary instruction if explicit instruction is used (Coyne et al., 2004, 2009; Elley, 1989; Justice et al., 2005; Marulis and Neuman, 2010). However, the current analysis extends this finding by evaluating the separable impact of the combined approach, whereas other studies evaluated performance of high- and low-vocabulary pupils under a semantic pedagogy.

Elision and nonword reading performance both displayed the same pattern of partial correlation, i.e. initial ability was significantly related to progress on these measures to a moderate extent. The combined intervention was therefore moderately more effective for those with lower initial scores on elision and nonword reading. Whilst the additional phonological facilitation cues experienced in combined vocabulary instruction may help children with weaker early literacy skills (phonemic awareness and nonword reading) to lessen the gap in these vital areas, it can also be argued that the approach is less effective for higher ability pupils who make respectively less gains. Given the age of the sample, however, this should not present a problem, as phonemic awareness and phonic decoding are typically nearing maturity by the end of Year One (Wagner et al., 1997).

14.6 Impact of Intervention on Teacher Confidence and Skill

Drawing upon questionnaire data, teachers' confidence in their ability to teach vocabulary increased by a significant 55% over the course of the programme. This can help to meet the staff development need identified in a survey of primary school teachers by the Communication Trust (2017) in which 100% agreed with the importance of oral language, yet only 4% felt adequately trained in this area. The underpinning pedagogy for both teaching groups incorporated effective principles of vocabulary instruction, so it is not surprising that teachers expressed equal gains in confidence, enjoyment of the programme and perceived skill development. Aspects of the intervention design that may have supported these positive results include the initial training session, regular contact with a language specialist, a manualised programme and the extra vocabulary teaching experience gained through participation in the intervention. Since a similar questionnaire was not administered to teachers of the control classes, a confounding effect of participation in the programme cannot be ruled out. The suggestion of bias may be lessened by the fact that intervention teachers consistently gave a score of 4 or 5 for this question.

14.7 Effectiveness of Programme Components

Teachers rated pupil vocabulary gains equally in both teaching conditions. High pupil impact scores were consistently given for the following components: storybook approach, direct instruction model, visual symbols, facilitation cues and pupil-friendly definitions. Interestingly, these features were drawn directly from the evidence base on effective vocabulary instruction (see Chapter 10). There was slightly more variation in responses regarding the teacher planning sheet, pupil games, self-rating scale, word spy record and word wall, all of which were developed by the researcher based on practical ideas drawn from the literature. These features were popular with some teachers though less valued by others. The teacher planning sheet was regarded as useful for quick access to the pupil-friendly definitions, but record keeping was considered superfluous. Year One pupils did not typically possess the metacognitive skills required for the self-evaluation task, particularly the nuanced understanding of the levels of word knowledge, i.e. what comprises word comprehension. The word spy record was utilised in some classrooms, whereas it was ignored in others. The word wall was seen as important for displaying the necessary resources and was used consistently during fidelity checks, but in the main it was not used interactively as intended.

Whilst not covered in the questionnaire, several other features of effective vocabulary instruction reviewed in Chapter 9 merit evaluation. The tier two word choices were successful on two levels: (1) they provided an adequate span of difficulty for the Year One age group, as evidenced by the lack of floor or ceiling effects on the bespoke vocabulary measure, and (2) fidelity visits indicated that pupils and teachers enjoyed engaging with the selected words. The high number of exposures for each item (facilitated by the teaching cue card) appears to have supported learning across the ability range, concurring with the literature that pupils with vocabulary difficulties require more encounters to encode and retain new vocabulary (Best, 2005; Dollaghan, 1987). The principles of distributed practice (Bahrick & Hall, 2005) and interleaved review (Kang, 2016) may also have contributed to the high retention of taught vocabulary in both teaching groups.

14.8 Feasibility of the Intervention

Teacher observation and discussions indicated that the intervention was realistic to deliver. Questionnaire responses on the practicality of specific programme components largely mirrors teacher views of how the intervention impacted on pupil performance. Further evidence of feasibility was observed on several occasions when the teacher was absent or away from the class, and the TA or supply teacher was able to continue with the programme with minimal support.

Teachers cited three main challenges to fidelity of delivery: timings, pupil games and the session structure. The 10-minute daily input was not always easy to fit in to the literacy lesson, resulting in two schools delivering the intervention during a separate story time. This created less time pressure but reduced consolidation opportunities during literacy activities. Two teachers suggested splitting the STAR structure such that the direct teaching input could occur in the literacy session and the practice game at another time of the day. In line with principles of distributed practice, this method could actually enhance learning by leaving a time gap between initial input and review. Nonetheless, most teachers saw value in the practice games, as exhibited in their ratings and comments in Chapter 13 section 13.8.2.1. Identified issues with the games were linked to pupil behaviour and the need to refresh resources regularly to maintain pupil interest and attention.

14.9 Strengths and Contributions of the Vocabulary Intervention Study

The present study extends the evidence base in several important ways. The mainstream sample enabled data to be collected across a wide and representative mainstream ability range, thus enriching previous experimental literature that concentrated on language-impaired pupils and a small number of school-based investigations. The mainstream sample lends confidence to extrapolating results to the wider Year One population. On the basis of the prediction evidence described in Chapter 3 (e.g. Wagner et al., 1997), current results could also have relevance for Foundation and Year Two classrooms, although further enquiry would be needed to more firmly establish whether this is the case.

The mixed within-between design allowed the research questions to be addressed concisely. Incorporating three experimental groups facilitated a systematic build-up and manipulation of vocabulary and phonemic awareness variables during teaching. The inclusion of both a comparison and control group built upon previous designs that included one but not both conditions. The use of three testing points augments the earlier pretest post-test designs characteristic of the current vocabulary intervention literature. The delayed maintenance test enabled observation of valuable effects that were not present at T2, including the steady increase in vocabulary learning over time for the combined group and the dissociation of phoneme-related outcomes between T2 and T3 once the Phonics Screening Check had finished. Importantly, the experimental design enabled the role of vocabulary as a predictor of phonemic awareness and decoding to be elevated to a causal mechanism, i.e. combined vocabulary instruction improved these distal outcomes.

A main contribution of the experimental intervention study stems from the measures used in the assessment battery. The inclusion of phonemic awareness and decoding measures extended the literature, as these were rarely included in vocabulary intervention studies. The current investigation was the first to assess decoding outcomes resulting from oral vocabulary instruction, although the results of Munro et al. (2008) suggested this as the next step for empirical enquiry. As in the cross-sectional study, separate rather than composite assessments enabled greater specificity of reporting. Splitting vocabulary into its receptive and expressive components and phonological awareness into rhyme, alliteration and elision provided more fine-grained results than the previous literature. The inclusion of a novel spelling measure has provided the evidence base with some initial data on the lack of impact of vocabulary teaching on this variable.

Whilst not meeting all criteria for a RCT design (see limitations in section 15.4), the intervention study was able to incorporate many control features to enhance the methodological rigour and the reliability of vocabulary intervention research. RCT features included the trialling of all materials, random allocation to experimental conditions, clear inclusion criteria, blinding of testers and participating teachers, triple marking of assessments, fidelity checks, manualised interventions, strict teaching protocols and evidence-based vocabulary principles. Importantly, both studies achieved a sample of adequate power to lend authenticity to results.

The intervention study has a high level of ecological validity, according to teacher ratings and researcher observations indicating feasibility for whole-class vocabulary teaching in the classroom setting. It can be delivered within existing curriculum scheduling and at little extra cost. Two aspects which appeared to enhance effectiveness include the manualised resources and access to a SLCN specialist to provide training and monitoring. Schools can access this type of oral language support from a range of professionals, including SALTs, Advisory Teachers, Educational Psychologists and Literacy Advisors. Manualised programmes can be purchased or developed by schools inhouse.

14.10 Limitations of the Vocabulary Intervention Study

The experimental investigation also carries several limitations. The main limitation is the divergent levels of randomisation and analysis, i.e. randomisation took place at the school level, whereas data analysis was based on pupil-level outcomes. The lack of hierarchical nesting could possibly result in uncontrolled school-level variables influencing results. Educational programme evaluations require a large number of clusters (schools or classes) to achieve adequate statistical power for multi-level modelling (Raudenbush & Schwartz, 2020). On a practical level, it was not feasible to include a larger number of schools in the current PhD study, leading to the decision to accept incongruent levels of analysis, consistent with other vocabulary researchers such as Janssen et al. (2018) and Silverman (2007). Raudenbush and Schwartz (2020) describe the main issue of this design as the tendency to overinflate statistically significant differences due to the sample size overstating power, necessitating cautious interpretation of p values. Effect sizes, on the other hand, are widely considered to be accurate (Raudenbush & Bryk, 2002; Slavin, 2008), enabling the current study to provide valuable evidence to guide educational practice.

Compensatory measures were implemented to counteract the bias stemming from the incongruent units of analysis (school randomisation vs. pupil analysis). SES blocks were created, and the pre-existing group difference on the BPVS3 was controlled to reduce variance (Raudensbush & Schwartz, 2020). As an additional precaution, statistically significant T3 results were scrutinised to ensure that alpha values did not approach the customary cut-off point of p=.05 (ibid). The taught definitions interaction term and all post hoc results were highly significant (p<.001). Elision results were highly significant for the interaction term (p<.001) and for post hoc comparison of the combined vs. control condition (p=.002). Nonword reading results showed highly significant differences for the interaction term (p<.001), for post hoc comparison of the combined vs. semantic groups (p<.001) and for the combined vs. control groups (p=.005). Since in each case results were highly significant with no borderline values, findings can be regarded as valid despite the design limitation.

A related limitation was the inability to adequately measure the variance linked to individual teacher styles. As a teaching intervention, the skill of the class teacher is an important variable. A hierarchical model may have been able to control for some of this variance. Nevertheless, several factors may serve to mitigate this effect. Most schools contained more than one participating class, and each condition included multiple teachers, thus dissipating some of the bias linked to individuals. Also, the SES variable assumed some of the variance for the school, in effect controlling for this element. The scripted programme and manualised resources also assisted towards minimising the impact of differing pedagogical styles. Nevertheless, it must be recognised that the diversity of teaching styles may have influenced results, and these could not be fully controlled. The influence of the teaching condition (semantic vs. combined) over and above individual teacher styles can be observed in the class-level results for taught definitions in Chapter 13 Figure 13.3, which displays a similar pattern of results for each of the combined classrooms, and a different but homogenous pattern of results for the semantic classrooms and control classes. The similar trajectories within each intervention condition discounts a strong effect of individual teaching styles.

The method of school randomisation fell short of the standard for an RCT, since all schools were not randomly allocated to intervention groups. The schools that entered the study earlier wished to be included in the teaching interventions from the outset, rather than waiting until the following year as the control group. The early-applying schools were therefore randomly allocated to the two teaching groups, and the later-applying schools were allocated to the waiting control condition. This could introduce an element of participant bias with the initial recruits potentially having a higher level of motivation, with concomitant impact on results. Responses to the pre-intervention teacher questionnaire (in Appendix S) did not bear this out, however, since ratings for the priority of vocabulary teaching were similar between the intervention and control groups.

The available sample in the region was largely monolingual (see section 12.4.2) and thus not entirely representative of the school population across the UK. It is not known whether results would be different if higher numbers of multilingual pupils had been included, so it could be beneficial to replicate the current study with a sample in line with national norms for EAL learners.

The presence of ceiling effects on two of the phonological awareness assessments meant that potential group differences on rhyme and alliteration could not be fully determined. Although it would have been preferable to find tests with older age norms, this was not possible for such early developing skills. The more complex skill of elision provided a more accurate portrayal of skill development, although the additional task demand of working memory is acknowledged. Elision was a novel measure in the vocabulary intervention literature, leading to limitations in comparing this result to other studies.

The unforeseen influence of the phonics test preparation is a likely factor in the equal group performance at T2 for all phonemic and literacy variables. Throughout Year One, pupils received daily instruction on several skills also being measured in the current study, i.e. phonemic awareness and nonword reading, culminating in the Phonics Screening Check. T3 results provide a more valid picture of the effects of the vocabulary intervention, presumably in the context of typical levels of phonics teaching.

A further limitation is the conceivable influence of orthography on the intervention outcomes. Evidence demonstrates the vital role of graphemes (letters) for teaching phonic decoding, yielding double the effect size (*d*=.67 vs. *d*=.38) of oral-only approaches (National Reading Panel, 2000). Given the strong role of print, it was decided to limit this influence so that the effect of oral vocabulary intervention on decoding outcomes could be ascertained. The printed word was available underneath the Widgit symbol, although teachers were asked not to refer to the written word or its constituent graphemes. Nonetheless, it must be acknowledged that orthographic facilitation may have enhanced nonword reading outcomes. However, since the same symbol cards were used in both teaching approaches, the effect would be observed in both conditions, which was not the case. Results comparing the two instructional groups are therefore not affected. In terms of best practice, the printed word should be included and highlighted during teaching in future interventions.

A main limitation, endemic to vocabulary intervention studies, is the lack of generalised improvement on standardised tests despite significant gains in target vocabulary. This remains an area for future investigation, as the goal of intervention should be to improve wider word learning strategies. In the meantime, however, the accumulation of tier 2 vocabulary through direct word-a -day teaching could amount to sizeable increases over the course of a child's schooling. In the UK, children attend school for 40 weeks per year, or 200 days per annum. Multiplying this by seven years of primary education could yield nearly 1500 academically useful words through direct instruction alone, apart from the other vocabulary learning opportunities that could arise from a whole-school vocabulary focus.

14.11 Chapter Summary

The superior taught vocabulary performance in the combined intervention group at T3 confirms the benefit of incorporating sound-based input into vocabulary instruction. The lack of group differences on standardised vocabulary measures is in keeping with the overall literature showing that vocabulary training has the greatest impact on taught items with limited generalisation to new words.

In addition to greater vocabulary gains, combined vocabulary instruction was more beneficial than semantic teaching an controls for supporting nonword reading and better than controls for high-level phonemic awareness (elision). Phonics training, taught intensively in Year One, may have supported equal gains in all groups at T2 on phonemic awareness and nonword reading measures, thus preventing scrutiny of possible group differences immediately after intervention. Differences were however observed at T3 when the combined group superseded both other conditions. At T3, continued ceiling performance on the rhyme and alliteration variables clouded observation of potential group differences. The absence of ceiling effects on the elision and nonword reading tasks signified that these were developmentally more appropriate measures that provided more valid results.

A tentative post hoc analysis assessed the influence of different starting points on taught vocabulary definitions, elision and nonword reading. Pupils with low starting vocabulary made slightly more vocabulary gains than those with high starting points. Children with weak initial phonemic awareness (elision) and nonword reading made moderately more gains than peers with stronger skills. The analysis provides an initial indication that the combined approach provides an inclusive teaching strategy relevant for a wide spectrum of Year One learners with the potential for some catchup on phonemic skills.

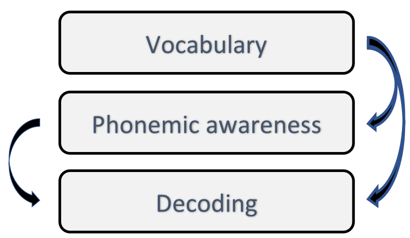
Teachers' confidence in their ability to teach vocabulary rose by a significant 55% over the course of the programme. The evidence-based pedagogical principles were regarded as highly effective by teachers, albeit with some difference of opinion regarding certain programme features. The vocabulary teaching protocol was rated as feasible for future delivery, with useful suggestions made regarding scheduling and the practice games.

In terms of theoretical mechanisms, it is proposed that combined vocabulary teaching may enhance the quality of lexical representations through multimodal input and accurate vocabulary encoding and storage (Perfetti & Hart, 2002), which in turn appeared to support long-term retention more than semantic and usual instruction. Considering the impact of vocabulary gains on distal outcomes of phonemic awareness and nonword reading, the data do not strongly support a role of lexical restructuring (Metsala & Walley, 1998) on these outcomes. Instead, the superior elision performance of the combined group (over controls) and on nonword reading (compared to both groups) suggests an effect consistent with enhanced lexical quality linked to the additional phonological awareness input in the combined teaching condition. Results should be evaluated in the context of ceiling effects on the rhyme and alliteration variables and the potential impact of intensive phonics instruction on phonemic and literacy outcomes at T2. The data does however confirm a role for explicit phonemic awareness cues and multimodal input for enhancing lexical representation quality (Perfetti & Hart, 2002) and connections to print (Harm & Seidenberg, 1999).

A strategy that delivers dual improvement in oral vocabulary and phonemic aspects of literacy can maximise instructional time by delivering greater efficiency than existing semantic teaching focussed solely on word meaning. The current study provides causal evidence that whole-class vocabulary teaching of phonological and meaning aspects of words is an inclusive, effective and feasible approach for a wide range of Year One pupils. Reflecting back to the predictive model of the relationships between vocabulary, phonemic awareness and decoding presented in Chapter 3, we can now adapt the nature of vocabulary's influence on phonemic awareness and decoding to depict its causal impact (see Figure 14.1). Findings point to the important role that vocabulary instruction can play in word-level literacy outcomes.

Figure 14.1

Vocabulary, Phonemic Awareness and Decoding - Causal Relationships



Chapter 15: General Discussion and Conclusion

15.1 Chapter Introduction

Two empirical studies were presented in the current thesis. The first cross-sectional study confirmed receptive vocabulary as a significant predictor of phonemic awareness and decoding in the early stages of primary education. Based on these results, a second experimental study explored whether boosting vocabulary (with and without phonemic awareness cues) might impact on vocabulary as well as distal phonemic awareness and nonword reading skills.

Chapter 15 commences with a synopsis of the findings of both empirical investigations. This is followed by practical implications arising from the current research for professionals in the field of education and speech and language therapy. The thesis raises some further questions for the field of vocabulary intervention research. Finally, the overall conclusions of the thesis will be presented.

15.2 Summary of Research Questions and Findings

A succinct overview of the findings of each study will now be presented.

15.2.1 Summary of Results for the Cross-Sectional Study

The primary questions addressed in the cross-sectional study concern the concurrent relationships between vocabulary, phonemic awareness and phonic reading. Based on a thorough literature review, predictors contributing to phonemic awareness, nonword reading and nonword spelling outcomes were investigated through multiple regression analyses. A three-step hierarchical regression model then endeavoured to quantify the successive contributions of background variables, vocabulary and phonemic awareness for predicting nonword reading outcomes.

15.2.1.1 Predictors of Phonemic Awareness

<u>Cross-sectional RQ 1</u>: *Does receptive vocabulary uniquely predict phonemic awareness (alliteration and phoneme segmentation) in 5-6 year olds, accounting for rhyme awareness and speech perception?*

Moderate correlations were confirmed between receptive vocabulary and phonemic awareness (alliteration, *r*=.48), in line with the reviewed literature (e.g. Bowey & Patel, 1988).

Accounting for precursor skills of rhyme and speech perception, receptive vocabulary was a highly significant predictor of alliteration ($p<.001$), explaining 47% of the variance in performance. Receptive vocabulary also predicted segmentation ($p<.001$), though providing a lesser 23% of the variance in outcomes. Cross-sectional results confirmed receptive vocabulary's contribution to phonemic awareness performance in the current Year One sample.

15.2.1.2 Predictors of Nonword Reading

Cross-sectional RQ 2: Does receptive vocabulary uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?

Cross-sectional RQ 3: Does expressive vocabulary uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?

Cross-sectional RQ 4: Does alliteration uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?

Cross-sectional RQ 5: Does segmentation uniquely predict nonword reading in 5-6 year olds, accounting for other predictors in the model?

The moderate correlations discovered between receptive vocabulary and nonword reading ($r=.49$) were consistent with previous findings (e.g. Bowey & Patel, 1988). Experimental research also confirms a causal impact of phonemic awareness on decoding with a medium effect size (Ehri et al., 2001). The regression model was highly significant ($p<.001$), accounting for 47% of the variance in nonword reading outcomes. Even after accounting for all other variables, receptive vocabulary and alliteration made significant contributions to decoding outcomes. Expressive vocabulary and segmentation did not significantly predict nonword reading performance, although RAN and nonword repetition were significant predictors. Individual differences in vocabulary and phonemic awareness were thus predictive of decoding performance.

15.2.1.3 Predictors of Nonword Spelling

The same multiple regression model was performed for the nonword spelling outcome.

Cross-sectional RQ 6: Does receptive vocabulary uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

Cross-sectional RQ 7: Does expressive vocabulary uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

Cross-sectional RQ 8: Does alliteration uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

Cross-sectional RQ 9: Does segmentation uniquely predict nonword spelling in 5-6 year olds, accounting for other predictors in the model?

Moderate correlations were found between the spelling task and receptive vocabulary (r=.55) and between spelling and alliteration (r=.57), corresponding to the limited research available (Kim et al., 2013; Sénéchal & LeFevre, 2002). The model was highly significant, accounting for 48% of the variation in spelling outcomes. Vocabulary was not found to be a significant contributor to spelling performance, although both phonemic awareness variables emerged as significant predictors. The RAN variable predicted further significant variance in spelling skill.

15.2.1.4 Hierarchical Contributions to Decoding Outcomes

A three-step hierarchical regression model was created to quantify the contributions of vocabulary and phonemic awareness to decoding outcomes. The model was intended to mirror the experimental conditions of the forthcoming intervention. Step one entered relevant within-child background variables to mirror the control group, step two included the vocabulary measures to emulate the additional vocabulary learning in the semantic condition, and step three added phonemic awareness variables to determine their additional influence on nonword reading performance.

Cross-sectional RQ 10: How much contribution do control variables (speech perception, RAN, nonword repetition, rhyme and nonverbal ability) provide to nonword reading in 5-6 year olds?

Cross-sectional RQ 11: *How much additional contribution does vocabulary provide to nonword reading in 5-6 year olds, accounting for control variables?*

Cross-sectional RQ 12: *How much additional contribution does phonemic awareness provide to nonword reading in 5-6 year olds, accounting for the control variables and vocabulary?*

The full model was highly significant (p<.001), explaining 47% of the variance in decoding outcomes. Background control variables accounted for 39%, vocabulary explained a further 3.5%, and phonemic awareness entered last still added 4.7%. Results point towards the possibility that both vocabulary and phonemic awareness could boost decoding performance, albeit to a moderate extent, providing a rationale to test the influence of these variables through a vocabulary intervention.

15.2.2 Summary of Results for the Vocabulary Intervention Study

Based on vocabulary's ability to predict outcomes of phonemic awareness and decoding, an experiment was designed to test the effects of vocabulary instruction for improving vocabulary and distal outcomes of phonemic awareness and nonword reading. A Year One sample was investigated on the basis that the predictive relationships are strongest before age 7-8 (Wagner et al., 1997). Mixed ANCOVA was used to answer the research questions at three timepoints (pre-intervention, post-intervention, maintenance) to ascertain which type of instruction (semantic, combined, control) would lead to the most improvement in vocabulary, phonemic awareness and decoding, adjusting for differences in T1 BPVS scores. A post hoc partial correlation analysis assessed whether all pupils in the combined group gained similarly from the significantly higher gains in vocabulary, elision and nonword reading in this teaching condition.

15.2.2.1 Impact on Vocabulary

Experimental RQ 1: *Which teaching approach (combined, semantic, control) improves target vocabulary most in 5-6 year olds?*

Directly after intervention at T2, both teaching approaches boosted taught vocabulary significantly more than controls with a large effect size (p<.001; combined, d=1.06; semantic, *d*=1.01). No significant difference was found between the intervention groups at this point (*p*=1.00). A different pattern emerged four months later at T3, when the combined group displayed significantly more progress than both other groups, significantly outperforming controls with a very large effect size (*p*<.001, *d*=1.78) and the semantic group with a medium effect size (*p*<.001, *d*=.54). The semantic group also outperformed controls with a large effect size (*p*<.001, *d*=1.15).

<u>Experimental RQ 2</u>: Which teaching approach (combined, semantic, control) improves outcomes on standardised vocabulary most in 5-6 year olds?

In line with most vocabulary intervention studies, no significant interaction was found on the standardised receptive or expressive vocabulary measures, showing that the groups performed equally at T2 and T3.

15.2.2.2 Impact on Phonemic Awareness

<u>Experimental RQ 3</u>: Which teaching approach (combined, semantic, control) improves rhyme most in 5-6 year olds?

Rhyme awareness, a precursor skill to phonemic awareness (Muter et al., 2004), did not display significant group differences at any timepoint. Sizeable ceiling effects at T2 and T3 led to the need to transform this variable, however visual histogram analysis indicated continued skewed data.

<u>Experimental RQ 4</u>: Which teaching approach (combined, semantic, control) improves alliteration most in 5-6 year olds?

Alliteration results were equivalent for all groups at T2 directly after intervention. At T3, group differences emerged, such that both intervention groups were significantly better than controls: semantic vs. controls (*p*=.04, *d*=0.3) and combined vs. controls (*p*=.001, *d*=0.5). No significant differences were apparent between the intervention groups (*p*=.93). Once again, the data at T2 and T3 were subject to ceiling effects, even after transformation. As a developmentally more difficult phonemic awareness skill, elision can provide more reliable results.

<u>Experimental RQ 5</u>: Which teaching approach (combined, semantic, control) improves elision most in 5-6 year olds?

Elision also showed no significant group differences at T2, potentially linked to intensive phonics instruction in Year One (NFER, 2013). At T3, the combined group demonstrated a significant advantage over controls with a medium effect size (p=.002, d=0.5). No significant difference emerged between the semantic and control groups (p=.51) or between the two intervention groups (p=.12).

15.2.2.3 Impact on Nonword Literacy

<u>Experimental RQ 6</u>: *Which teaching approach (combined, semantic, control) improves nonword reading most in 5-6 year olds?*

The influence of phonics training is once again evident at T2, when all three groups performed equally on the nonword reading measure, which was the focus of the Year One Phonics Screening Check. In Year Two (T3), when phonics instruction may have resumed to its more typical level, the combined group superseded both other groups on nonword reading. Combined group scores were significantly higher than results in the semantic group (p<.001, d=.67) and also significantly better than controls (p=.005, d=.49), both with a medium effect size. The semantic intervention was not significantly better than controls (p=.81).

No significant interaction was found for the nonword spelling task, with all groups demonstrating equal performance at all timepoints.

15.2.2.4 Further Analyses: The Effect of Different Starting Points

Post-hoc analyses were performed on taught definitions, elision and nonword reading, all of which displayed significantly higher gains in the combined condition. Partial correlations tested whether pupil progress differed according to initial scores by correlating the T1 score and the progress score, adjusted for the effect of T1 receptive vocabulary differences. Initial scores were significantly related to progress on all three variables to a small to moderate extent.

Pupils in the combined group made significantly higher progress on taught definitions than both other groups. Pearson's partial correlation found a significant but weak negative relationship between starting vocabulary and the level of improvement, $r_{partial}(91) = -.297$, $p=.004$, indicating that pupils with lower prior scores made slightly more progress than those with higher starting points. The combined group performed significantly better than controls on the elision task. The significant partial correlation shows a moderate negative relationship between starting scores and progress on this variable, $r_{partial}(91) = -.407$, $p<.001$, implying that children with lower start points made moderately more progress than those with higher initial ability. The combined group performed significantly better than both the semantic and control groups on this variable. The significant partial correlation once again portrays a moderate negative relationship between initial nonword reading and progress, $r_{partial}(91) = -.407$, $p<.001$, demonstrating the same pattern as the elision variable, i.e. children who started with lower nonword reading scores made relatively more gains than better phonic readers.

15.2.2.5 The Teacher Questionnaire

Participating teachers completed a questionnaire before and after teaching the vocabulary programme to ascertain perceived changes in teacher confidence, vocabulary teaching skill, impact on pupils and programme feasibility. Data included (1) descriptive information (mean score out of five possible points on the Likert scale), (2) nonparametric analyses due to small teacher sample size and (3) qualitative comments.

15.2.2.5.1 Confidence and Professional Development

How confident do you feel about teaching new vocabulary?

To what extent did you enjoy delivering the Year 1 vocabulary programme?

To what extent has participation in the Year 1 vocabulary programme changed your vocabulary teaching practice?

Teacher confidence in their ability to teach vocabulary showed a significant 55% increase ($p=.01$). Teacher enjoyment of the programme was high ($M=4.67$ out of 5 points) as well as perceived teacher development ($M=4.78$). No significant differences were evident between responses for the semantic and combined group teachers on any measure, suggesting equal benefit and enjoyment of the programme, regardless of teaching condition.

15.2.2.5.2 Perceived Impact on Pupils' Taught Vocabulary

How effective was the programme for improving these aspects of the taught vocabulary: understanding, use, writing?

How effective were these elements for improving children's taught vocabulary: storybooks, direct instruction, visual symbols, strategy card, pupil-friendly definitions planning sheet, games, self-rating scale, word spy record, word wall?

Teachers rated pupil impact highly (*M*=4.00-4.66) for vocabulary understanding, expressive use and written application. Qualitative comments indicated that applying vocabulary in written work was the most difficult area for Year One pupils. Teachers scored five design features (storybooks, explicit instruction, visual symbols, strategy teaching card and pupil-friendly definitions) as effective in promoting pupils' vocabulary learning (*M*=4.44-4.88) with little variation in responses. The other four features (planning sheet, games, self-rating scale, word spy record and the word wall) received more moderate scores (*M*=2.33-3.89) with a wider range of ratings. Qualitative comments provided suggestions for future iterations.

15.2.2.5.3 Feasibility of Delivering the Vocabulary Programme

How likely is it that you will use these strategies again in classroom vocabulary teaching (elements stated above)?

Teachers rated the feasibility of using elements of the programme in future. Results mirrored the above findings on pupil impact. The five evidence-based elements were more likely to be used again (*M*=4.78-5.00), whereas the other four features were valued to a more moderate extent (*M*=2.22-3.89). Challenges identified from teachers' qualitative comments centred around timing/scheduling and the pupil games.

15.2.3 Summary of Key Research Findings

- Vocabulary size significantly predicted outcomes of alliteration and nonword reading

- Both semantic and combined vocabulary instruction significantly improved taught vocabulary more than controls
- Combined vocabulary teaching improved taught vocabulary definitions significantly more than semantic teaching with a medium effect size and significantly more than controls with a very large effect size
- In addition to vocabulary gains, combined vocabulary instruction improved higher-level phonemic awareness (elision) significantly more than controls with a medium effect size
- In addition to vocabulary gains, combined vocabulary instruction improved nonword reading significantly more than semantic teaching and significantly more than controls, both with a medium effect size
- The significantly superior gains made in the combined condition on taught definitions were mainly homogeneous across the range of pupils, although those with lower starting scores made slightly more progress and pupils with higher initial scores made slightly less progress
- The significantly superior gains made in the combined condition on elision (better than controls) and on nonword reading (better than both groups) moderately favoured those with lower starting scores
- Teachers gave high ratings for programme effectiveness, pupil impact and feasibility of delivery.

Overall, combined instruction was more effective than customary semantic vocabulary teaching for improving vocabulary and nonword reading and more effective than the control condition on vocabulary, phonemic awareness and nonword reading.

15.3 Implications for Educational and Clinical Practice

The current thesis adds to the evidence base on vocabulary instruction as well as the wider field of speech and language development. Key findings of the experimental intervention study can support school decision-making when choosing effective vocabulary approaches for younger classes. Findings of both studies can provide lessons for initial teacher training, continuing professional development (CPD), as well as for Speech and Language Therapy course content. Finally, several options for programme implementation will be considered, particularly in terms of collaborative practice models between education and speech and language professionals and the manualised resources.

15.3.1 Implications for School Decision-Making

The widespread impact of weak vocabulary development described in Chapter 1 places children's attainment, behaviour and well-being at risk. The economic prosperity of the nation is also affected (ICAN, 2007), as evidenced in a survey of UK employers (Confederation of British Industry, 2015) disclosing that almost half (49%) were not satisfied with the communication skills of young people entering the workplace.

These concerns have accelerated vocabulary up the government agenda, leading to large-scale funding for oral language schemes in primary schools (Bienkiewicz, 2017; Education Endowment Fund, 2017). Since the COVID-19 pandemic, a further downward trend in pupils' oral vocabulary has been documented by the Education Endowment Fund (Bowyer-Crane et al., 2021). Out of 58 primary schools surveyed across the UK, 76% reported that pupils starting school in September 2020 had lower communication skills than in previous years, and 96% said they were concerned about pupils' speech and language development. Post-COVID concerns have prompted further extensive government initiatives, such as Catch Up Premium funding to support language development through the Nuffield Early Language Intervention (Fricke et al., 2017) and for Core Skills funding to boost phonics (GOV.UK, 2021). In light of current results demonstrating a causal impact of combined vocabulary instruction on dual aspects of language and literacy, it is worth considering whether this would be a useful approach to boost low-attaining pupils with early levels of vocabulary and phonics. In this way, the opportunity to increase vocabulary size could also lead to growth in phonemic awareness and phonics.

Due to the large scale of the vocabulary concern (described in Chapter 1), schools are increasingly seeking out whole-school models of vocabulary development (Literacy Trust, 2017).

There is growing awareness that low vocabulary levels reach far beyond pupils with SLCN, confirmed by research showing that at school entry, 75% of economically disadvantaged children in the UK are below average in language development, in comparison to 35% who never experienced poverty (Communication Trust, 2017). These outcomes highlight a large group of children (75%) living in low-income households who may require vocabulary support, but they also hint at the needs of a further sizeable cohort (35%) not living in poverty. Oracy specialists Gaunt and Scott describe the importance of vocabulary instruction for these pupils: “school is their second chance to acquire the rich and varied vocabulary they will need for success both in life and academically” (2019, p.75).

To best facilitate educational decision making, the impact of combined vocabulary instruction needs to be quantified in a format useful to schools. School leaders often consult the Education Endowment Fund (EEF) for information regarding the most effective intervention approaches. It is therefore worth translating the current results to the effect size metric used by the EEF. Two points have a bearing on this interpretation: (1) the results of a single study are not directly comparable to EEF analysis, which operates at a higher level of meta-analysis and systematic review and (2) the EEF mandates rigorous hierarchical designs, which was not met by the quasi-randomised design used in the current investigation. It is nonetheless practical to tentatively situate current results into this comparative context. The EEF Teaching and Learning Toolkit (2017) expresses effect sizes in terms of the number of additional months progress over a year. An effect size of 1.0 would therefore yield on average 12 months of extra progress. Using this metric, oral language approaches in the EEF Toolkit score +6, indicating an additional six months of progress, and phonics interventions score +5. In the current intervention study, taught vocabulary outcomes at T3 equate to well above the highest score of +12 for both groups compared to controls. The combined group also scored +5 compared to semantic intervention. Turning to phonemic awareness gains, the combined group gain was rated at +5 compared to controls and +3 compared to the semantic group. The nonword reading gain was +7 compared to the semantic group and +5 compared to controls. Overall, gains for the combined intervention compare favourably to effective oral language and phonics interventions reported by the EEF.

The main outcome of interest for educators will be the potential for boosting vocabulary. The combined sound-meaning approach provided an effective and efficient method of explicit vocabulary instruction for Year One pupils, leading to the best long-term retention. In addition to the combined sound-meaning cues, it is important to incorporate the evidence-based features of vocabulary instruction that supported these strong gains, including high-quality texts, a focus on tier two words, explicit instruction of child-friendly definitions, rich instruction, active learning and systematic review.

The efficiency of combined instruction lies in its unique ability to produce supplementary learning gains beyond vocabulary over the same instructional time as semantic instruction. Combined instruction provided an additional opportunity to engage with phonemic awareness and segmentation of words outside the daily phonics lesson. This could be beneficial for many Year One pupils, but particularly those whose lexical representations are still holistic in form, since pupils with low phonemic sensitivity struggle with phonics learning. A dual emphasis on boosting vocabulary size and orally breaking down words may support the creation of segmental representations needed for phonemic understanding and the connection to print. The effect of combined vocabulary instruction on distal outcomes of phonemic awareness (elision) and nonword reading are, as anticipated, more modest than the strong effect of intensive phonics instruction (NFER, 2013) on all phoneme-related measures observed at T2.

Practical matters such as feasibility of delivery and resourcing are important criteria affecting choices of vocabulary teaching methods. Teacher ratings on the questionnaire suggest that the programme was realistic and enjoyable to deliver. Achieving the current gains in a short 10-12 minute daily session constitutes a time-efficient approach. Apart from the training implications to be discussed in the next section, there are few additional costs. High-quality texts are already the mainstay of the primary English lesson, and peripheral supplies such as plastic wallets, whiteboards and stationery are commonly available in schools. Symbol software would be beneficial to simplify the creation of vocabulary images by teachers.

Educational leaders require finely tuned evidence to specify which vocabulary approaches can optimally support specific pupil age groups and abilities (Blachowitz et al., 2014). In terms of which cohorts can benefit most from the combined approach to improve instructed vocabulary and wider outcomes, the prediction literature strongly suggests that vocabulary continues to play a role in phonemic awareness and decoding development until around age 8 (Wagner et al., 1997). Based on this evidence, it would be reasonable to utilise combined vocabulary instruction until the end of Key Stage One, pending further studies. Mainstream research beyond Key Stage One is still lacking, but existing experimental studies confirm that combined instruction continues to provide an effective targeted intervention for pupils with persistent vocabulary difficulties at Key Stage Two (St John & Vance, 2014) and Key Stage Three (Joffe et al., 2019; Lowe et al., 2019; Murphy et al., 2017). Once phonological skills reach maturity, vocabulary increasingly starts to predict comprehension skills (Nation & Snowling, 1998; Oulette & Beers, 2010). Semantic-based approaches may therefore be more appropriate for whole-class delivery in Key Stage Two, using the evidence-based tenets of rich instruction (Beck et al., 2013). However, the broad spread of pupil ability suggests that combined instruction will continue to offer a suitable approach for those with vocabulary delay. Morphemic input can also be considered for older classes, due to its important influence on vocabulary and literacy development in the later primary years (Breadmore et al., 2021).

15.3.2 Training Implications

The results of the current intervention study can inform professional training and development opportunities for education and speech and language professionals. On the basis of the cross-sectional data presented in Chapter 7 and supported by the wider prediction literature, staff can feel secure in the knowledge that vocabulary contributes towards performance not only in reading comprehension but also in phonemic awareness and phonic reading ability. The Simple View of Reading (Gough & Tunmer, 1986) has traditionally been conceptualised as two dichotomous skills - oral language proficiency and word reading accuracy. In contrast, current evidence suggests that oral vocabulary and word reading are highly inter-related skills (Duke & Cartwright, 2021; Snowling & Hulme, 2020; Wegener et al., 2022), comprising important information that is not captured in the Simple View. Teaching both skills in an integrated way through combined vocabulary instruction may help to cross this divide.

Initial teacher training would be enhanced by language and literacy modules that consider the overlap between vocabulary and phonemic awareness skills. As part of ongoing vocabulary-related CPD, educators should be made aware of vocabulary's relationship with phonemic segmentation and decoding (Wagner et al., 1997). Teachers of Foundation and Key Stage One could additionally learn to effectively deliver explicit combined vocabulary instruction to supplement current implicit approaches linked to storybooks. SALT training already includes enhanced content on phonemic awareness and vocabulary but could additionally benefit from input on the theoretical and empirical links between these areas. This knowledge would supplement existing SALT expertise to further support effective advice to school staff on vocabulary instruction, particularly in early primary classes.

15.3.3 Dissemination Considerations

Two aspects of the current intervention study offer lessons to support successful dissemination in mainstream education. The existing thesis and a wide literature suggest that collaborative practice between education staff and speech and language professionals can effectively enhance language and literacy outcomes. Once teachers learn about the relevant background linguistic constructs and how to deliver the teaching approach, structured manualised resources can also greatly facilitate effective classroom implementation.

An increasing number of research studies have demonstrated the effectiveness of collaborative models for improving language and literacy teaching (Ciesielski & Greaghead, 2020; Dickinson & Caswell, 2007; Flynn et al., 2021; Goldfield et al., 2021; Neuman & Cunningham, 2009; Wasik & Hindman, 2011). In addition to gains in teacher skills, an RCT by Wasik and Hindman (2011) presents evidence of significantly higher pupil outcomes in vocabulary and phonemic awareness as a result of a collaborative coaching model. The expertise of a speech and language specialist (e.g. Speech and Language Therapist, Advisory Teacher, Educational Psychologist) can be instrumental in increasing staff knowledge through inservice training, as well as supporting effective implementation through guided reflection and individualised feedback (Goldfield et al., 2021).

Another successful method to support rollout is to use a manualised protocol, as shown in the current study. Alternatively, senior leaders and English subject specialists can develop a bespoke structure for their school utilising the STAR approach. A published programme can be a useful tool to support teacher confidence, consistency and fidelity. At present, few programmes based on the combined approach are available for whole-school use (Parsons & Branagan, 2014) or for targeted intervention (ICAN, 2012).

15.4 Suggested Directions for Future Research

Results of the vocabulary intervention study present several exciting directions for future research. The combined sound-meaning vocabulary teaching approach continues to be an under-researched area, particularly in terms of its mainstream application and effect on wider outcomes beyond vocabulary. Initial findings of several studies (Droop et al., 2005; Janssen et al., 2018, Munro et al., 2008), including the current investigation, suggest that phonemic awareness gains are possible as a result of a combined vocabulary teaching approach. Only the current study has so far shown gains in word reading as a result of combined instruction, however a number of experiments based on connectionist models (e.g Harm & Seidenberg, 2004) have demonstrated that teaching word meaning alongside decoding promotes both skills significantly more than decoding intervention alone, indicating a role for both phonological and semantic elements for enhancing phonic reading. It would be useful next to observe the impact of the combined and semantic approaches on phonic reading.

The dearth of studies on spelling suggests this as a useful area to address in future. The current study found that vocabulary plays a lesser role in phonic spelling than reading, however further investigations would be useful to expand upon this result.

Confidence in the generalisability of the intervention study results can be enhanced further through replication using a full RCT design. An upscaled sample of schools would be needed for hierarchical linear modelling, which can account for variance in outcomes linked to school and classroom variables. This type of study design aligns well with current trends in educational research (EEF, 2017).

Another useful avenue of enquiry would be to test the impact of combined vocabulary instruction on a cohort of pupils with poor reading skills, often termed *nonresponders*, i.e. those who do not make progress even after receiving classroom phonics instruction and specific intervention. Research by Savage et al. (2009) revealed that in Year One, 30% of pupils did not improve even after targeted phonics support, indicating that some learners may require an alternative approach to boost early word reading. A number of researchers have demonstrated that nonresponders typically have low levels of vocabulary, leading to the suggestion that early vocabulary may provide a critical foundation for the development of explicit phonological awareness skills (Austin et al., 2021; Dickinson et al., 2003; Whiteley et al., 2007). Phonic interventions may consequently be less successful for pupils with insufficient vocabulary. Support for this argument can be found in the results of the current cross-sectional study and the wider literature confirming vocabulary and phonological awareness as unique predictors of decoding. Using the lens of lexical restructuring theory (Metsala & Walley, 1998), words in a smaller lexicon tend to be represented as whole words, rather than as sublexical units that can support explicit phonemic awareness and decoding. Nonresponders, who lack readiness for phonics instruction, may derive benefit from vocabulary teaching to build their lexical knowledge. Alongside vocabulary learning, the combined approach could provide a focus on the sound structure of new words to enhance the quality of lexical representations, which in turn can serve as the basis for explicit phonemic awareness and word reading.

The role of visual support in vocabulary learning also warrants further exploration. As described in Chapter 10, numerous effective interventions have included visual symbols, however studies have not used sufficient experimental control to determine the precise contribution of this variable (Lawson-Adams et al., 2020). An experimental design that teaches vocabulary with and without the use of symbols will be able to demonstrate any potential causal gains.

In addition to exploring outcomes of combined vocabulary intervention, further research is needed to establish the views of participating teachers and students. This work has begun in the secondary school sector, where the perceptions of therapists and teachers have been documented (Lowe et al., 2019), as well as the student experience (Lowe & Joffe, 2017). Similar studies need to be replicated in primary schools.

A further challenge for the research community would be to develop a way to measure gains in terms of school attainment. Current school-based assessment tools are not adequately fine-grained to provide this data, so this is an important area for development if explicit vocabulary teaching becomes more prevalent. In a similar vein, there is a need to consider why vocabulary teaching interventions so rarely impact on standardised language measures, as this is a necessary step towards promoting independent word learners.

15.5 Conclusion

The present thesis adds to the experimental evidence base on effective vocabulary instruction for mainstream pupils during the initial years of schooling. Schools are reliant on high quality evidence about the effectiveness of intervention approaches that they wish to adopt. It additionally adds to the emerging theoretical base suggesting that oral language has a role to play in word level reading development (Duke & Cartwright, 2021; Snowling & Hulme, 2020; Wegener et al., 2022).

The first empirical aim was to confirm the ability of vocabulary to predict distal outcomes of phonemic awareness and decoding in Year One pupils to inform the experimental research phase. The second aim was to evaluate methods of vocabulary teaching that would lead to the best vocabulary outcomes for young learners. The large cohort of children with low levels of vocabulary prompted the search for a feasible and effective whole-class model of instruction that could fit readily into the school day. By synthesising theory and empirical literature across the fields of education and speech and language therapy, the current experimental study determined a causal role for combined vocabulary teaching in improving phonemic awareness (elision) and nonword reading.

In conclusion, combined instruction that simultaneously attends to the sound structure and meaning of new words provides an inclusive and efficient way to teach vocabulary to Year One mainstream pupils, offering supplementary benefits for phonemic awareness and phonic reading. Outcomes of combined sound-meaning vocabulary instruction outweigh traditional semantic-based approaches in this age group, thereby providing optimal early support to boost children's school attainment and hopefully also their social-emotional wellbeing and future employment prospects.

References

Ainsworth, S., Welbourne, S., & Hesketh, A. (2015). Lexical restructuring in preliterate children: Evidence from novel measures of phonological representation. *Applied Psycholinguistics, 37*(04), 997-1023. https://doi.org/10.1017/s0142716415000338

Alt, M., & Plante, E. (2006). Factors that influence lexical and semantic fast mapping of young children with Specific Language Impairment. *Journal of Speech Language and Hearing Research, 49*(5), 941-954. https://doi.org/10.1044/1092-4388(2006/068)

Altman, D. G. (1998). Adjustment for covariate imbalance. In P. Armtage & T. Colton (Eds.), *Encyclopedia of biostatistics* (pp. 1000-1005). John Wiley.

Anthony, J. L., Williams, J. M., Aghara, R. G., Dunkelberger, M., Novak, B., & Mukherjee, A. D. (2010). Assessment of individual differences in phonological representation. *Reading and Writing: An Interdisciplinary Journal, 23*(8), 969-994. https://doi.org/10.1007/s11145-009-9185-7

Araújo, S., Reis, A., Petersson, K. M., & Faísca, L. (2015). Rapid automatized naming and reading performance: A meta-analysis. *Journal of Educational Psychology, 107*(3), 868-883. https://doi.org/10.1037/edu0000006

Armstrong, R., Arnott, W., Copland, D. A., McMahon, K., Khan, A., Najman, J. M., & Scott, J. G. (2017). Change in receptive vocabulary from childhood to adulthood: Associated mental health, education and employment outcomes. *International Journal of Language and Communication Disorders, 52*(5), 561-572. https://doi.org/10.1111/1460-6984.12301

Austin, C. R., Vaughn, S., Clemens, N. H., Pustejovsky, J. E., & Boucher, A. N. (2021). The relative effects of instruction linking word reading and word meaning compared to word reading instruction alone on the accuracy, fluency, and word meaning knowledge of 4th-5th grade students with Dyslexia. *Scientific Studies of Reading*, 1-19. https://doi.org/10.1080/10888438.2021.1947294

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158-173. https://doi.org/10.1037/0033-295X.105.1.158

Bahrick, H. P., & Hall, L. K. (2005). The importance of retrieval failures to long-term retention: A metacognitive explanation of the spacing effect. *Journal of Memory and Language, 52*(4), 566-577. https://doi.org/10.1016/j.jml.2005.01.012

Bailey, T. M., & Plunkett, K. (2002). Phonological specificity in early words. *Cognitive Development, 17*(2), 1265-1282. https://doi.org/10.1016/S0885-2014(02)00116-8

Ball, E. W., & Blachman, B. A. (1991). Does phoneme awareness training in kindergarten make a difference in early word recognition and developmental spelling? *Reading Research Quarterly, 26*(1), 49-66. https://doi.org/10.1598/RRQ.26.1.3

Beck, I. L., McKeown, M., & Kucan, L. (2002). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.

Beck, I. L., McKeown, M., & Kucan, L. (2013). *Bringing words to life: Robust vocabulary instruction*. Guilford Press.

Beck, I. L., & McKeown, M. (2007). Increasing young low-income children's oral vocabulary repertoires through rich and focused instruction. *The Elementary School Journal, 107*(3), 251-271. https://doi.org/10.1086/511706

Best, W., Hughes, L., Masterson, J., Thomas, M. S. C., Howard, D., Kapikian, A., & Shobbrook, K. (2021). Understanding differing outcomes from semantic and phonological interventions with children with word-finding difficulties: A group and case series study. *Cortex, 134*, 145-161. https://doi.org/10.1016/j.cortex.2020.09.030

Bialystok, E., Luk, G., Peets, K. F., & Yang, S. (2010). Receptive vocabulary differences in monolingual and bilingual children. *Bilingualism: Language and Cognition, 13*(4), 525-531. https://doi.org/10.1017/S1366728909990423

Biemiller, A. (2001). Teaching vocabulary: Early, direct, and sequential. *American Educator, 25*(1), 24-28.

Biemiller, A. (2003). Vocabulary: Needed if more children are to read well. *Reading Psychology, 24*(3-4), 323-335. https://doi.org/10.1080/02702710390227297

Biemiller, A. (2015). Which words are worth teaching? *Perspectives on Language and Literacy, 41*(3), 9-13.

Biemiller, A., & Boote, C. (2006). An effective method for building meaning vocabulary in primary grades. *Journal of Educational Psychology, 98*(1), 44-62. https://doi.org/10.1037/0022-0663.98.1.44

Biemiller, A., & Slonim, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93*(3), 498-520. https://doi.org/10.1037/0022-0663.93.3.498

Bienkiewicz, A. (2017). Up to £45 million up for grabs in latest round of Teaching and Leadership Innovation Fund. *Times Educational Supplement*.

Blachowicz, C. L., Fisher, P. J., Ogle, D., & Watts-Taffe, S. (2006). Vocabulary: Questions from the classroom. *Reading Research Quarterly, 41*(4), 524-539. https://doi.org/10.2307/1510288

Blachowicz, C. L., & Fisher, P. J. (2014). *Teaching vocabulary in all classrooms (5th ed.)*. Pearson.

Blanden, J. (2006). *'Bucking the trend': What enables those who are disadvantaged in childhood to succeed later in life?* Department for Work and Pensions.

Bortfeld, H., Morgan, J. L., Golinkoff, R. M., & Rathbun, K. (2005). Mommy and me: Familiar names help launch babies into speech-stream segmentation. *Psychological Science, 16*(4), 298-304. https://doi.org/10.1111/j.0956-7976.2005.01531.x

Bowey, J. (2001). Does poor exception word reading reflect vocabulary deficits? *Australian Journal of Psychology, 53*(3), 177-177.

Bowey, J. A., & Hirakis, E. (2006). Testing the protracted lexical restructuring hypothesis: The effects of position and acoustic–phonetic clarity on sensitivity to mispronunciations in children and adults. *Journal of Experimental Child Psychology, 95*(1), 1-17. https://doi.org/doi:10.1016/j.jecp.2006.02.001.

Bowey, J. A., & Patel, R. K. (1988). Metalinguistic ability and early reading achievement. *Applied Psycholinguistics, 9*(4), 367-383. https://doi.org/10.1017/S0142716400008067

Bowyer-Crane, C., Bonetti, S., Compton, S., Nielsen, D., D'Apice, K., & Tracey, L. (2021). *The impact of Covid-19 on school starters: interim briefing 1: Parent and school concerns about children starting school*. Education Endowment Foundation.

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 phonology with reading intervention. *Journal of Child Psychology and Psychiatry, 49*(4), 422-432. https://doi.org/10.1111/j.1469-7610.2007.01849.x

Brabham, E., & Lynch-Brown, C. (2002). Effects of teachers' reading-aloud styles on vocabulary acquisition and comprehension of students in the early elementary grades. *Journal of Educational Psychology, 94*(3), 465-473. https://doi.org/10.1037//0022-0663.94.3.465

Brackenbury, T., & Pye, C. (2005). Semantic deficits in children with language impairments: Issues for clinical assessment. *Language, Speech and Hearing Services in Schools, 36*(1), 5-16. https://doi.org/10.1044/0161-1461(2005/002)

Bragard, A., Schelstraete, M., Snyers, P., & James, D. G. (2012). Word-finding intervention for children with Specific Language Impairment: A multiple single-case study. *Language Speech and Hearing Services in Schools, 43*(2), 222-234. https://doi.org/10.1044/0161-1461(2011/10-0090)

Breadmore, H. L., Levesque, K., & Deacon, S. H. (2021). Special issue editorial: Advances in understanding the role of morphemes in literacy development. *Journal of Research in Reading, 44*(1), 1-9. https://doi.org/https://doi.org/10.1111/1467-9817.12346

Bryan, K., Freer, J., & Furlong, C. (2007). Language and communication difficulties in juvenile offenders. *International Journal of Language and Communication Disorders, 42*(5), 505-520. https://doi.org/10.1080/13682820601053977

Bryan, K., Garvani, G., Gregory, J., & Kilner, K. (2015). Language difficulties and criminal justice: the need for earlier identification. *International Journal of Language and Communication Disorders, 50*(6), 763-775. https://doi.org/10.1111/1460-6984.12183

Bryant, P. E., Maclean, M., Bradley, L. L., & Crossland, J. (1990). Rhyme and alliteration, phoneme detection, and learning to read. *Developmental Psychology, 26*(3), 429-438. https://doi.org/10.1037/0012-1649.26.3.429

Burgoyne, K., Kelly, J. M., Whiteley, H. E., & Spooner, A. (2009). The comprehension skills of children learning English as an additional language. *British Journal of Educational Psychology, 79*(4), 735-747. https://doi.org/10.1348/000709909X422530

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. https://doi.org/10.1037/0022-0663.91.3.403

Byrne, B., & Fielding-Barnsley, R. (1989). Phonemic awareness and letter knowledge in the child's acquisition of the alphabetic principle. *Journal of Educational Psychology, 81*(3), 313-321. https://doi.org/10.1037/0022-0663.81.3.313

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology, 33*(2), 111-153. https://doi.org/10.1006/cogp.1997.0649

Callaghan, T. C. (2000). Factors affecting children's graphic symbol use in the third year: Language, similarity, and iconicity. *Cognitive Development, 15*(2), 185-214. https://doi.org/10.1016/s0885-2014(00)00026-5

Cameron, L. (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research, 6*(2), 145-173. https://doi.org/10.1191/1362168802lr103oa

Caravolas, M., Hulme, C., & Snowling, M. J. (2001). The foundations of spelling ability: Evidence from a 3-year longitudinal study. *Journal of Memory and Language, 45*(4), 751-774. https://doi.org/10.1006/jmla.2000.2785

Carey, S. (1978). The child as word learner. In M. Halle, J. Bresnan, & M. GA (Eds.), *Linguistic theory and psychological reality* (pp.264-293). The MIT Press.

Carlsen, K. D., & Herdman, A. O. (2012). Understanding the impact of convergent validity on research results. *Organisational Research Methods*, 15(*1*), 17-32. https://doi.org/10.1177/1094428110392383

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review, 24*(3), 369-378. https://doi.org/10.1007/s10648-012-9205-z

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology, 19*(5), 619-636. https://doi.org/10.1002/acp.1101

Carroll, J., Snowling, M., Hulme, C., & Stevenson, J. (2003). The development of phonological awareness in preschool children. *Developmental Psychology, 39*(5), 913-923. https://doi.org/10.1037/0012-1649.39.5.913

Cassano, C. M., & Schickedanz, J. A. (2015). An examination of the relations between oral vocabulary and phonological awareness in early childhood. *Literacy Research: Theory, Method and Practice, 64*(1), 227-248. https://doi.org/10.1177/2381336915617608

Cassano, C. M., & Steiner, L. (2016). Exploring assessment demands and task supports in early childhood phonological awareness assessments. *Literacy Research: Theory, Method and Practice, 65*(1), 217-235. https://doi.org/10.1177/2381336916661521

Castles, A., & Coltheart, M. (2004). Is there a causal link from phonological awareness to success in learning to read? *Cognition, 91*(1), 77-111. https://doi.org/10.1016/S0010-0277(03)00164-1

Catling, J. C., Pymont, C., Johnston, R. A., Elsherif, M. M., Clark, R., & Kendall, E. (2021). Age of acquisition effects in recognition without identification tasks. *Memory, 29*(5), 662-674. https://doi.org/10.1080/09658211.2021.1931695

Catts, H. W., Fey, M. E., Zhang, X., & Tomblin, J. B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. *Scientific Studies of Reading, 3*(4), 331-361. https://doi.org/10.1207/s1532799xssr0304_2

Chaney, C. (1998). Preschool language and metalinguistic skills are links to reading success. *Applied Psycholinguistics, 19*(3), 433-446. https://doi.org/10.1017/S0142716400010250

Charles-Luce, J., & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language, 22*(3), 727-735. https://doi.org/10.1017/S0305000900010023

Christian, K., Morrison, F. J., Frazier, J. A., & Massetti, G. (2000). Specificity in the nature and timing of cognitive growth in kindergarten and first grade. *Journal of Cognition and Development, 1*(4), 429-448. https://doi.org/10.1207/S15327647JCD0104_04

Ciesielski, E. J. M., & Creaghead, N. A. (2020). The effectiveness of professional development on the phonological awareness outcomes of preschool children: A systematic review. *Literacy Research and Instruction, 59*(2), 121-147. https://doi.org/10.1080/19388071.2019.1710785

Claessen, M., Heath, S., Fletcher, J., Hogben, J., & Leitao, S. (2009). Quality of phonological representations: A window into the lexicon? *International Journal of Language and Communication Disorders, 44*(2), 121-144. https://doi.org/10.1080/13682820801966317

Claessen, M., & Leitao, S. (2012). The relationship between stored phonological representations and speech output. *International Journal of Speech and Language Pathology, 14*(3), 226-234. https://doi.org/10.3109/17549507.2012.679312

Clark, E. V. (2015). Lexical meaning. In E. L. Bavin & L. R. Naigles (Eds.), *The Cambridge handbook of child language* (2nd ed., pp. 351-368). Cambridge University Press. https://doi.org/10.1017/CBO9781316095829.016

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149-210. https://doi.org/10.1007/bf01320076

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science, 21*(8), 1106-1116. https://doi.org/10.1177/0956797610375449

Clayton, F. J., West, G., Sears, C., Hulme, C., & Lervåg, A. (2020). A longitudinal study of early reading development: Letter-sound knowledge, phoneme awareness and ran, but not letter-sound integration, predict variations in reading development. *Scientific Studies of Reading, 24*(2), 91-107. https://doi.org/10.1080/10888438.2019.1622546

Clegg, J. (2014). Curriculum vocabulary learning intervention for children with social, emotional and behavioural difficulties (SEBD): Findings from a case study series. *Emotional and Behavioural Difficulties, 19*(1), 106-127. https://doi.org/10.1080/13632752.2013.854958

Clegg, J., Stackhouse, J., Finch, K., Murphy, C., & Nicholls, S. (2009). Language abilities of secondary age pupils at risk of school exclusion: A preliminary report. *Child Language Teaching and Therapy, 25*(1), 123-139. https://doi.org/10.1177/0265659008098664

CLPE. (2018, June 3). *Book lists*. Centre for Literacy in Primary Education. https://www.clpe.org.uk.

Coady, J. A., & Aslin, R. N. (2003). Phonological neighbourhoods in the developing lexicon. *Journal of Child Language, 30*(2), 441-469. https://doi.org/10.1017/S0305000903005579

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.

Cohen, N. J., Barwick, M. A., Horodezky, N. B., Vallance, D. D., & Im, N. (1998). Language, achievement, and cognitive processing in psychiatrically disturbed children with previously identified and unsuspected language impairments. *Journal of Child Psychology and Psychiatry, 39*(6), 865-877. https://doi.org/10.1111/1469-7610.00387

Coles, H., Gillett, K., Murray, G., & Turner, K. (2017). *RCSLT justice evidence base, consolidation 2017*. The Royal College of Speech and Language Therapists.

Collins UK. (1984). *Cobuild primary learner's dictionary*. Harper Collins.

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology, 33*(4), 497-505.

Communication Matters. (2013). *Shining a light on Augmentative and Alternative Communication*. Communication Matters.

Communication Trust. (2017). *Professional development in speech, language and communication: Findings from a national survey*. The Communication Trust.

Confederation of British Industry. (2015). *Inspiring growth: CBI/Pearson education and skills survey 2015*. Pearson Education.

Conti-Ramsden, G., Durkin, K., Simkin, Z., & Knox, E. (2009). Specific Language Impairment and school outcomes: Identifying and explaining variability at the end of compulsory education. *International Journal of Language and Communication Disorders, 44*(1), 15-35. https://doi.org/10.1080/13682820801921601

Cooper, J., Moodley, M., & Reynell, J. (1978). *Helping language development: A developmental programme for children with early language handicaps*. Arnold.

Coyne, M. D., Simmons, D. C., Kame'enui, E. J., & Stoolmiller, M. (2004). Teaching vocabulary during shared storybook readings: An examination of differential effects. *Exceptionality, 12*(3), 145-162. https://doi.org/10.1207/s15327035ex1203_3

Coyne, M. D., McCoach, D .B., Loftus, S., Zipoli, R., & Kapp, S. (2009). Direct vocabulary instruction in kindergarten: Teaching for breadth versus depth. *The Elementary School Journal*, 110(*1*), 1-18. https://doi.org/10.1086/598840

Coyne, M. D., McCoach, D. B., Ware, S. M., Loftus-Rattan, S. M., Baker, D. L., Santoro, L. E., & Oldham, A. C. (2022). Supporting vocabulary development within a multitiered system of support: Evaluating the efficacy of supplementary kindergarten vocabulary intervention. *Journal of Educational Psychology*. Advance Online Publication. https://doi.org/10.1037/edu0000724

Cramer, D., & Howitt, D. (2004). *The Sage dictionary of statistics: A practical resource for students in the social sciences*. Sage.

Crawford, C., & Greaves, E. (2013). *A comparison of commonly used socioeconomic indicators: their relationship to educational disadvantage and relevance to Teach First*. Institute for Fiscal Studies.

Cutler, A. (2008). The abstract representations in speech processing. *The Quarterly Journal of Experimental Psychology, 61*(11), 1601-1619. https://doi.org/10.1080/13803390802218542

Damhuis, C., Segers, E., Scheltinga, F., & Verhoeven, L. (2016). Effects of individualized word retrieval in kindergarten vocabulary intervention. *School Effectiveness and School Improvement, 27*(3), 441-454. https://doi.org/10.1080/09243453.2015.1114500

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42*(1), 287-314. https://doi.org/10.1016/0010-0277(92)90046-K

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review, 104*(4), 801-838. https://doi.org/10.1037/0033-295X.104.4.801

Department for Children, Families and Schools. (2008). *Teaching effective vocabulary: What can teachers do to increase the vocabulary of children who start education with a limited vocabulary? DCFS-00376-2008*. DCSF.

Department for Education. (2016). *Schools, pupils and their characteristics: January 2016. SFR 20/2016*. DFE.

Department for Education. (2016a). *English proficiency of pupils with English as an Additional Language*. DFE.

Department for Education. (2020). *English proficiency of pupils with English as an Additional Language*. DFE.

Department for Education and Skills. (2007). *Letters and Sounds. DFES-00281-2007*. DFES.

Dickinson, D. K. (1984). First impressions: Children's knowledge of words gained from a single exposure. *Applied Psycholinguistics, 5*(4), 359-373. https://doi.org/10.1017/S0142716400005233

Dickinson, D. K., & Caswell, L. (2007). Building support for language and early literacy in preschool classrooms through in-service professional development: Effects of the Literacy Environment Enrichment Program (LEEP). *Early Childhood Research Quarterly, 22*(2), 243-260.

Dickinson, D. K., McCabe, A., Anastasopoulos, L., Peisner-feinberg, E. S., & Poe, M. D. (2003). The comprehensive language approach to early literacy: The interrelationships among vocabulary, phonological sensitivity, and print knowledge among preschool-aged children. *Journal of Educational Psychology, 95*(3), 465-481. https://doi.org/10.1037/0022-0663.95.3.465

Dickinson, D. K., Nesbitt, K. T., Collins, M. F., Hadley, E. B., Newman, K., Rivera, B. L., Ilgez, H., Nicolopoulou, A., Golinkoff, R. M., & Hirsh-Pasek, K. (2019). Teaching for breadth and depth of vocabulary knowledge: Learning from explicit and implicit instruction and the storybook texts. *Early Childhood Research Quarterly, 47*, 341-356. https://doi.org/10.1016/j.ecresq.2018.07.012

Dineen, L. C., & Blakesley, B. C. (1973). Algorithm AS62: A generator for the sampling distribution of the Mann-Whitney U statistic. *Applied Statistics*, 22, 269-273.

Dollaghan, C. (1985). Child meets word: Fast mapping in preschool children. *Journal of Speech, Language and Hearing Research, 28*(3), 449-454. https://doi.org/10.1044/jshr.2803.454

Dollaghan, C. A. (1987). Fast mapping in normal and language-impaired children. *Journal of Speech and Hearing Research, 52*(3), 218-222. https://doi.org/10.1044/jshd.5203.218

Droop, W., Peters, S., Aarnoutse, C., & Verhoeven, L. (2005). Effecten van stimulering van beginnende geletterdheid in groep 2. *Pedagogische Studien, 82*, 160-180.

Duff, F. J., & Hulme, C. (2012). The role of children's phonological and semantic knowledge in learning to read words. *Scientific Studies of Reading, 16*(6), 504-525. https://doi.org/10.1080/10888438.2011.598199

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry, 56*(8), 848-856. https://doi.org/10.1111/jcpp.12378

Duke, N. K., & Cartwright, K. B. (2021). The science of reading progresses: Communicating advances beyond the Simple View of Reading. *Reading Research Quarterly*, 56(1), S25–S44. https://doi.org/10.1002/rrq.411

Dumay, N., & Gaskell, G. (2007). Sleep-associated changes in the mental representation of spoken words. *Psychological Science, 18*(1), 35-39. https://doi.org/10.1111/j.1467-9280.2007.01845.x

Duncan, L. G., Seymour, P. H. K., & Hill, S. (1997). How important are rhyme and analogy in beginning reading? *Cognition, 63*(2), 171-208. https://doi.org/10.1016/S0010-0277(97)00001-2

Dunn, L., & Dunn, D. (2007). *The Peabody Picture Vocabulary Test (4th ed.)*. Pearson.

Dunn, L., Dunn, D., Styles, B., & Swells, J. (2009). *British Picture Vocabulary Scales (3rd ed.)*. GL Assessment.

Dyson, H., Best, W., Solity, J., & Hulme, C. (2017). Training mispronunciation correction and word meanings improves children's ability to learn to read words. *Scientific Studies of Reading, 21*(5), 392-407. https://doi.org/10.1080/10888438.2017.1315424

Easton, C., Sheach, S., & Easton, S. (1997). Teaching vocabulary to children with wordfinding difficulties using a combined semantic and phonological approach: An efficacy study. *Child Language Teaching and Therapy, 13*(2), 125-142. https://doi.org/10.1177/026565909701300202

Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of educational measurement*. Prentice-Hall.

Echols, C. H., & Newport, E. L. (1992). The role of stress and position in determining first words. *Language Acquisition, 2*(3), 189-220. https://doi.org/10.1207/s15327817la0203_1

Education and Skills Funding Agency. (2018). *Pupil premium: Allocations and conditions of grant 2018-19. allocations (school level)*. GOV.UK.

Education Endowment Foundation. (2017). *Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit*. EEF.

Ehri, L. C., Nunes, S. R., Willows, D. M., Schuster, B. V., Yaghoub-Zadeh, Z., & Shanahan, T. (2001). Phonemic awareness instruction helps children learn to read: Evidence from the National Reading Panel's meta-analysis. *Reading Research Quarterly, 36*(3), 250-287. https://doi.org/10.1598/RRQ.36.3.2

Elbro, C. (1996). Early linguistic abilities and reading development: A review and a hypothesis. *Reading and Writing: An Interdisciplinary Journal, 8*(6), 453-485. https://doi.org/10.1007/BF00577023

Elbro, C. (1998). When reading is" readn" or somthn. Distinctness of phonological representations of lexical items in normal and disabled readers. *Scandinavian Journal of Psychology, 39*(3), 149-153. https://doi.org/10.1111/1467-9450.393070

Elbro, C., & Palleson, B. (2002). The quality of phonological representations and phonological awareness: A causal link. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (Vol. 11, pp. 17-32). John Benjamins Publishing.

Elleman, A. M., Lindo, E. J., Morphy, P., & Compton, D. L. (2009). The impact of vocabulary instruction on passage-level comprehension of school-age children: A meta-analysis. *Journal of Research on Educational Effectiveness, 2*(1), 1-44. https://doi.org/10.1080/19345740802539200

Elley, W. B. (1989). Vocabulary acquisition from listening to stories. *Reading Research Quarterly, 24*(2), 174-187. https://doi.org/10.2307/747863

Elliott, C., & Smith, P. (2011). *The British Ability Scales (3rd ed.)*. GL Assessment.

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. https://doi.org/10.3758/BF03193146

Fazio, L. K., & Marsh, E. J. (2019). Retrieval-Based Learning in Children. *Current Directions in Psychological Science*: a Journal of the American Psychological Society, 28(*2*), 111–116. https://doi.org/10.1177/0963721418806673

Fernald, A., Marchman, V. A., & Weisleder, A. (2013). SES differences in language processing skill and vocabulary are evident at 18 months. *Developmental Science, 16*(2), 234-248. https://doi.org/10.1111/desc.12019

Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. Sage.

Flack, Z. M., Field, A. P., & Horst, J. S. (2018). The effects of shared storybook reading on word learning: A meta-analysis. *Developmental Psychology, 54*(7), 1334-1346. https://doi.org/10.1037/dev0000512

Flynn, N., Powell, D., Stainthorp, R., & Stuart, M. (2021). Training teachers for phonics and early reading: Developing research-informed practice. *Journal of Research in Reading, 44*(2), 301-318. https://doi.org/10.1111/1467-9817.12336

Fowler, A., Swainson, B., & Scarborough, H. (2004). Relationships of naming skills to reading, memory, and receptive vocabulary: Evidence for imprecise phonological representations of words by poor readers. *An Interdisciplinary Journal of the International Dyslexia Association, 54*(2), 247-280. https://doi.org/10.1007/s11881-004-0013-0

Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry, 54*(3), 280-290. https://doi.org/10.1111/jcpp.12010

Fukkink, R. G., & de Glopper, K. (1998). Effects of instruction in deriving word meaning from context: A meta-analysis. *Review of Educational Research, 68*(4), 450-469. https://doi.org/10.2307/1170735

Gallagher, M. A., Barber, A. T., Beck, J. S., & Buehl, M. M. (2019). Academic vocabulary: Explicit and incidental instruction for students of diverse language backgrounds. *Reading and Writing Quarterly, 35*(2), 84-102. https://doi.org/10.1080/10573569.2018.1510796

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language, 45*(3), 468-492. https://doi.org/10.1006/jmla.2000.2784

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes, 12*(5-6), 613-656. https://doi.org/10.1080/016909697386646

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*(2), 200-213. https://doi.org/10.1016/0749-596X(89)90044-2

Gathercole, S. E., & Baddeley, A. D. (1990). The role of phonological memory in vocabulary acquisition: A study of young children learning new names. *British Journal of Psychology, 81*(4), 439-454. https://doi.org/10.1111/j.2044-8295.1990.tb02371.x

Gathercole, S. E., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33*(6), 966-979. https://doi.org/10.1037/0012-1649.33.6.966

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A. M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology, 13*(1), 65-77. https://doi.org/10.1002/(SICI)1099-0720(199902)13:1<65::AID-ACP548>3.0.CO;2-O

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*(5), 887-892. https://doi.org/10.1037//0012-1649.28.5.887

Gaunt, A., & Scott, A. (2019). *Transform teaching and learning through talk: The oracy imperative.* Rowman and Littlefield.

Gentner, D. (1982). *Why nouns are learned before verbs: Linguistic relativity versus natural partitioning. Center for the Study of Reading technical report no. 257*. National Institute for Education.

German, D. J. (1986). *Test of Word Finding*. Developmental Learning Materials.

German, D. J. (2002). A phonologically based strategy to improve word-finding abilities in children. *Communication Disorders Quarterly, 23*(4), 179-192. https://doi.org/10.1177/15257401020230040301

German, D. J., Schwanke, J. H., & Ravid, R. (2012). Word finding difficulties: Differentiated vocabulary instruction in the speech and language room. *Communication Disorders Quarterly, 33*(3), 146-156. https://doi.org/10.1177/1525740111405840

Gibbs, S., & Bodman, S. (2014). *Phonological Assessment Battery Primary (2nd ed.)*. GL Assessment.

Gilhooly, K. J., & Logie, R. H. (1980). Age of acquisition, imagery, concreteness, familiarity and ambiguity measures for 1944 words. *Behaviour Research Methods and Instrumentation*, 12(4), 395-427. https://doi.org/10.3758/bf03201693

Goldfeld, S., Snow, P., Eadie, P., Munro, J., Gold, L., Orsini, F., Connell, J., Stark, H., Watts, A., & Shingles, B. (2021). Teacher knowledge of oral language and literacy constructs: results of a randomized controlled trial evaluating the effectiveness of a professional learning intervention. *Scientific Studies of Reading, 25*(1), 1-30. https://doi.org/10.1080/10888438.2020.1714629

Golinkoff, R. M., Mervis, C. B., & Hirsh-Pasek, K. (1994). Early object labels: The case for a developmental lexical principles framework. *Journal of Child Language, 21*(1), 125-155. https://doi.org/10.1017/S0305000900008692

Goswami, U. (1999). Causal connections in beginning reading: The importance of rhyme. *Journal of Research in Reading, 22*(3), 217-240. https://doi.org/10.1111/1467-9817.00087

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*(1), 6-10. https://doi.org/10.1177/074193258600700104

GOV.UK. (2015). *English indices of deprivation*. https://www.gov.uk/government/collections/english-indices-of-deprivation

GOV.UK. (2019). *Compare schools*. https://www.compare-school-performance.service.gov.uk/compare-schools.

GOV.UK (2021). *£10 million scheme to help pupils boost core skills.* https://www.gov.uk/government/news/10-million-scheme-to-help-pupils-boost-core-skills

Graphpad. (2018, August 6). *Quickcalcs*. www.graphpad.com.

Graves, M. F. (2006). *The vocabulary book: Learning and instruction*. Columbia University.

Gray, S. (2005). Word learning by preschoolers with specific language impairment: Effect of phonological or semantic cues. *Journal of Speech, Language and Hearing Research, 48*(6), 1452-1467. https://doi.org/10.1044/1092-4388(2005/101)

Gupta, P. (2005). What's in a word? A functional analysis of word learning. *SIG 1 Perspectives on Language Learning and Education, 12*(3), 4-8.

Hadley, E. B., Dickinson, D. K., Hirsh-Pasek, K., Golinkoff, R. M., & Nesbitt, K. T. (2016). Examining the acquisition of vocabulary knowledge depth among preschool students. *Reading Research Quarterly, 51*(2), 181-198. https://doi.org/10.1002/rrq.130

Hair, J., Black, W. C., Babin, B., Anderson, R., & Tatham, R. (2014). *Pearson new international edition. Multivariate data analysis (7th ed.)*. Pearson Education

Hairrell, A., Rupley, W., & Simmons, D. (2011). The state of vocabulary research. *Literacy Research and Instruction, 50*(4), 253-271. https://doi.org/10.1080/19388071.2010.514036

Hammill, D. D., Pearson, N. A., Hresko, W. P., & Hoover, J. J. (2012). *Early Reading Assessment*. Pro-Ed.

Hammond, S. (2006). Using psychometric tests. In G. M. Breakwell, S. Hammond & C. Fife-Schaw (Eds.) *Research Methods in Psychology* (pp. 182-209). Sage.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106*(3), 491-528. https://doi.org/10.1037/0033-295X.106.3.491

Hart, B., & Risley, T. (2003). The early catastrophe: The 30 million word gap by age 3. *American Educator, 27*, 4-9.

Heibeck, T. H., & Markman, E. M. (1987). Word learning in children: An examination of fast mapping. *Child Development, 58*(4), 1021-1034. https://doi.org/10.2307/1130543

Hiebert, E. H., & Kamil, M. L. (2005). *Teaching and learning vocabulary: Bringing research to practice*. Routledge, Taylor & Francis Group.

Hogan, T. P. (2010). A short report: Word-level phonological and lexical characteristics interact to influence phoneme awareness. *Journal of Learning Disabilities, 43*(4), 346-356. https://doi.org/10.1177/0022219410369083

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention by 24-month-old infants. *Infancy, 13*(2), 128-157. https://doi.org/10.1080/15250000701795598

Howitt, D., & Cramer, D. (2014). *Introduction to research methods in psychology*. Pearson Education.

Huber, P. J. (1996). *Robust statistical procedures (2nd ed.)*. Siam.

Huitema, B. (2011). *The analysis of covariance and alternatives: Statistical methods for experiments, quasi-experiments and single-case studies (2nd ed.)*. John Wiley & Sons.

Hulme, C., Goetz, K., Gooch, D., Adams, J., & Snowling, M. J. (2007). Paired-associate learning, phoneme awareness, and learning to read. *Journal of Experimental Child Psychology, 96*(2), 150-166. https://doi.org/10.1016/j.jecp.2006.09.002

Hulme, C., Stothard, S.E., Clarke, P., Bowyer-Crane, C., Harrington, A., Truelove, E. & Snowling, M.J. (2009). *York Assessment of Reading for Comprehension (YARC): Early Reading*. GL Assessment.

Hulme, C., Nash, H. M., Gooch, D., Lervåg, A., & Snowling, M. J. (2015). The foundations of literacy development in children at familial risk of dyslexia. *Psychological Science, 26*(12), 1877-1886. https://doi.org/10.1177/0956797615603702

Hyde-Wright, S., Gorrie, B., Hayes, C., & Shipman, A. (1993). Teaching word-finding strategies to severely language-impaired children. *International Journal of Language and Communication Disorders, 28*(2), 165-175. https://doi.org/10.3109/13682829309041464

ICAN. (2007). *The human cost of children's poor communication*. ICAN.

ICAN. (2012). *Talk Boost Key Stage 1*. ICAN.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*(1), 101-144. https://doi.org/10.1016/j.cognition.2002.06.001

Jaccard, J. (1998). *Interaction effects in factorial analysis of variance*. Sage.

Janssen, C., Segers, E., McQueen, J. M., & Verhoeven, L. (2018). Comparing effects of instruction on word meaning and word form on early literacy abilities in kindergarten. *Early Education and Development, 30*(3), 375-399. https://doi.org/10.1080/10409289.2018.1547563

Joffe, V. (2011). *Vocabulary enrichment intervention programme*. Speechmark.

Joffe, V., Rixon, L., & Hulme, C. (2019). Improving storytelling and vocabulary in secondary school students with language disorder: A randomized controlled trial. *International Journal of Language and Communication Disorders, 54*(4), 656-672. https://doi.org/10.1111/1460-6984.12471

Johnson, T. R. (2016). Violation of the homogeneity of regression slopes assumption in ANCOVA for two-group pre-post designs: Tutorial on a modified Johnson-Neyman procedure. *The Quantitative Methods for Psychology, 12*(3), 253-263. https://doi.org/10.20982/tqmp.12.3.p253

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology., 29*(1), 1-23. https://doi.org/10.1006/cogp.1995.1010

Jusczyk, P. W., Houston, D. M., & Newsome, M. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology, 39*(3-4), 159-207. https://doi.org/10.1006/cogp.1999.0716

Justice, L., Chow, S.-M., Capellini, C., Flanigan, K., & Colton, S. (2003). Emergent literacy intervention for vulnerable preschoolers: Relative effects of two approaches. *American Journal of Speech-Language Pathology, 12*(3), 320-332. https://doi.org/10.1044/1058-0360(2003/078)

Justice, L., Meier, J., & Walpole, S. (2005). Learning new words from storybooks: An efficacy study with at-risk kindergartners. *Language, Speech and Hearing Services in Schools, 36*(1), 17-32. https://doi.org/10.1044/0161-1461(2005/003)

Kameenui, E. J., & Baumann, J. F. (2012). *Vocabulary instruction: research to practice (2nd ed.)*. The Guilford Press.

Kameenui, E. J., Dixon, R. C., & Carnine, D. W. (1987). Issues in the design of vocabulary instruction. In M. McKeown & M. Curtis (Eds.), *The nature of vocabulary acquisition* (pp. 129-145). Lawrence Erlbaum Associates.

Kamil, M. (2004). Vocabulary and comprehension instruction: Summary and implications of the National Reading Panel. In P. McCardle & V. Chhabra (Eds.), *The voice of evidence in reading research* (pp. 213-234). Paul H. Brookes Publishing Co.

Kang, S. H. K. (2016). Spaced repetition promotes efficient and effective learning: Policy implications for instruction. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 12-19. https://doi.org/10.1177/2372732215624708

Karpicke, J. D., Blunt, J. R., & Smith, M. A. (2016). Retrieval-based learning: Positive effects of retrieval practice in elementary school children. *Frontiers in Psychology*, 7, 350–350. https://doi.org/10.3389/fpsyg.2016.00350

Kim, H.Y. (2013). Statistical notes for clinical researchers: assessing normal distribution using skewness and kurtosis. *Restorative Dentistry and Endodontics*, 38(1), 52-54. https://doi.org/10.5395/rde.2013.38.1.52

Kim, Y.-S., Apel, K., & Al Otaiba, S. (2013). The relation of linguistic awareness and vocabulary to word reading and spelling for first-grade students participating in response to intervention. *Language, Speech and Hearing Services in Schools, 44*(4), 337-347. https://doi.org/10.1044/0161-1461(2013/12-0013)

King, T. M., Rosenberg, L. A., Fuddy, L., McFarlane, E., Sia, C., & Duggan, A. K. (2005). Prevalence and early identification of language delays among at-risk three year olds. *Journal of Developmental and Behavioral Pediatrics, 26*(4), 293-303.

Kline, P. (1998). *The new psychometrics: Science, psychology and measurement*. Psychology Press.

Kucera, H., & Francis, W.N. (1967). *Computational analysis of present-day American English*. Brown University Press.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44*(4), 978-990. https://doi.org/10.3758/s13428-012-0210-4

Kurdek, L. A., & Sinclair, R. J. (2001). Predicting reading and mathematics achievement in fourth-grade children from kindergarten readiness scores. *Journal of Educational Psychology, 93*(3), 451-455. https://doi.org/10.1037/0022-0663.93.3.451

Laing, E., & Hulme, C. (1999). Phonological and semantic processes influence beginning readers' ability to learn to read words. *Journal of Experimental Child Psychology, 73*(3), 183-207. https://doi.org/0.1006/jecp.1999.2500

Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: The construct of task-induced involvement. *Applied Linguistics, 22*(1), 1-26. https://doi.org/10.1093/applin/22.1.1

Law, J., Rush, R., Schoon, I., & Parsons, S. (2009). Modeling developmental language difficulties from school entry into adulthood: Literacy, mental health, and employment outcomes. *Journal of Speech Language and Hearing Research*, 52(6), 1401-1416. https://doi.org/10.1044/1092-4388(2009/08-0142)

Lawson-Adams, J., & Dickinson, D. K. (2020). Building lexical representations with nonverbal supports. *Reading Research Quarterly, 56*(3), 603-622. https://doi.org/10.1002/rrq.326

Lee, J. (2011). Size matters: Early vocabulary as a predictor of language and literacy competence. *Applied Psycholinguistics, 32*(01), 69-92. https://doi.org/10.1017/s0142716410000299

Leppink, J. (2018). Analysis of Covariance (ANCOVA) vs. moderated regression (MODREG): Why the interaction matters. *Health Professions Education, 4*(3), 225-232. https://doi.org/https://doi.org/10.1016/j.hpe.2018.04.001

Lerner, M. D., & Lonigan, C. J. (2016). Bidirectional relations between phonological awareness and letter knowledge in preschool revisited: A growth curve analysis of the relation between two code-related skills. *Journal of Experimental Child Psychology, 144*, 166-183. https://doi.org/10.1016/j.jecp.2015.09.023

Lervåg, A., Bråten, I., & Hulme, C. (2009). The cognitive and linguistic foundations of early reading development: A Norwegian latent variable longitudinal study. *Developmental Psychology, 45*(3), 764-781. https://doi.org/10.1037/a0014132

Levlin, M., Wiklund-Hornqvist, C., Sandgren, O., Karlsson, S., & Jonsson, B. (2022). Evaluating the effect of rich vocabulary instruction and retrieval practice on the classroom vocabulary skills of children with (Developmental) Language Disorder. *Language, Speech & Hearing Services in Schools,* 53(2), 542–560. https://doi.org/10.1044/2021_LSHSS-21-00101

Lewis-Beck, M., Bryman, A., & Liao, T. (2004). *The SAGE encyclopedia of social science research methods*. SAGE. https://doi.org/10.4135/9781412950589

Liberman, I. Y., Shankweiler, D., & Liberman, A. M. (1989). *The alphabetic principle and learning to read*. US Dept of Health and Human Services.

Literacy Trust. (2017). *Developing a whole-school approach to teaching vocabulary in primary schools.* https://literacytrust.org.uk/resources/developing-whole-school-approach-teaching-vocabulary-primary-schools/.

Locke, A., Ginsborg, J., & Peers, I. (2002). Development and disadvantage: Implications for the early years and beyond. *International Journal of Language and Communication Disorders, 37*(1), 3-15. https://doi.org/10.1080/1368282011008991

Lohman, D. F., & Hagen, E. P. (2001). *Cognitive Abilities Test.* Riverside Publisher.

Lonigan, C. (2007). Vocabulary development and the development of phonological awareness skills in preschool children. In R. K. Wagner, A. E. Muse, & K. R. Tannenbaum (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 15-31). Guilford Press.

Lonigan, C. J., Burgess, S. R., Anthony, J. L., & Barker, T. A. (1998). Development of phonological sensitivity in 2- to 5-year- old children. *Journal of Educational Psychology, 90*(2), 294-311. https://doi.org/10.1037/0022-0663.90.2.294

Lowe, H. & Joffe, V. (2017) Exploring the feasibility of a classroom-based vocabulary intervention for mainstream secondary school students with language disorder. *Support for Learning*, 32(2), 110–128. https: doi:10.1111/1467-9604.12157

Lowe, H., Henry, L., & Joffe, V. (2019). The effectiveness of classroom vocabulary intervention for adolescents with language disorder. *Journal of Speech Language and Hearing Research, 62*(8), 2829-2846. https://doi.org/10.1044/2019_JSLHR-L-18-0337

Lowe, H., Henry, L., Wallinger, J., & Joffe, V. (2022). Teaching vocabulary to adolescents with language disorder: Perspectives from teachers and speech and language therapists. *Child Language Teaching and Therapy*, 38(1), 95–114. https://doi.org/10.1177/02656590211064541

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear Hear, 19*(1), 1-36. https://doi.org/10.1097/00003446-199802000-00001

Lundberg, I., Olofsson, Å., & Wall, S. (1980). Reading and spelling skills in the first school years predicted from phonemic awareness skills in kindergarten. *Scandinavian Journal of Psychology, 21*(1), 159-173. https://doi.org/10.1080/10888438.2019.1615491

Maratsos, M. (1990). Are actions to verbs as objects are to nouns? On the differential semantic bases of form, class, category. *Linguistics, 28*(6), 1351-1380. https://doi.org/doi:10.1515/ling.1990.28.6.1351

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science, 14*(1), 57-77. https://doi.org/10.1207/s15516709cog1401_4

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology, 20*(2), 121-157. https://doi.org/10.1016/0010-0285(88)90017-5

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1), 71-102. https://doi.org/10.1016/0010-0277(87)90005-9

Marulis, L. M., & Neuman, S. B. (2010). The effects of vocabulary intervention on young children's word learning: A meta-analysis. *Review of Educational Research, 80*(3), 300-335. https://doi.org/10.3102/0034654310377087

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*(2), 91-121. https://doi.org/10.1016/S0010-0277(00)00109-8

Maynard, K. L., Pullen, P. C., & Coyne, M. D. (2010). Teaching vocabulary to first-grade students through repeated shared storybook reading: A comparison of rich and basic instruction to incidental exposure. *Literacy Research and Instruction*, 49(3), 209–242. https://doi.org/10.1080/19388070902943245

McBride-Chang, C., Wagner, R. K., & Chang, L. (1997). Growth modeling of phonological awareness. *Journal of Educational Psychology, 89*(4), 621-630. https://doi.org/10.1037/0022-0663.89.4.621

McDermott, K. B. (2021). Practicing retrieval facilitates learning. *Annual Review of Psychology*, 72(*1*), 609–633. https://doi.org/10.1146/annurev-psych-010419-051019

McGregor, K. K. (1994). Use of phonological information in a word-finding treatment for children. *Journal of Speech and Hearing Research, 37*(6), 1381-1393 https://doi.org/10.1044/jshr.3706.1381

McGregor, K. K., Van Horne, A. O., Curran, M., Cook, S. W., & Cole, R. (2021). The challenge of rich vocabulary instruction for children with Developmental Language Disorder. *Language, Speech & Hearing Services in Schools*, 52(2), 467–484. https://doi.org/10.1044/2020_LSHSS-20-00110

McKague, M., Pratt, C., & Johnston, M. B. (2001). The effect of oral vocabulary on reading visually novel words: a comparison of the dual-route-cascaded and triangle frameworks. *Cognition, 80*(3), 231-262. https://doi.org/10.1016/S0010-0277(00)00150-5

McNeill, B. C., & Hesketh, A. (2010). Developmental complexity of the stimuli included in mispronunciation detection tasks. *International Journal of Language and Communication Disorders, 45*(1), 72-82. https://doi.org/10.3109/13682820902745479

Metsala, J. L. (1997). Spoken word recognition in reading disabled children. *Journal of Educational Psychology, 89*(1), 159-169. https://doi.org/10.1037/0022-0663.89.1.159

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*(1), 3-19. https://doi.org/10.1037/0022-0663.91.1.3

Metsala, J. L., Stavrinos, D., & Walley, A. C. (2009). Children's spoken word recognition and contributions to phonological awareness and nonword repetition: A 1-year follow-up. *Applied Psycholinguistics, 30*(1), 101-121. https://doi.org/10.1017/S014271640809005X

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J.

Metsala & L. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89-120). Lawrence Erlbaum Associates.

Mitchell, A., & Brady, S. (2013). The effect of vocabulary knowledge on novel word identification. *An Interdisciplinary Journal of The International Dyslexia Association, 63*(3), 201-216. https://doi.org/10.1007/s11881-013-0080-1

Mol, S. E., Bus, A. G., & de Jong, M. T. (2009). Interactive book reading in early education: A tool to stimulate print knowledge as well as oral language. *Review of Educational Research, 79*(2), 979-1007. https://doi.org/10.3102/0034654309332561

Morais, J., Bertelson, P., Cary, L., & Alegria, J. (1986). Literacy training and speech segmentation. *Cognition, 24*(1-2), 45-64. https://doi.org/10.1016/0010-0277(86)90004-1.

Moran, E., & Moir, J. (2018). Closing the vocabulary gap in early years: Is' Word Aware' a possible approach? *Educational and Child Psychology, 35*(1), 51-64.

Munro, N., Lee, K., & Baker, E. (2008). Building vocabulary knowledge and phonological awareness skills in children with specific language impairment through hybrid language intervention: A feasibility study. *International Journal of Language and Communication Disorders, 43*(6), 662-682. https://doi.org/10.1080/13682820701806308

Murphy, A., Franklin, S., Breen, A., Hanlon, M., McNamara, A., Bogue, A., & James, E. (2017). A whole class teaching approach to improve the vocabulary skills of adolescents attending mainstream secondary school, in areas of socioeconomic disadvantage. *Child Language Teaching and Therapy, 33*(2), 129-144. https://doi.org/10.1177/0265659016656906

Murphy, M. L. (2010). *Lexical meaning*. Cambridge University Press.

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. https://doi.org/10.1037/0012-1649.40.5.665

Nagy, W. (2005). Why vocabulary instruction needs to be long-term and comprehensive. In E. H. Hiebert & M. L. Kamil (Eds.), *Teaching and learning vocabulary: Bringing research to practice* (pp. 27-44). Routledge, Taylor & Francis

Nagy, W. E., Herman, P. A., & Anderson, R. C. (1985). Learning words from context. *Reading Research Quarterly, 20*(2), 233-253. https://doi.org/10.2307/747758

Nash, M., & Donaldson, M. L. (2005). Word learning in children with vocabulary deficits. *Journal of Speech, Language and Hearing Research, 48*(2), 439-458. https://doi.org/10.1044/1092-4388(2005/030)

NASUWT. (2018). *Taking control of your performance management: A practical guide for teachers*. National Association of Schoolmasters Union of Women Teachers.

Nation, I., & Webb, S. A. (2011). *Researching and analyzing vocabulary*. Cengage Learning.

Nation, K., & Cocksey, J. (2009). The relationship between knowing a word and reading it aloud in children's word reading development. *Journal of Experimental Child Psychology, 103*(3), 296-308. https://doi.org/10.1016/j.jecp.2009.03.004

Nation, K., & Snowling, M. J. (1998). Semantic processing and the development of word-recognition skills: Evidence from children with reading comprehension difficulties. *Journal of Memory and Language, 39*(1), 85-101. https://doi.org/10.1006/jmla.1998.2564

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: Broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342-356. https://doi.org/10.1111/j.1467-9817.2004.00238.x

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. NRP.

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal, 46*(2), 532-566. https://doi.org/10.3102/0002831208328088

NFER (National Foundation for Educational Research). (2013). *Evaluation of the Phonics Screening Check: First Interim Report. May 2013*. Department for Education.

Nittrouer, S., & Studdert-Kennedy, M. (1987). The role of coarticulatory effects in the perception of fricatives by children and adults. *Journal of Speech and Hearing Research, 30*(3), 319-329. https://doi.org/10.1044/jshr.3003.319

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

Oetting, J. B., Rice, M. L., & Swank, L. K. (1995). Quick incidental learning (QUIL) of words by school-age children with and without SLI. *Journal of Speech, Language and Hearing Research, 38*(2), 434-445. https://doi.org/10.1044/jshr.3802.434

OFSTED (Office for Standards in Education). (2015). *School reports*. https://reports.ofsted.gov.uk.

Ortiz, M., Folsom, J. S., Al Otaiba, S., Greulich, L., Thomas-Tate, S., & Connor, C. M. (2012). The componential model of reading: Predicting first grade reading performance of culturally diverse students from ecological, psychological, and cognitive factors assessed at kindergarten entry. *Journal of learning Disabilities, 45*(5), 406-417. https://doi.org/10.1177/0022219411431242

Ouellette, G. P. (2006). What's meaning got to do with it: The role of vocabulary in word reading and reading comprehension. *Journal of Educational Psychology, 98*(3), 554-566. https://doi.org/10.1037/0022-0663.98.3.554

Ouellette, G., & Beers, A. (2010). A not-so-simple view of reading: how oral vocabulary and visual-word recognition complicate the story. *Reading and Writing: An Interdisciplinary Journal, 23*(2), 189-208. https://doi.org/10.1007/s11145-008-9159-1

Ouellette, G., & Fraser, J. R. (2009). What exactly is a yait anyway: The role of semantics in orthographic learning. *Journal of Experimental Child Psychology, 104*(2), 239-251. https://doi.org/10.1016/j.jecp.2009.05.001

Oxford University Press. (2018). *Why closing the word gap matters: Oxford language report*. Oxford University Press.

Pallant, J. (2013). *SPSS survival manual*. McGraw-Hill Education

Pallante, D. H., & Kim, Y.-S. (2013). The effect of a multicomponent literacy instruction model on literacy growth for kindergartners and first-grade students in Chile. *International Journal of Psychology, 48*(5), 747-761. https://doi.org/10.1080/00207594.2012.719628

Parsons, S., & Branagan, A. (2014). *Word Aware: Teaching vocabulary across the day, across the curriculum*. Speechmark.

Parsons, S., Law, J., & Gascoigne, M. (2005). Teaching receptive vocabulary to children with specific language impairment: A curriculum-based approach. *Child Language Teaching and Therapy, 21*(1), 39-59. https://doi.org/10.1191/0265659005ct280oa

Pavlik, P. I., & Anderson, J. R. (2005). Practice and forgetting effects on vocabulary memory: an activation-based model of the spacing effect. *Cognitive Science, 29*(4), 559-586. https://doi.org/10.1207/s15516709cog0000_14

Penno, J. F., Wilkinson, I. A. G., & Moore, D. W. (2002). Vocabulary acquisition from teacher explanation and repeated listening to stories: Do they overcome the Matthew effect? *Journal of Educational Psychology, 94*(1), 23. https://doi.org/10.1037/0022-0663.94.1.23

Perfetti, C. (1985). *Reading ability*. Oxford University Press.

Perfetti, C., & Hart, L. (2002). The lexical quality hypothesis. In L. Verhoeven, C. Elbro, & P. Reitsma (Eds.), *Precursors of functional literacy* (Vol. 11, pp. 189-213). John Benjamins Publishing.

Perfetti, C. A., Beck, I., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal: A longitudinal study of first grade children. *Merrill-Palmer Quarterly 33*(3), 283-319.

Perin, D. (1983). Phonemic segmentation and spelling. *British Journal of Psychology, 74*(1), 129-144. https://doi.org/10.1111/j.2044-8295.1983.tb01849.x

Pollard-Durodola, S. D., Gonzalez, J. E., Simmons, D. C., Kwok, O., Taylor, A. B., Davis, M. J., Kim, M., & Simmons, L. (2011). The effects of an intensive shared book-reading intervention for preschool children at risk for vocabulary delay. *Exceptional Children, 77*(2), 161-183. https://doi.org/10.1177/001440291107700202

Porta, M. E., & Ramirez, G. (2019). The impact of an early intervention on vocabulary, phonological awareness, and letter–sound knowledge among Spanish-speaking kindergarteners. *International Journal of School and Educational Psychology, 8*(1), 65-79. https://doi.org/10.1080/21683603.2018.1558137

Random.org. (2016, October 16; 2019, February 12). *Sequences*. https://www.random.org/sequences/?min=1&max=100&col=1&format=html&rnd=new.

Rapp, B., & Goldrick, M. (2000). Discreteness and interactivity in spoken word production. *Psychological Review, 107*(3), 460-499. https://doi.org/10.1037/0033-295X.107.3.460

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods. Volume 1*. Sage.

Raudenbush, S. W., & Schwartz, D. (2020). Randomized experiments in education, with implications for multilevel causal inference. *Annual Review of Statistics and Its Application, 7*, 177-208. https://doi.org/10.1146/annurev-statistics-031219-041205

Renfrew, C. E. (1988). *The Renfrew battery of language assessments (Revised)*. C.E. Renfrew.

Rice, M. L., Buhr, J. C., & Nemeth, M. (1990). Fast mapping word-learning abilities of language-delayed preschoolers. *Journal of Speech and Hearing Disorders, 55*(1), 33-42. https://doi.org/10.1044/jshd.5501.33

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. https://doi.org/10.1080/10888430701344306

Ripley, K., & Yuill, N. (2005). Patterns of language impairment and behaviour in boys excluded from school. *British Journal of Educational Psychology, 75*(1), 37-50. https://doi.org/doi:10.1348/000709905X27696

Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(1), 29-33. https://doi.org/10.1037/0278-7393.28.1.29

Rosenthal, J., & Ehri, L. C. (2011). Pronouncing new words aloud during the silent reading of text enhances fifth graders' memory for vocabulary words and their spellings. *Reading and Writing, 24*(8), 921-950. https://doi.org/10.1007/s11145-010-9239-x

Roulstone, S., Law, J, Rush, R, Clegg, J and Peters, T. (2011). *Investigating the role of language in children's early educational outcomes. Research report DFE-RR134.* Department for Education.

Russell, G., Ukoumunne, O. C., Ryder, D., Golding, J., & Norwich, B. (2018). Predictors of word-reading ability in 7-year-olds: Analysis of data from a U.K. cohort study. *Journal of Research in Reading, 41*(1), 58-78. https://doi.org/10.1111/1467-9817.12087

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926-1928. https://doi.org/10.1126/science.274.5294.1926.

Savage, R., Carless, S. U. E., & Erten, O. (2009). The longer-term effects of reading interventions delivered by experienced teaching assistants. *Support for Learning*, 24(2), 95-100. https://doi.org/10.1111/j.1467-9604.2009.01405.x

Scarborough, H. S. (2001). Connecting early language and literacy to later reading (dis) abilities: Evidence, theory, and practice. In S. Neuman & D. Dickinson (Eds.), *Handbook of early literacy research* (pp. 97-110). Guildford Press.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology, 96*(2), 265-282. https://doi.org/10.1037/0022-0663.96.2.265

Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Palgrave Macmillan.

Schoon, I., Parsons, S., Rush, R., & Law, J. (2010). Children's language ability and psychosocial development: A 29-year follow-up study. *Pediatrics, 126*(1), e73-e80. https://doi.org/10.1542/peds.2009-3282

Scott, J. A., Jamieson-Noel, D., & Asselin, M. (2003). Vocabulary instruction throughout the day in twenty-three Canadian upper-elementary classrooms. *The Elementary School Journal, 103*(3), 269. https://doi.org/10.1086/499726

Semel, E., Wiig, E. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals (4th ed.)*. The Psychological Corporation.

Sénéchal, M., & LeFevre, J. A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73*(2), 445-460. https://doi.org/0.1111/1467-8624.00417

Sénéchal, M., Ouellette, G., & Rodney, D. (2006). The misunderstood giant: On the predictive role of early vocabulary to future reading. In D. Dickinson & S. Neuman (Eds.), *Handbook of Early Literacy Research* (Vol. 2, pp. 173-182). Guilford.

Silverman, R., & Crandell, J. D. (2010). Vocabulary practices in prekindergarten and kindergarten classrooms. *Reading Research Quarterly, 45*(3), 318-340. https://doi.org/10.1598/rrq.45.3.3

Silverman, R. (2007). Vocabulary development of English-language and English-only learners in kindergarten. *The Elementary School Journal, 107*(4), 365-383. https://doi.org/10.1086/516669

Slavin, R. E. (2008). Perspectives on evidence-based research in education—what works? Issues in synthesizing educational program evaluations. *Educational Researcher, 37*(1), 5-14. https://doi.org/10.3102/0013189X08314117

Smith, F. R. H., Gaskell, M. G., Weighall, A. R., Warmington, M., Reid, A. M., & Henderson, L. M. (2017). Consolidation of vocabulary is associated with sleep in typically developing children, but not in children with dyslexia. *Developmental Science, 21*(5), 1-16. https://doi.org/10.1111/desc.12639

Snowling, M. J., & Hulme, C. (2021). Annual research review: Reading disorders revisited – the critical importance of oral language. *Journal of Child Psychology and Psychiatry*, 62(5), 635–653. https://doi.org/10.1111/jcpp.13324

Spencer, S., Clegg, J., Lowe, H., & Stackhouse, J. (2017). Increasing adolescents' depth of understanding of cross-curriculum words: an intervention study: Evaluating adolescent cross-curriculum vocabulary intervention. *International Journal of Language and Communication Disorders, 52*(5), 652-668. https://doi.org/10.1111/1460-6984.12309

Spencer, S., Clegg, J., & Stackhouse, J. (2012). Language and disadvantage: A comparison of the language abilities of adolescents from two different socioeconomic areas. *International Journal of Language and Communication Disorders, 47*(3), 274-284. https://doi.org/10.1111/j.1460-6984.2011.00104.x

St John, P., & Vance, M. (2014). Evaluation of a principled approach to vocabulary learning in mainstream classes. *Child Language Teaching and Therapy, 30*(3), 255-271. https://doi.org/10.1177/0265659013516474

Stackhouse, J., Vance, M, Pascoe, M, Wells, B. (2007). *Compendium of auditory and speech tasks: Children's speech and literacy difficulties 4*. John Wiley & Sons.

Stackhouse, J., & Wells, B. (1997). *Children’s speech and literacy difficulties: A psycholinguistic framework*. Whurr.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56*(1), 72-110. https://doi.org/10.3102/00346543056001072

Standards and Testing Agency. (2012). *Year One Phonics Screening Check. Pilot: 2011 technical report*. Department for Education.

Steele, S. C., Willoughby, L. M., & Mills, M. T. (2013). Learning word meanings during reading: Effects of phonological and semantic cues on children with language impairment. *International Journal of Speech and Language Pathology, 15*(2), 184-197. https://doi.org/10.3109/17549507.2012.700322

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934-947. https://doi.org/10.1037//0012-1649.38.6.934

Storkel, H., Armbrüster, J., & Hogan, T. (2006). Differentiating phonotactic probability and neighborhood density in adult word learning. *Journal of Speech, Language, and Hearing Research, 49*(6), 1175-1192. https://doi.org/10.1044/1092-4388(2006/085)

Storkel, H., & Morrisette, M. (2002). The lexicon and phonology: Interactions in language acquisition. *Language, Speech and Hearing Services in Schools, 33*(1), 24-36. https://doi.org/10.1044/0161-1461(2002/003

Storkel, H., & Rogers, M. (2000). The effect of probabilistic phonotactics on lexical acquisition. *Clinical Linguistics and Phonetics, 14*(6), 407-425. https://doi.org/10.1080/026992000415859

Storkel, H. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech Language and Hearing Research, 44*(6), 1321-1337. https://doi.org/10.1044/1092-4388(2001/103)

Storkel, H. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics, 25*(2), 201-221. https://doi.org/10.1017/S0142716404001109

Storkel, H., Voelmle, K., Fierro, V., Flake, K., Fleming, K. K., & Romine, R. S. (2017). Interactive book reading to accelerate word learning by kindergarten children with Specific Language Impairment: Identifying an adequate intensity and variation in treatment response. *Language, Speech and Hearing Services in Schools, 48*(1), 16-30. https://doi.org/10.1044/2016_LSHSS-16-0014

Stothard, S., Snowling, M., Bishop, D., Chipchase, B., & Kaplan, C. (1998). Language-impaired preschoolers: A follow-up into adolescence. *Journal of Speech, Language and Hearing Research, 41*(2), 407-418. https://doi.org/10.1044/jslhr.4102.407

Swan, D., & Goswami, U. (1997). Phonological awareness deficits in developmental dyslexia and the phonological representations hypothesis. *Journal of Experimental Child Psychology, 66*(1), 18-41. https://doi.org/10.1006/jecp.1997.2375

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. *Cognition, 76*(2), 147-166. https://doi.org/10.1016/S0010-0277(00)00081-0

Thomson, J. M., Richardson, U., & Goswami, U. (2005). Phonological similarity neighborhoods and children's short-term memory: Typical development and dyslexia. *Memory and Cognition, 33*(7), 1210-1219. https://doi.org/10.3758/BF03193223

Torgesen, J. K., & Davis, C. (1996). Individual difference variables that predict response to training in phonological awareness. *Journal of Experimental Child Psychology, 63*(1), 1-21. https://doi.org/10.1006/jecp.1996.0040

Treiman, R., & Breaux, A. M. (1982). Common phoneme and overall similarity relations among spoken syllables: Their use by children and adults. *Journal of Psycholinguistic Research, 11*(6), 569-598. https://doi.org/10.1007/BF01067613

Treiman, R., & Zukowski, A. (1996). Children's sensitivity to syllables, onsets, rimes, and phonemes. *Journal of Experimental Child Psychology, 61*, 193-215. https://doi.org/10.1006/jecp.1996.0014

Treiman, R., Gordon, J., Boada, R., Peterson, R. L., & Pennington, B. F. (2014). Statistical learning, letter reversals, and reading. *Scientific Studies of Reading*, 18(6), 383-394. https://doi.org/10.1080/10888438.2013.873937

Tunmer, W. E., & Chapman, J. W. (2012). Does set for variability mediate the influence of vocabulary knowledge on the development of word recognition skills? *Scientific Studies of Reading, 16*(2), 122-140. https://doi.org/10.1080/10888438.2010.542527

Turner, J. E., Henry, L. A., Smith, P. T., & Brown, P. A. (2004). Redintegration and lexicality effects in children: do they depend upon the demands of the memory task? *Memory and Cognition, 32*(3), 501-510. https://doi.org/10.3758/BF03195842

Vadasy, P. F., & Sanders, E. A. (2016). Attention to orthographic and phonological word forms in vocabulary instruction for kindergarten English learners. *Reading Psychology., 37*(6), 833-866. https://doi.org/10.1080/02702711.2015.1116477

Verhoeven, L., van Leeuwe, J., & Vermeer, A. (2011). Vocabulary growth and reading development across the elementary school years. *Scientific Studies of Reading., 15*(1), 8-25. https://doi.org/10.1080/10888438.2011.536125

Vickers, A. J., & Altman, D. G. (2001). Analysing controlled trials with baseline and follow up measurements. *British Medical Journal, 323*, 1123-1124. https://doi.org/10.1136/bmj.323.7321.1123

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language, 68*(1-2), 306-311. https://doi.org/10.1006/brln.1999.2116

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading skills. *Psychological Bulletin*, 101(2), 192. https://doi.org/10.1037/0033-2909.101.2.192

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., Donahue, J., & Garon, T. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33*(3), 468-479. https://doi.org/10.1037/0012-1649.33.3.468

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., & Pearson, N. A. (2013). *Comprehensive Test of Phonological Processing (2nd ed.)*. Pro-Ed.

Waldfogel, J., & Washbrook, E. (2010). *Low income and early cognitive development in the UK*. Sutton Trust.

Walley, A., Metsala, J., & Garlock, V. (2003). Spoken vocabulary growth: Its role in the development of phoneme awareness and early reading ability. *Reading and Writing: An Interdisciplinary Journal, 16*(1), 5-20. https://doi.org/10.1023/A:1021789804977

Walley, A. C. (1993). The role of vocabulary development in children's spoken word recognition and segmentation ability. *Developmental Review, 13*, 286-350. https://doi.org/10.1006/drev.1993.1

Walley, A. C., & Metsala, J. L. (1990). The growth of lexical constraints on spoken word recognition. *Attention, Perception and Psychophysics, 47*(3), 267-280. https://doi.org/10.3758/bf03205001

Wasik, B. A., & Hindman, A. H. (2011). Improving vocabulary and pre-literacy skills of at-risk preschoolers through teacher professional development. *Journal of Educational Psychology, 103*(2), 455-469. https://doi.org/10.1037/a0023067

Watts, S. M. (1995). Vocabulary instruction during reading lessons in six classrooms. *Journal of Reading Behavior, 27*(3), 399-424. https://doi.org/10.1007/s10643-015-0762-x

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics, 28*(2), 46-65. https://doi.org/10.1093/applin/aml048

Wegener, S., Beyersmann, E., Wang, H., & Castles, A. (2022). Oral vocabulary knowledge and learning to read new words: A theoretical review. *Australian Journal of Learning Difficulties*, Online first, 1–26. https://doi.org/10.1080/19404158.2022.2097717

Werker, J. F., & Curtin, S. (2005). Primir: A developmental framework of infant speech processing. *Language Learning and Development, 1*(2), 197-234. https://doi.org/10.1080/15475441.2005.9684216

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy, 3*(1), 1-30. https://doi.org/10.1207/S15327078IN0301_1

Werker, J. F., & Tees, R. C. (1987). Speech perception in severely disabled and average reading children. *Canadian Journal of Psychology, 41*(1), 48-61. https://doi.org/10.1037/h0084150

Werker, J. F., & Tees, R. C. (1999). Influences on infant speech processing: Toward a new synthesis. *Annual Review of Psychology, 50*, 509-535. https://doi.org/10.1146/annurev.psych.50.1.509

West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and applications* (pp. 56-75). Sage.

Whiteley, H. E., Smith, C. D., & Connors, L. (2007). Young children at risk of literacy difficulties: Factors predicting recovery from risk following phonologically based intervention. *Journal of Research in Reading, 30*(3), 249-269. https://doi.org/10.1111/j.1467-9817.2007.00342.x

Widgit. (2007*). WidgitOnline. Widgit Multimedia.* www.widgit.com.

Williams, K. (2007). *Expressive Vocabulary Test (2nd ed.)*. Pearson.

Wilson, M. (1988). The MRC psycho-linguistic database: machine readable dictionary, version 2. *Behavioural Research Methods, Instruments and Computers, 20*(1), 6-11.

Wing, C. S. (1990). A preliminary investigation of generalization to untrained words following two treatments of children's word-finding problems. *Language Speech and Hearing Services in Schools, 21*(3), 151-156. https://doi.org/10.1044/0161-1461.2103.151

Winstanley, M., Webb, R. T., & Conti-Ramsden, G. (2018). More or less likely to offend? Young adults with a history of identified developmental language disorders. *International Journal of Language and Communication Disorders, 53*(2), 256-270. https://doi.org/10.1111/1460-6984.12339

Wisenburn, B., & Mahoney, K. (2009). A meta-analysis of word-finding treatments for Aphasia. *Aphasiology, 23*(11), 1338-1352. https://doi.org/10.1080/02687030902732745

Wordsmyth. (2014, January 24). *Children's dictionary*. http://www.wordsmyth.net.

Yazici, B., & Yolacan, S. (2007). A comparison of various tests of normality. *Journal of Statistical Computation and Simulation, 77*(2), 175-183. https://doi.org/10.1080/10629360600678310

Yopp, H. (1988). The validity and reliability of phonemic awareness tests. *Reading Research Quarterly, 23*(2), 159-177. https://doi.org/10.2307/747800

Zens, N. K., Gillon, G. T., & Moran, C. (2009). Effects of phonological awareness and semantic intervention on word-learning in children with SLI. *International Journal of Speech and Language Pathology, 11*(6), 509-524. https://doi.org/10.3109/17549500902926881

Zimmerman, D. (2004). A note on preliminary tests of equality of variances. *British Journal of Mathematical and Statistical Psychology, 57*, 173-181. https://doi.org/10.1348/000711004849222

Appendices

Appendix A

Ethical Approval for the Cross-Sectional Study

The University Of Sheffield.

Downloaded: 09/08/2021
Approved: 20/09/2016

Rosemarie Brooks
Registration number: 150111506
Human Communication Sciences
Programme: Department for Human Communication Sciences

Dear Rosemarie

PROJECT TITLE: Trial of an Assessment Tool to Measure Phonological Representations
APPLICATION: Reference Number 009173

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 20/09/2016 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 009173 (form submission date: 20/09/2016); (expected project end date: 22/07/2017).
- Participant information sheet 1018653 version 5 (20/09/2016).
- Participant information sheet 1018652 version 2 (18/08/2016).
- Participant consent form 1018654 version 2 (18/08/2016).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

John Mason
Ethics Administrator
Human Communication Sciences

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy: https://www.sheffield.ac.uk/rs/ethicsandintegrity/ethicspolicy/approval-procedure
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Administrator (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.

Ethical Approval for the Vocabulary Intervention Study

The University Of Sheffield.

Downloaded: 09/08/2021
Approved: 06/07/2018

Rosemarie Brooks
Registration number: 150111506
Human Communication Sciences
Programme: PhD

Dear Rosemarie

PROJECT TITLE: A comparison of combined sound-meaning vocabulary instruction versus semantic instruction on vocabulary, phonemic awareness, nonword reading and nonword spelling in 5-6 year olds: a randomised control trial
APPLICATION: Reference Number 019719

On behalf of the University ethics reviewers who reviewed your project, I am pleased to inform you that on 06/07/2018 the above-named project was **approved** on ethics grounds, on the basis that you will adhere to the following documentation that you submitted for ethics review:

- University research ethics application form 019719 (form submission date: 01/07/2018); (expected project end date: 30/07/2020).
- Participant information sheet 1045479 version 3 (01/07/2018).
- Participant consent form 1045482 version 4 (01/07/2018).

If during the course of the project you need to deviate significantly from the above-approved documentation please inform me since written approval will be required.

Your responsibilities in delivering this research project are set out at the end of this letter.

Yours sincerely

Traci Walker
Ethics Administrator
Health Sciences School

Please note the following responsibilities of the researcher in delivering the research project:

- The project must abide by the University's Research Ethics Policy: https://www.sheffield.ac.uk/rs/ethicsandintegrity/ethicspolicy/approval-procedure
- The project must abide by the University's Good Research & Innovation Practices Policy: https://www.sheffield.ac.uk/polopoly_fs/1.671066!/file/GRIPPolicy.pdf
- The researcher must inform their supervisor (in the case of a student) or Ethics Administrator (in the case of a member of staff) of any significant changes to the project or the approved documentation.
- The researcher must comply with the requirements of the law and relevant guidelines relating to security and confidentiality of personal data.
- The researcher is responsible for effectively managing the data collected both during and after the end of the project in line with best practice, and any relevant legislative, regulatory or contractual requirements.

Appendix B

Test Descriptions

B.1 Overview

Psychometric information on reliability and convergent validity is provided for each test used in the cross-sectional and experimental studies.

B.2 Reliability and Validity

The reliability of a test refers to how likely it is to produce consistent scores (Hammond, 2006). *Alpha reliability* takes into account all possible ways of splitting the scale to indicate the level of consistency between all items. Alpha is reported as a correlation between 0 and 1. For research purposes, an alpha of .70 is considered acceptable, .80 is good and .90 is excellent (ibid). A further frequently reported metric of reliability is the standard error of measurement, which provides an estimate of the error component within the scale (ibid). The participant's actual score minus the standard error of measurement would indicate their true score band. Standard error of measurement is reported in the same units as the test. The smaller the error, the more reliability can be attributed to the test scale.

Construct validity measures the extent to which the scale is representative of the domain it is intended to cover (ibid). One relevant subtype is *convergent validity* to demonstrate how well two measures capture the same construct. According to research by Carlson and Herdman (2010), convergent validity of .70 and above is optimal for research purposes, with the further recommendation that correlations of less than .50 should be avoided.

B.3 Tests Used in the Cross-Sectional Study

B.3.1 BPVS3 (British Picture Vocabulary Scale, 3rd Edition)

The BPVS3 (Dunn et al., 2009) is a receptive vocabulary test standardised in the UK in 2008 with 2027 pupils from 105 schools. The manual for the BPVS3 reports reliability in terms of the standard error of measurement linked to raw score bands and age (see Table B.1). It states that it was not possible to calculate Cronbach's alpha reliability because children engage with different test items within their critical range. Convergent validity was calculated in relation to the verbal battery of the Cognitive Abilities Test (Lohman et al., 2001) yielding an acceptable level of validity for research purposes (r=.72).

Table B.1

BPVS3 Reliability Based on Standard Error of Measurement (168 Possible Items) 95% Confidence Interval

Mean Age	*Raw Score Band*	*Standard Error Of Measurement*
5.2	73-96	7-13
7.5	97-120	7-13

Note. Information taken from the BPVS3 test manual (Dunn et al., 2009, p. 32).

B.3.2 CELF4 (Clinical Evaluation of Language Fundamentals UK, 4th Edition, Expressive vocabulary Subtest)

The *CELF4* (Semel et al., 2003) Expressive vocabulary subtest is a referential naming test standardised in 2002 in the US with over 4,500 individuals aged 5-21. Additional UK norming took place in 2004 with 1400 pupils between ages 3-16.

The test manual reports good reliability for the subscale in terms of Cronbach's alpha for the age group of interest to the current research project (see Table B.2). Convergent validity of the scale was not possible to ascertain, as there is no equivalent clinical diagnostic tool.

Table B.2

CELF4 Expressive Vocabulary - Cronbach's Alpha Reliability

Age	*Cronbach's α*
5;0-5;6	.83
5;7-5;12	.84
6;0-6;6	.84
6;7-6;12	.85

Note. Based on US norms. Information taken from the CELF4 test manual (Semel et al., 2003, p. 219).

B.3.3 PhAB2 (Phonological Assessment Battery, 2nd Edition)

The PhAB2 (Gibbs & Bodman, 2014) is a set of tests normed on children aged 5-11 to index phonological awareness and other phonological processing skills. The battery was trialled in the UK with 550 children, and the final standardisation sample consisted of a further 773 children. The test manual lists participant numbers in the 5-6 years age range (*N*=101) and the 6-7 years age range (*N*=122).

The Cronbach's alpha coefficients provided in the test manual indicate good or very good reliability for all subtests used in the current study (see Table B.3), apart from the 'acceptable' alpha for Phonological Working Memory. Data is not provided in the manual for the individual age ranges, presumably due to the limited size of subgroups.

Table B.3

PhAB2 Subtests - Cronbach's Alpha Reliability

PhAB2 Subtest	*Cronbach's α*
Rhyme	.80
Alliteration	.86
Phoneme Segmentation	.90
Nonword Reading	.84
Naming Speed Digits	.95
Phonological Working Memory	.76

Note. Information taken from PhAB2 test manual (Gibbs & Bodman, 2014, p. 80).

Convergent validity in the test manual is calculated in relation to the YARC (York

Assessment of Reading for Comprehension) Single Word Reading Test (Snowling et al., 2009), which has an adequate level of validity for research purposes (*r*=.72). This single word reading test provides a useful comparison with the PhAB2 Nonword Reading subtest, however comparing the PhAB2 battery to another test of phonological processing such as the CTOPP2 with similar subscales would have provided even more useful convergent validity. CTOPP2 reports convergent validity above 0.7 for all subtests in comparison to other similar scales, in line with the original result linked to the YARC.

B.3.4 BAS3 (British Ability Scales, 3rd Edition) Matrices Subtest

The *BAS3* (Elliot & Smith, 2011) *Matrices* subtest was administered as an index of non-verbal cognition. The BAS3 was standardised on 1360 children across the UK, with 193 in the 5 year old norming sample and 189 in the 6 year old sample. Two versions of the assessment are required to cover the intended age range: the Early Years version for ages 3-5;11, and the School Age version for ages 6-18. If the child reaches criterion at the end of the early version, then testing progresses to the school version.

Reliability is similar (good) for the early years and school age tests (see Table B.4). No convergent validity information was provided in the technical manual to compare the BAS3 Matrices subtest to other similar tools.

Table B.4

BAS3 - Cronbach's Alpha Reliability

Age	*Cronbach's α*
5	.86
6	.83

B.3.5 Quality of Phonological Representations (QPR) Task

The *Quality of Phonological Representations* task (Claessen et al., 2009) is a norm-referenced mispronunciation detection task, standardised in Australia with 235 mainstream children aged 5-6. The QPR task is made up of correct pronunciations and pseudoword mispronunciations. A computerised presentation was developed for the current study to provide a standard format for administration and scoring. The original QPR scale (*N*=100) used in the pilot study contained 10 stimuli for each target word: four correct pronunciations and six mispronunciations (see Claessen et al., 2009, pp. 140-141). A reduced scale was used for the cross-sectional study (*N*=50) containing five stimuli for each target word - three correct and two incorrect pronunciations. The randomised 50 item scale is available in Appendix C.

The QPR task displayed good levels of reliability for the original 100 item test (α=.84) as well as the reduced 50 item test (α=.84). It is not possible to provide information on convergent validity, as no similar standardised tests were found for comparison.

B.3.6 Bespoke Nonword Spelling Task

A nonword spelling task was created based on phases 2-5 of the Letters and Sounds programme (DFES, 2007). A reliability analysis using the cross-sectional data yielded a good alpha of .85. Convergent validity is difficult to ascertain, since no other nonword spelling tests were found for comparison.

B.4 Additional Tests Used in the Experimental Intervention Study

Most of the above tests were also used in the experimental test battery. The two further tests are presented below.

B.4.1 CTOPP2 (Comprehensive Test of Phonological Processing, 2nd Edition)

The CTOPP2 (Wagner et al., 2013) is used for ages 4 through to adult to assess phonological processing skills linked to literacy. The battery was normed in the USA in 2007 on 1900 participants. The test was normed on large samples for the age ranges of the current study: age 4-5 (N=251) and age 6-7 (N=295).

The manual reports very good reliability for the elision task in the age range of the cross-sectional sample (see Table B.5). Good convergent validity (r=.85) was established against the US-normed Early Reading Assessment subtest (Hammill et al., 2012) for 4-6 year olds.

Table B.5

CTOPP2 Elision - Cronbach's Alpha Reliability

Age	Cronbach's α
4-5	.90
5-6	.92

Note. Information taken from the CTOPP2 test manual (Wagner et al., 2013, p. 50).

B.4.2 Taught Vocabulary Definitions

A bespoke test of taught vocabulary definitions was developed to measure increases in taught vocabulary. An overview of the test construction, including randomisation, equivalence of the word sets, test administration and scoring can be found in Chapter 12, section 12.5.1.

Reliability of the 21-item probe was good (α=.78), suggesting that the test has good internal consistency. It was not possible to quantify convergent validity, although it was modelled on existing standardised tests such as the WISC4, 2004) and WASI-II (Wechsler, 2011).

Appendix C

Quality of Representations Task

Record Sheet

Naming of QPR items

Instructions: *What are these are called? (Show pictures; provide names for any unknown items. Count shortened words rhino and TV as correct but provide correct word. If the pupil says alligator, count as correct but say 'crocodile'.*

		Score 0 or 1
A	ambulance	
1	helicopter	
2	telescope	
3	dominoes	
4	crocodile	
5	television	

		Score 0 or 1
6	hippopotamus	
7	binoculars	
8	microphone	
9	rhinocerus	
10	spaghetti	
	Total correct	

QPR Mispronunciation Detection Task

Instructions: *You will hear some words on the computer. If it sounds right, click the tick; if it sounds wrong, click the cross. We will practise first. (Tester to click mouse for practice items. Draw a line to show when pupil attention wanes.)*

Item	pronounce	√ or X
Practice A	**ambulance**	√ X
Practice B	armbulance	√ X
Practice C	**ambulance**	√ X
Practice D	ampulance	√ X
Practice E	umbulance	√ X
1	hibopotamus	√ X
2	**crocodile**	√ X
3	**hippopotamus**	√ X
4	binorculars	√ X
5	domanoes	√ X
6	**dominoes**	√ X
7	spagiti	√ X
8	hilicopter	√ X
9	rhinoterus	√ X
10	**dominoes**	√ X
11	helicupter	√ X
12	televisiom	√ X
13	**rhinoceros**	√ X
14	teleslope	√ X

Item	pronounce	√ or X
15	**telescope**	√ X
16	**binoculars**	√ X
17	bimoculars	√ X
18	**television**	√ X
19	rhinorcerus	√ X
20	telescobe	√ X
21	microphome	√ X
22	migrophone	√ X
23	**microphone**	√ X
24	rhimocerus	√ X
25	**crocodile**	√ X
26	binukulars	√ X
27	duminoes	√ X
28	**hippopotamus**	√ X
29	tilavision	√ X
30	**telescope**	√ X
31	crookadile	√ X
32	**spaghetti**	√ X
33	**spaghetti**	√ X

Item	pronounce	√ or X
34	micwophone	√ X
35	spaketti	√ X
36	cocodile	√ X
37	tewescope	√ X
38	**helicopter**	√ X
39	crocodill	√ X
40	**helicopter**	√ X
41	**microphone**	√ X
42	hepapotamus	√ X
43	televusion	√ X
44	**rhinoceros**	√ X
45	domimoes	√ X
46	hipopitamus	√ X
47	**binoculars**	√ X
48	helicobter	√ X
49	pasgetti	√ X
50	**television**	√ X
Total Correct		/50
Total Correct Rejections		/30

Item Reduction of the QPR Scale

	C or V	Mean	SD	Consonant Change Place (P), Manner (M), Voice (V), Cluster Reduction (CR)	Vowel Change High-Low (HL) Front-back (FB) Long - short (LS)	Syllable Where Change Occurs
helicobter	C	.60	.503	V		3
hilicopter	V	.55	.510		FB	1
helicupter	V	.45	.510		HL	3
televusion	V	.90	.308		FB	3
televisiom	C	.65	.489	P (nasal)		4
tilavision	V	.50	.513		HL	1
domanoes	V	.80	.410		HL	2
domimoes	C	.35	.489	P (nasal)		3
duminoes	V	.20	.410		HL	1
crocodill	V	.90	.308		LS	3
crookadile	V	.70	.470		HL	1
cocodile	C	.65	.489	CR		1
teleslope	C	.90	.308	V		3
telescobe	C	.75	.444	V		3
tewescope	C	.50	.513	P		1
hipopitamus	V	.90	.308		FB	3
hepapotamus	V	.70	.470		HL	1
hibopotamus	C	.40	.503	V		1
binorculars	V	.95	.224		HL	2
bimoculars	C	.65	.489	P		2
binukulars	V	.40	.503		HL	2
microphome	C	.75	.444	P		3
migrophone	C	.70	.470	V		1
micwophone	C	.60	.503	P		2
rhinoterus	C	.80	.410	M		3
rhinorcerus	V	.70	.470		HL	3
rhimocerus	C	.55	.510	P (nasal)		2
spaketti	C	.75	.444	V		2
pasgetti	C	.70	.470	CR		1
spagiti	V	.45	.510		HL	2
Totals	16C 14V			6 Voice 1 Manner 7 Place (incl. 2 nasal) 2 Cluster reduc	3 FB 10HL 1LS	1st syllable – 10 2nd syllable - 8 3rd syllable - 11 4th syllable - 1

Appendix D

Nonword Spelling Task Marking Grid

	Word	Phonemes Correct	Acceptable Alternatives	Spellings Not Accepted
1	cag	3	kag	
2	reb	3		red
3	pim	3		
4	nud	3		
5	lof/f	3		
6	hish	3		
7	chee	2		
8	zight	3	zite	
9	woat	3	wote	wowt (ow not a medial digraph)
10	yain	3	yane	yayn (ay not a medial digraph)
11	vorb	3	vaub/vawb	
12	grib	4		
13	jant	4		
14	pronk	5	pronc	
15	spunch	5		
Total spellings correct	**/15**			
Total phonemes correct		**/ 50**		

Developmental considerations in marking:

Accepted:

- Capitals
- Script letters, e.g. ‘f’ (n.B. script letters may be harder to discern)
- Backward letters that don’t look like another letter: a, r, h, f, s, c, y, z
- Backward ‘g’ and ‘i ‘accepted if written as a descender to distinguish from ‘e’ and ’j’

Not accepted:

- Backward letters that resemble another letter: backward ‘g’ looks like ‘e’, backward ‘p’ looks like ‘q’, b and d reversals, backward capital ‘J’ looks like ‘t’
- Reversed patterns, e.g. ‘hc’ instead of ‘ch’
- Letter cannot be discerned, e.g. unclosed ‘g’ looks like ‘y’

In terms of phonemes correct, not penalised for extra letters (e.g. loft for lof) or wrong order of phonemes (e.g. girb for grib)

Appendix E

Research Information for Schools– Cross-Sectional Study

Title of Research Project: Assessment trial to investigate children's vocabulary and early reading skills

Background

Rose Brooks is the Advisory Teacher from the Communication and Interaction team who supports pupils with speech, language and communication needs. She is carrying out research for a PhD in the Human Communication Sciences Department at the University of Sheffield. The focus of the research will be to evaluate vocabulary teaching methods. Vocabulary has been shown to have a high impact on school attainment, behaviour, mental well-being and future employment prospects. The current phase of research will trial the assessments to be used in the future study. Children from Devon schools are now being recruited for the trial, depending upon parent/carer consent.

Please take time to read the following information in order to understand why the trial is being done and what the project involves. Staff and parents should feel free to contact Rose for further information or if you have a question.

What is the project´s purpose?

An assessment tool has been developed to measure a range of language and literacy skills, including vocabulary, pre-phonic skills (rhyme, sound awareness), memory, nonverbal ability and phonics. The trial will enable Rose to ensure that the tests will accurately measure children's progress in the later vocabulary teaching project.

Why has my child been chosen?

Up to 150 year one children are being asked to participate in this trial study in order to ensure that the assessment is relevant to a wide range of pupils. Year one children have been chosen because they are already learning phonics, which is linked to the skills being assessed.

Do I have to take part?

Involvement in the project is completely voluntary. It will first be cleared with school staff - the head teacher, SENCo and teacher. Next a *Parent/carer Consent Form* will be sent out to all pupils in the class. If parental permission is granted, the child will also be asked if they wish to take part. Participation can be withdrawn at any time without question. It will not affect the support the child or school receives in any way.

What does the study involve?

The child will have two short assessment sessions over the course of a week to measure oral vocabulary, pre-phonic skills and literacy. The testers (Rose and other employees from Babcock LDP) will spend time in the classroom before the sessions begin so they are familiar to the children. Each session will last approximately 25 minutes, and these will be made as interesting and fun as possible. Parental permission will be sought for audio recording some assessments in order to transcribe the child's speech, although these will not be used in presentations. The Advisory Teacher and other testers have a full DBS check (Disclosure and Barring Service). Assessment will take place in a quiet area of the school with an open-door policy.

What are the possible benefits?

The information will be used to develop an assessment to help the PhD research to identify effective vocabulary teaching strategies. While there is no direct benefit to your child, it may help to improve vocabulary teaching methods. A set of language and literacy support activities will be provided. Some of the research data may be published, in order to further the education knowledge base. This would be anonymised, so that no child can be identified.

Handling complaints should they arise

Please feel free to contact Rose Brooks via email (see below) with any concerns or queries, or phone her via the secretary if that is easier. In the event of a complaint, you may contact Dr. Traci Walker, Ethics Lead (traci.walker@sheffield.ac.uk) and then Professor Patti Cowell, Head of Human Communication Sciences Department (p.e.cowell@sheffield.ac.uk) tel: (0114) 2222418.

Confidentiality / Ethics

All details about pupils, families and staff will be kept strictly confidential. Only information gathered in this assessment project will be used. You will not be able to identify any child in the final report. This confidential information includes the child´s name, date of birth, home post code, contact details, school and test results. *The Parent/carer Consent Form* asks for permission to use the anonymous data in future publications. The University of Sheffield, Department of Human Communication Sciences has granted ethics approval for the research project.

Sponsor

Sheffield University Department for Human Communication Sciences has awarded a departmental scholarship. Babcock LDP has provided support to carry out the study.

Thank you very much for your time and consideration.

Contact details:
Secretary **-** Sue Vanstone on (01392) 287355;
Email: r.brooks@sheffield.ac.uk 19/1/17

The University Of Sheffield.

Appendix F

Parent/Carer Consent Form – Cross-Sectional Study

Title of Project: Trial of an assessment tool for a vocabulary teaching study
Name of Researcher: Rose Brooks

______________________	_______________	___________________	_____________
Child's Name	**Date of birth**	**School**	**Class**

Please initial each box:

☐ 1. I confirm that I have read and understand the Project Information Sheet dated 19.1.17 explaining the above assessment trial, and I have had the opportunity to ask questions about the project.

☐ 2. I understand that my child´s participation is voluntary and that I am free to withdraw him or her at any time without giving any reason and without any negative consequences. Please feel free to contact Rose Brooks through her secretary on (01392) 287355.

☐ 3. I give permission for some assessments to be audio-recorded, which will be anonymous and stored securely; these will not be used in presentations.

☐ 4. I understand that my child´s details will be kept strictly confidential, and his or her name will not be linked with the research materials. S/he will not be identified or identifiable in the research report. Only information collected from the study will be used.

☐ 5. I agree for the data to be used anonymously in the PhD study to follow, in future research publications, conferences, published resources and research projects.

☐ 6. I agree for my child to take part in the above assessment trial.

_________________________	_________________	_____________________
Parent / Carer Signature	**Post Code**	**Date**

The parent/carer received a copy Project Information Sheet. A copy of the signed and dated consent form will be placed in the project's main secure location.

Appendix G

Details of Reviewed Phonological-Semantic Intervention Studies

Paper	**Study Design**	**Sample** *Size* *Age* *diagnosis*	**Random-isation of Participants**	**Model of Delivery** *Who delivered* *Grouping*	**Dosage**	**Fidelity Measures**	**Control Measures**	**Statistical Analyses**
Specialist language settings (vocabulary outcomes)								
Best et al. (2021)	Quasi-experimental	*N*=20 6-8 DLD	Yes	SALT Individual	4.5 hours	Check on 10% of intervention sessions	Matching of vocabulary sets Blind assessors	ANOVA
Bragard et al. (2012)	Case series	*N*=4 9;6-13;9 SLI	No	SALT Individual	2.5 hours	No	No	McNemar test
Easton et al. (1997)	Case series	*N*=4 10-11	No	SALT Individual	5 hours	No	No	Descriptive % gain
German et al. (2012)	Quasi-experimental	*N*=10 7;11-8;9 SLI	No	SALT Small group	2 hours in each condition	No	--	Wilcoxon signed rank test
Gray (2005)	Quasi-experimental	*N*=24 4;0-5;11 SLI	No	SALT Individual	7.5 hours	No	Control group matched for age and gender IRR checks	ANOVA
Hyde-Wright et al. (1993)	Quasi-experimental	*N*=15 8;1-14;6 SLI	No	SALT Individual	5 hours – Phonological 7.5 hours - Semantic	No	No	Mann Whitney U test
Steele et al. (2013)	Quasi-experimental	*N*=15 9-11	No	SALT Individual	4 sessions Dosage not specified	Researcher observed 25% of sessions	Control group matched for language age IRR checks	Mann Whitey U test
Wing (1990)	Quasi-experimental	*N*=10 5;11-7.1 SLI	No	SALT Small group	12.5 hours	No	Control group matched for age	t tests

Wisenburn & Mahoney (2009)	Meta-analysis	44 studies Sample *N*>1	--	SALT Different group sizes	--	--	--	Effect size d
Zens et al. (2009)	Quasi-experimental	*N*=19 6;2-8;3 SLI	Yes	SALT Individual	12 hours	Sessions videotaped 20% viewed	Control group matched for age and gender IRR checks	ANOVA
Mainstream setting: Targeted invtervention (vocabulary outcomes)								
Clegg (2014)	Case series	*N*=5 6-8 Behaviour	No	SALT Individual	3 hours	No	Control word list	Sign test
Joffe et al. (2019)	RCT	*N*=358 11-12 Language difficulties	Yes	TA Small groups	4.5-6 hours	Researcher observed each TA 3 times Manual	Control group matched for age Blind assessment	Structural Equation Modelling (SEM)
Lowe et al. (2019)	Quasi-experimental	*N*=78 11-14 Language disorder	No	Teacher Whole class	2 hours	Teacher records Researcher observed 20 lessons	Control word list Resources supplied	ANOVA
Murphy et al. (2017)	RCT	*N*=128 11-13 Low SES schools	Random allocation of schools	Teacher Whole class	16 hours	Fidelity checklist completed by teachers Manual	Control group matched for age Blind assessment	ANCOVA
Parsons et al. (2005)	Case series	*N*=2 8;10 & 9;5 SLI	No	SALT Individual	10-12 hours	No	Control word list	Chi square
Spencer et al. (2017)	Quasi-experimental	*N*=35 12-14 Low ability classes	Yes	SALT Small groups of 3-5	10 hours	Lesson plans provided Weekly meeting	Control group matched for age Control word list Blind assessment	ANOVA
St John & Vance (2014)	Quasi-experimental	*N*=18 5-6 SLCN, EAL, SEN	No	Teacher Groups of 6	4 hours	No	Control word list	t tests

Mainstream setting: Whole class approach (vocabulary outcomes)								
Damhuis et al. (2016)	Quasi-experimental	*N*=87 4-6	No	Teacher Whole class	24 hours	Resources provided Researcher visits Weekly teacher reports	Control group matched for age	ANCOVA
Droop et al. (2005)	Quasi-experimental	*N*=263 4-6 All pupils	Information not available in English	Teacher Whole class	80 hours	Information not available in English	Control group matched or age	ANCOVA
Janssen et al. (2018)	Quasi-experimental	*N*=85 4-6 All pupils	Random allocation of schools	Teacher Whole class	10 hours	Resources provided Researcher visits Teacher records	No	ANCOVA
Moran & Moir (2018)	Pretest post-test	*N*=91 3-5 Low SES schools	No	Teacher Whole class	4 weeks (dosage not specified)	Researcher visits and coaching	No	t tests
Silverman (2007)	Quasi-experimental	*N*=94 5-6 Low SES, high EAL schools	Random allocation of classes within each school	Teacher Whole class	18 sessions (dosage not specified)	Two sessions per teacher videotaped and transcribed	No	Linear growth modelling
Mainstream setting (vocabulary and phonemic awareness outcomes)								
Droop et al. (2005)	As above							
Janssen et al. (2018)	As above							
Munro et al. (2008)	Pretest post-test	*N*=17 4-6 SLI	No	SALT Individual	6 hours	Scripted programme Six sessions videotaped	No	t tests

Appendix H

School Invitation Letter

Research project: The Impact of Vocabulary Teaching in 5-6 Year Old Children

For the attention of the Head Teacher

I would like to invite you to consider whether to take part in a vocabulary research project, which begins in September 2018. The *Participant Information Sheet for Schools* accompanying this letter provides further information about the research programme.

Background to the Research

Rose Brooks is the Advisory Teacher from the Communication and Interaction team who supports pupils with speech, language and communication needs. She is carrying out research for a PhD in the Human Communication Sciences Department at the University of Sheffield. The focus of the research will be to evaluate effective vocabulary teaching. Children from south Devon schools are now being invited to take part in the study, pending parent/carer consent.

Vocabulary is an important predictor of children's academic attainment and social outcomes. Current teaching of vocabulary needs to be improved in order to raise outcomes for all pupils. The link between vocabulary and literacy is especially important, and Year One (Y1) has been chosen as an important year for this development. The project will compare several methods of vocabulary instruction to see which provides the best outcomes for pupils. The findings could help schools to make more informed choices about their vocabulary teaching.

What does the study involve?

Each Y1 pupil who receives written parent/carer consent will have two short assessment sessions for language and literacy skills over the course of a week in September 2018 (week beginning 10th, 17th and 24th). Each session will last up to 25 minutes, and these will be made as interesting and fun as possible. All information will be anonymous and will not be linked to the child's name in any research outputs. The testers will be experienced in working with children and will have a full DBS check. Assessment will take place in a quiet area of the school with an open-door policy. Post-testing will take place in July 2019, after the teaching programme, and a further check will be made in November 2019 to ascertain whether the skills were maintained.

Y1 pupils will take part in a vocabulary teaching programme from 1st October 2018 to 5th July 2019. Teachers will receive full training, resources and support to implement the vocabulary teaching programme, free of charge. After training, teachers will provide a daily vocabulary lesson lasting 10 minutes, linked to a literacy text. The books are drawn from the approved list of high-quality texts for Y1 by the Babcock Literacy Team and the Centre for Literacy in Primary Education (CLPE). It is requested that the set text forms the basis of the teaching block (2-2½ weeks each) to reduce variability and to allow fair comparison between the teaching methods.

Vocabulary sessions will follow principles of effective vocabulary teaching. The daily vocabulary teaching structure is as follows: *select* the target word, *teach* it using the relevant vocabulary cue card, *activate* the learning through a multisensory activity on the task menu provided and *review* the word. A word wall in the classroom will display the pictorial vocabulary cards and associated resources.

Recruitment to the project

Up to 250 Y1 children are being asked to participate in this study. Schools have been invited if they have a Y1 class of at least 25 pupils on their Pupil Admission Numbers (PAN). Mixed year groups are not suitable for the assessment and teaching programme, which has been designed for Y1 participants and class routines.

Schools will be accepted on a first-come basis as responses arrive. As it is a randomised control trial, schools will be randomly allocated to either an intervention group or a waiting control group. The control group is vital to determine the real impact of the vocabulary programme compared to usual teaching. The control group will receive the same testing, and they will get the more effective teaching programme in December 2019, when they are in Y2 after all testing has been completed. The two teaching groups will not be informed of the contrasting approach in order not to influence results. Since socio-economic factors are important to vocabulary acquisition, each of group will contain a mix of schools at the upper and lower end of the socio-economic index.

School involvement in the project is completely voluntary. Participation can be withdrawn at any time without question by contacting Rose directly. It will not affect the support the child or school receives. A *Parent/carer Consent Form* will be sent out to all pupils in the class to seek permission for testing and audio recording of some assessments, alongside a *Participant Information Sheet for Parents/carers*. Rose is happy to meet with parents and to provide an information session if required. Once parental permission is granted, pupils will be asked if they wish to take part.

Teacher training and support

A half-day training session will be provided for Y1 teachers in September. Each teacher will also receive an individual planning and practice session with Rose in September before the start of the teaching programme. During the first week of the vocabulary teaching (in October), Rose will provide an informal observation and constructive guidance. She will also provide a visit in January and April 2019 to make an observation using a standard checklist and to provide further support.

A questionnaire will be given to Y1 teachers before their training session in September and again after the teaching programme has ended in July. They will also be asked to give their views in a small focus group of teachers after the teaching programme. The aims are to glean information about their views, barriers to implementation and professional development over the course of the project. All responses will be anonymised in the final report. Participating teachers will be asked to sign a consent form to check whether they accept their information to be used in this way.

Thank you for considering whether your Y1 class may join the vocabulary project. Please take time to discuss this proposal within school, and particularly with the Y1 teacher/s. You can express your interest via email on r.brooks@sheffield.ac.uk or through the Babcock administrator - Sue Vanstone on (01392) 287355. If your school takes part in the project, a Head teacher consent form will also be sent out to you. Once again, thank you for your time and consideration.

Kind regards, Rose Brooks *28/5/18*

The University Of Sheffield.

Appendix I

Research Information for Schools – Vocabulary Intervention Study

Invitation to the project
Rose Brooks is the Advisory Teacher from the Communication and Interaction team who supports pupils with speech, language and communication needs. She is carrying out research to evaluate effective vocabulary teaching. Children from south Devon schools are now being invited to take part in the study, with parent/carer consent. Please take the time to read the following information in order to understand why the research is being done and what the project involves. Staff should feel free to ask Rose for further information or with any questions using the contact details overleaf.

What is the project´s purpose?
Vocabulary is important for children's academic and social development. Current teaching of vocabulary needs to be improved in order to raise outcomes for all pupils. The link between vocabulary and literacy is especially important. Assessment before and after the project will help to demonstrate how different vocabulary teaching methods can affect progress. Rose is completing the research as part of a PhD in the Human Communication Sciences Department at the University of Sheffield.

Why have the children been chosen?
Up to 250 Year One children are being asked to participate in this study. Year One has been chosen because this is an important period for both vocabulary and reading development.

Do children have to take part?
The testing is completely voluntary; it is up to the school and ultimately the parent/carer to decide whether or not the child should take part. It will first be cleared with school staff, including the Head teacher and the Year One class teacher. A *Parent/carer Consent Form* and Participant Information Sheet for Parents/carers will then be sent out to all pupils in the class. Permission will be sought for audio recording some of the assessments. The children will also be asked if they wish to take part. Parents/carers can withdraw a child at any time by contacting the teacher or Rose directly. No questions will be asked, and it will not affect the support the child or school receives.

What does the study involve?
The child will have two short assessment sessions for language and literacy skills over a week in September 2018, July 2019 and November 2019. Each session will last up to 25 minutes, and these will be made as interesting and fun as possible. All information will be anonymous and will not be linked to the child's name. The testers will be experienced in working with children and will have a full DBS (Disclosure and Barring Service) check. Assessment will take place in a quiet area of the school with an open door policy. Children will receive vocabulary teaching beginning in September 2018, or in December 2019, depending on whether they are in the intervention or waiting control group (this is necessary to compare results fairly). Year One teachers in the project will complete an open-ended questionnaire before and after the teaching programme and will be asked to join a focus group to gain their views. This data will be anonymous and will not be linked to the teacher's name in any report or research output.

What are the possible disadvantages and risks of taking part?
Some children are reluctant to work with an unfamiliar adult in school. To reduce this potential difficulty, testers will spend time in the classroom in order to get to know the children beforehand.

Testing sessions have time built in for a chat to put children at their ease. The tests themselves are designed to be fun and interactive. The child may take a break, and they can also decide at any time not to take part.

What are the possible benefits of taking part?
The information will be used to check progress after vocabulary teaching. The findings could help schools to make more informed choices about their vocabulary methods. Some of the research data may be published, in order to further the education knowledge base, but this would be anonymised, so that no child can be identified. A potential benefit to your child is the additional vocabulary instruction s/he will receive.

Will the child's information be kept confidential?
A pupil identification number will be used instead of the child's name on all research information. All information will be kept strictly confidential to Rose and her admin team. The child will not be identified in any written report or publication.

What is the legal basis for processing the child's personal data?
According to data protection legislation, we are required to inform you that the legal basis applied in order to process the child's personal data is that 'processing is necessary for the performance of a task carried out in the public interest' (Article 6(1)(e)).

What will happen to the data collected, and the results of the research project?
The anonymised information will be used by Rose and her admin team in order to evaluate the results of vocabulary teaching. The master list of confidential information such as pupil names, dates of birth, school and home postcode will be kept securely in a locked filing cabinet and in a password protected file on Rose's encrypted laptop. Research results are likely to be used in publication and training conferences, but without any reference to children's names. All data will be kept for three years after the PhD is completed, after which it will be destroyed.

Who is organising and funding the research?
Sheffield University Department for Human Communication Sciences has awarded Rose a departmental scholarship for the PhD research.

Who is the data controller?
The University of Sheffield will act as the Data Controller for this study and is responsible for looking after your information and using it properly. All data will be kept strictly confidential and held securely for the duration of the PhD project.

Who has ethically reviewed the project?
This project has been ethically approved via the University of Sheffield's Ethics Review Procedure, as administered by the Department of Human Communication Sciences.

Handling complaints should they arise
In the event of a complaint not being handled to your satisfaction, you may contact Rose's research supervisor – Jenny Thomson via email at j.m.thomson@sheffield.ac.uk or the Head of Department – Patricia Cowell at p.e. cowell@sheffield.ac.uk, who will then escalate the complaint through the appropriate channels. If the complaint relates to how the participants' personal data has been handled, information about how to raise a complaint can be found in the University's Privacy Notice: https://www.sheffield.ac.uk/govern/data-protection/privacy/general**.**

Contact for further information
Please feel free to contact Rose Brooks via email (r.brooks@sheffield.ac.uk) with any concerns or queries, or you can contact her through the Babcock administrator, Sue Vanstone on (01392) 287355, if that is easier.

Will the child be recorded, and how will the recorded media be used?
Audio recordings will be made for some pupil assessments, so they can be checked for accuracy. No other use will be made of the audio recordings, and no one outside the project will be able to access them. Audio files will be labelled with an anonymised pupil identification number and stored securely on a password-protected laptop. Audio files and all other data will be destroyed three years after the PhD project is completed.

Thank you very much for your time and consideration.

Once signed, each staff participant receives a copy of the signed and dated consent form and Project Information Sheet for Schools. A copy of the signed and dated consent form will be placed in the project's main secure location.

01/05/18

The University Of Sheffield.

Appendix J

Parent/Carer Consent Form

Research Project: The Impact of Vocabulary Teaching in 5-6 Year Old Children

Child's Name: **Date of Birth:** **School:**

Please initial the appropriate boxes	**Yes**	**No**
Taking part in the project		
I have read and understood the Participant Information Sheet for Parents/carers dated 30.6.18. (If you answer No to this question, please do not proceed with this consent form until you understand what participation in the project will mean).		
I have been given the opportunity to ask questions about the project.		
I agree to my child taking part in the assessments linked to the research. I understand that this includes two short assessment sessions in September/October 2018, July 2019 and November 2019. I understand that some tests will be audio-recorded in order to check for accuracy.		
I understand that my child is taking part on a voluntary basis and that I can withdraw from the assessments at any time; I do not have to give any reasons for why I no longer want to take part and there will be no consequences.		
How my information will be used during and after the project		
I understand that my child's personal details such as name, date of birth, postcode, special needs status, ethnicity, EAL status and assessment results will be collected, but they will not be revealed to people outside the project. All information about my child will be kept confidential and stored securely.		
I understand and agree that the assessment and research information can be used in conferences and publications linked to this research. My child will not be named. All data will be destroyed three years after the PhD is completed.		
So that the information you provide can be used legally by the researchers		
I agree to assign the copyright for any materials generated as part of this project to The University of Sheffield.		

Does your child have any hearing difficulties? yes/no (please circle)

Name of parent/carer: **Signature:** **Date:**
Name of researcher: Rose Brooks **Signature:** **Date:**

Contact details for further information or if you have any questions:
Please feel free to contact Rose Brooks through her secretary on (01392) 287355 or via email: r.brooks@sheffield.ac.uk. If the complaint Is not handled to your satisfaction, you may contact her PhD supervisor Jenny Thomson via email: j.m.thomson@sheffield.ac.uk**.**

Once signed, the parent/carer receives a copy of the signed and dated consent form and Project Information Sheet for Parents/carers. A copy of the signed and dated consent form will be placed in the project's main secure location.

The University Of Sheffield.

Appendix K

Research Information for Parents/Carers

Research project: The Impact of Vocabulary Teaching in 5-6 Year Old Children

Invitation to the research project
Rose Brooks is the Advisory Teacher from the Communication and Interaction team who supports pupils with speech, language and communication needs. She is carrying out research to evaluate effective vocabulary teaching for all children. Children from south Devon schools are now being invited to take part in the study, with parent/carer consent. Please take the time to read the following information in order to understand why the research is being done and what the project involves. Please feel free to ask for further information or ask any questions using the contact details overleaf.

What is the project´s purpose?
Vocabulary is important for children's development and teaching needs to be improved in this area. The project will compare different methods of vocabulary teaching over the course of Year One. Rose is completing the research as part of a PhD in the Human Communication Sciences Department at the University of Sheffield.

Why have the children been chosen?
Up to 250 Year One children are being asked to join the study. Year One has been chosen because this is an important time for vocabulary and reading development.

Do children have to take part?
The testing is completely voluntary; it is up to the parent/carer to decide whether or not the child should take part. The children will also be asked if they wish to take part. Parents/carers can withdraw a child at any time by contacting the teacher or Rose directly. No questions will be asked, and it will not affect the support the child or school receives.

What does the study involve?
The child will have two short sessions of language and literacy assessment in September 2018, July 2019 and November 2019. The two sessions will last up to 25 minutes each and will include tasks that are as interesting and fun as possible. All information will be anonymous and will not be linked to the child's name. The testers will be experienced in working with children and will have a full DBS (Disclosure and Barring Service) check. Assessment will take place in a quiet area of the school with an open door policy. Children will have vocabulary teaching beginning in September 2018, or in December 2019, depending on whether they are in the intervention or waiting control group (this is necessary to compare results fairly).

What are the possible disadvantages and risks of taking part?
Some children are reluctant to work with an adult they don't know. To help reduce this potential difficulty, testers will spend time in the classroom in order to get to know the children beforehand. Testing sessions have time for a chat to make children feel comfortable. The tests themselves are designed to be fun and interactive. The child may take a break, and they can also decide at any time not to take part.

What are the possible benefits of taking part?
The information will be used to compare vocabulary teaching methods. The findings could help schools to make better choices about their vocabulary methods. Some of the research data may be published to support schools in their vocabulary teaching. Another benefit is the extra vocabulary lessons your child will receive.

Will my child's information be kept confidential?
A pupil identification number will be used instead of the child's name on all research documents. All information will be kept strictly confidential to Rose and her admin team. The child will not be identified in any written report or publication.

What is the legal basis for processing my child's personal data?
According to data protection legislation, we are required to inform you that the legal basis applied in order to process the child's personal data is that 'processing is necessary for the performance of a task carried out in the public interest' (Article 6(1)(e)).

What will happen to the data collected, and the results of the research project?
The anonymised information will be used by Rose and her admin team in order to compare the results of vocabulary teaching. The master list of confidential information such as pupil names, dates of birth, school and home postcode will be kept securely in a locked filing cabinet and in a password protected file on Rose's encrypted laptop. Research results are likely to be used in publication and conferences but will not be linked to children's names. All data will be kept for three years after the PhD is completed, then it will be destroyed.

Who is organising and funding the research?
Sheffield University Department for Human Communication Sciences has awarded a departmental scholarship for the PhD research.

Who is the data controller?
The University of Sheffield will act as the Data Controller for this study and is responsible for looking after your information and using it properly. All data will be kept strictly confidential and held securely for the duration of the PhD project.

Who has ethically reviewed the project?
This project has been ethically approved via the University of Sheffield's Ethics Review Procedure, as managed by the Department of Human Communication Sciences.

Handling complaints should they arise
In the event of a complaint not being handled to your satisfaction, you may contact Rose's research supervisor – Jenny Thomson via email at j.m.thomson@sheffield.ac.uk or the Head of Department – Patricia Cowell at p.e. cowell@sheffield.ac.uk, who will then escalate the complaint through the appropriate channels. If the complaint relates to how the participants' personal data has been handled, information about how to raise a complaint can be found in the University's Privacy Notice: https://www.sheffield.ac.uk/govern/data-protection/privacy/general.

Contact for further information
Please feel free to contact Rose Brooks via email (r.brooks@sheffield.ac.uk) with any questions or concerns, or you can contact her through the Babcock administrator, Sue Vanstone on (01392) 287355, if that is easier.

Will my child be recorded, and how will the recorded media be used?
Audio recordings will be made for some assessments, so they can be checked for accuracy. The audio recordings will not be used for anything else, and no one outside the project will have access to them. Audio files will be labelled with the anonymised pupil identification number and stored securely on a password-protected laptop. Audio files will also be destroyed three years after the PhD project is completed.

Thank you very much for your time and consideration.

If you are happy for your child to take part in this project, please can you initial each section of the Parent/carer Consent Form and return it to school.

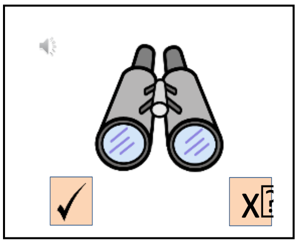
Appendix L

Supplemental Research Information for Parents/Carers

Research Project: The Impact of Vocabulary Teaching in 5-6 Year Old Children

We would like to invite your child to take part in some enjoyable assessment tasks linked to a vocabulary teaching project.

The assessments will help schools to improve vocabulary teaching, which is important for children's school and life outcomes. This is what some of the tests look like:



There will be two short sessions of around 20-25 minutes each before and after the project. Trained adults will come in to do the activities, under the direction of Rose Brooks, our Advisory Teacher from the Communication and Interaction Team. It does not matter how many answers your child gets right or wrong. The purpose is to see if the vocabulary teaching is effective.

The audio recording and results will be used in the project and later writing, but not with your child's name.

You have a chance to ask questions about the project if you wish – feel free to phone Rose Brooks on 07714363863. You can see what the teaching and assessments look like on Youtube:
https://www.youtube.com/watch?v=Q8aOmUpj4oE&feature=em-share_video_user

You have a choice whether you would like your child to be involved. It does not affect the support your child or school will get.

If you agree for your child to take part in the language assessment tasks, please can you sign the Parent/carer consent form and send it back to the teacher as soon as possible.

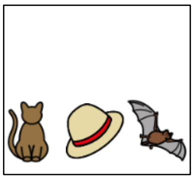
6/7/18

Appendix M

School Newsletter Blog

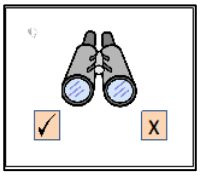
Sample Information for School Use

In September, Year One will be taking part in a project to develop children's vocabulary skills. Rose Brooks, our Advisory Teacher from the Communication and Interaction team, will be supporting Y1 with this project, which involves daily vocabulary teaching and practice games. Please can you read the project information sheet that will be sent home. Then please sign and return the Parent/carer consent form if you are happy for your child to take part in the language and literacy assessments related to the project. These are enjoyable tasks that are needed to see how well the vocabulary teaching is working for all children. Here are some pictures of the tasks the children will be doing:



Rhyming words

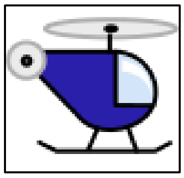
Timing how quickly you can say words



A computer game to listen to words carefully

ras	nam
san	fap
paz	vam
pam	pam
lat	vam
mab	san

Reading and spelling nonsense words



Pointing to pictures and naming them

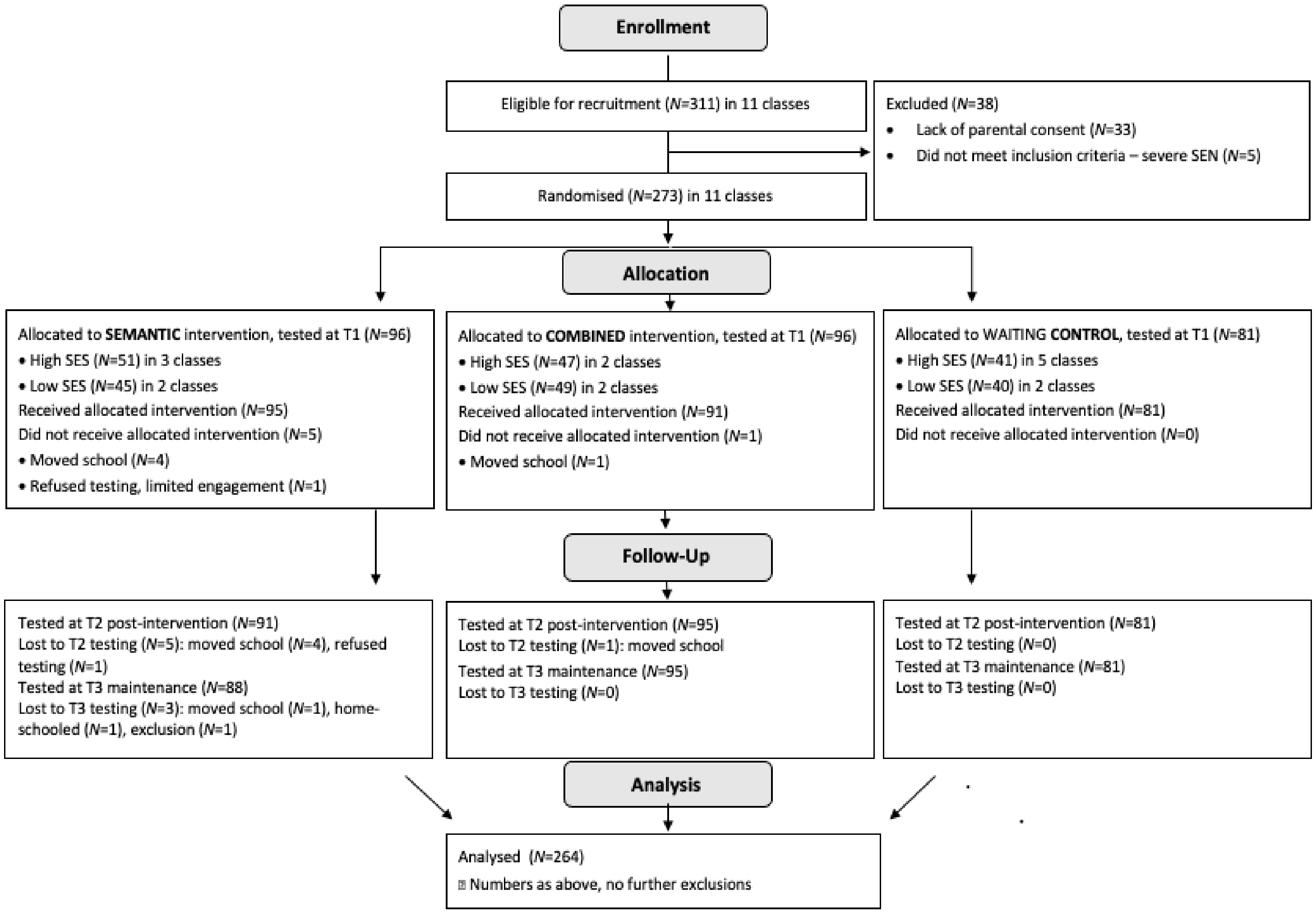
(Images from Widgit Symbols © Widgit Software 2002 – 2018)

You can hear more about the project and assessment tasks on Youtube:

https://www.youtube.com/watch?v=Q8aOmUpj4oE&feature=em-share_video_user

Appendix N

Flow of Participants Through the Intervention Study



Appendix O

Taught Vocabulary Definitions

Test Record

Word	Response	Score		
		2	**1**	**0**
1. searched				
2. realised				
3. spy				
4. muscles				
5. distracting				
6. wobble				
7. dunes				
8. noticed				
9. ancient				
10. choir				
11. squawk				
12. grab				
13. lair				
14. tumbling				
15. stroll				
16. personality				
17. disguise				
18. dive				
19. mysterious				
20. tool				
21. anchor				
	Column totals			
	Total			

Taught Vocabulary Definitions Scoring Protocol and Matrix

Scoring Protocol

Examples taken from the WISC4 administration manual (Wechsler, 2004)

- Award two points for: (1) a definition incorporating all points on the scoring matrix, (2) a good synonym (*cap* means a hat), (3) the function (an *umbrella* keeps off the rain), (4) a description (a *bike* has four wheels and a handlebar) or (5) another verbal response that demonstrates accurate understanding. The response must make sense to be awarded two points, even if both elements of the two-point matrix are present.

- Give one point if: (1) one of the elements from the two-point scoring matrix is present, (2) the answer corresponds to the one-point column of the scoring matrix, (3) a sentence which clearly shows the meaning, (4) a concrete example using the target word but not elaborated (e.g. My *clock* tells me when to go to school) or (5) if the meaning is evident but not clearly expressed (e.g. muscles = 'you get strong in your body').

- Score zero points for (1) a wrong answer, (2) if the child responds "I don't know", or "I never heard it", (3) no response after 10 seconds wait time, (4) a demonstration rather than verbal explanation, even after prompting or (5) a sentence using the target word that does not demonstrate meaning, for example **spy:** *I spy on someone*. Such constructions may be learned as repeated or formulaic responses and do not necessarily demonstrate comprehension.

Grammatical correctness is not a factor in scoring the definitions test.

The following matrix was used for more detailed scoring.

Taught Definitions Scoring Matrix

Simple definition	2 points	1 point (or one element of the 2 point column)	0 points
1. **searched** – looked carefully for something or someone **OR** looked for information on the internet	*Two points:* Look up information on the internet Look high and low *One point for each element:* (1) look, seek, explore, try to find (2) for something / someone/ stuff/ it	Find Go through your stuff Try to get something Search on a computer	Search for something See Spot
2. **realised** – suddenly understood something	*One point for each element:* (1) recognise, understand, discover, find out, figure out, think, notice, know, remember (2) suddenly, finally, just, now	Forget	See/look Surprised Spot Find
3. **spy** – to notice or spot something **OR** to watch someone secretly **OR** a person who tries to find out secret information	*Two points:* Creep up on someone A person who tries to find out secret information *One point for each element:* (1) watch, peep, peek, sneak up, creep up, tiptoe, look at, see (2) secret, without someone knowing, without being seen, surprise, hide, sneakily, quietly, disguised	Keep a look out/eye on Spot something Agent	Spy on someone/ something/ people/ you Search Look for something Find something Costume Trick Stare Look around Spy with your eye
4. **muscles** – part of the body that makes our bones move and makes us strong Alternative meaning: shellfish (mussels)	*One point for each element:* (1) part of/inside the body or arms (2) makes us strong, makes us move, pick up heavy things	Strength/strong Power On your bones Six pack Keep you fit Big arms	Something you eat Exercise/go to the gym Big lumps
5. **distracting** – to take a person's attention away from what they are doing	*One point for each element:* (1) take attention away, disturb, put them off (2) from a task/work, what they are doing, when busy	Stop/distract someone concentrating /working Get in your way Not letting you work Make them not learn Make silly noises	Talk Interrupt Mess around Annoy Bother

6. **wobble** - to move or tip from side to side in an unsteady way Alternative meaning: to have a wobble	*One point for each element:* (1) move, tip, go, fall, shake, tilt, go, wiggle, jiggle, wriggle, teeter, quiver, topple (2) from side to side, all over the place, sideways, back and forth, around, left and right, about	Fall (over) Seesaw Like a jelly Like a penguin About to fall off/over Wobbly tooth Out of balance Unsteady/unstable Nearly fall down Wibbly wobbly Feel unsure	Not still Balancing Dizzy Trip Like drunk Scared
7. **dunes** – mound or hill of sand created by the wind	*Two points:* sand + hill/lumps/mounds/piles/ pyramids/beach hill/mountains/bumps	Sand Near the beach Sand dunes In the desert	cliffs
8. **noticed** – to see or observe something not there before	*One point for each element*: (1) recognise, see, realise, figure out, spot, look at, observe (2) something not there before, different, surprise, suddenly, just appeared, quickly, then, looks different		Note Know Remember Forget Find / find out Not right Look Watch Think
9. **ancient** – very old; from a long time ago	*Two points:* Old Long time ago/long ago Years ago/ ages ago In the past Olden days	A long time Fossilised	Ancient Egyptians Mummies Artefacts
10. **choir** – a group of people who sing together	*One point for each element:* (1) a group, club, crowd, people, they, everybody, together, all (2) who sing	At church Sing/ing	Band Music Perform(ance)
11. **squawk** – to give a loud high scream like a parrot	*One point for each element:* (1) bird + sound/noise/call (2) noisy, loud, big, shout, scream, screech, shriek, high-pitched, squeak	A bird/parrot goes/tells/ does/says/squawks/ tweets/ squeaks, etc. Not a nice sound	A sound/ noise A bird Funny Weird
12. **grab** - to pick something up quickly and roughly	*Two points:* To snatch *One point for each element:* (1) to take away/off, get, grasp, hold, pinch, pull, tug, catch, pick up, reach, grip (2) suddenly, quickly, roughly, tight	To have Squeeze Move	Touch Steal
13. **lair** – a wild animal's shelter or den **OR** a hideout for bad people	*One point for each element:* (1) for animals, villains/ baddies (2) home, habitat, house, shelter, live, base, cave, place, space, sleeps, den, burrow, underground	Secret hidden place Hideaway Where people hide	Zoo Trap

14. **tumbling** – to roll over head first **OR** when something is falling	*One point for each element:* (1) roll, tip over, flip, roly poly, fall (2) down, forward, head over heels, off, for a long time, steep, down a hill	Wobble Up and down Forward roll Fall over Fall again and again	Trip Like Jack and Jill Go Tumble clothes Move
15. **stroll** – to go for a slow walk	*Two points:* Roam, wander, walk around, walk along, take a walk, go for a walk, have a walk, sense of rambling *One point for each element:* (1) to walk (2) slowly, not for a purpose, for leisure, little	Go somewhere Strolling along	Stroll a baby Go for a picnic Run Go on a trip
16. **personality** – what you are like as a person; your character	*Two points:* Identity, self, character What you are like Who/what/how you are What you act like All about you What makes you Makes us unique	Different than another person Yourself Core values Describes someone All the things you have inside Likes Interests Talents/what you are good at When it's your thing	Own stuff Nice Feelings Traits What you do What you think What you look like
17. **disguise** – to change the way you look so you cannot be recognised **OR** a costume, mask or camouflage, so that you are not recognised	*One point for each element:* (1) costume, mask, dress up, outfit, camouflage, pretend to be someone else, like someone else (2) no one will recognise you/work you out, as a different person, not be recognised, as someone else, to spy on people	Wear different things Blend in What a spy wears	Hide Invisible Trick Change clothes
18. **dive –** to swim deep under-neath the water **OR** to jump down head first from a high place	*Two points:* Swim + deep/under/down/below the water *One point for each element:* (1) jump, go into water/sea/pool/liquid (2) downward, headfirst, arms first, hands first, bend down/over, up, hands together	Dive in a pool/ water/sea Jump off a diving board Dive for the ball Plunge Jump in/into/off Splash Swoop Hands like a rocket Go off a cliff Like a dolphin Lean down	Go in water Jump (without 'in') Straight Sink

19. **mysterious**– something that is strange and can't be explained	*One point for each element:* (1) something + odd/strange/ weird/ funny/suspicious/confusing/unusual /different *(2)* can't understand, explain/ recognise/ never seen before	Curious about	Scary/spooky Not right Something is wrong Something you don't know Unsure
20. **tool** - something people use to fix and repair items	*Two points:* (Used to) mend /fix/repair/build/ make/work with/create	Things in the garage Made of metal That you build with Examples of individual tools, e.g. spanner Break, cut, chop, bang, drill, etc.	You use a tool
21. **anchor** – a heavy metal object that is attached to a long rope. It is dropped from a boat or ship to stop it moving	*One point for each element:* (1) dropped, lowered, pulled up, pick up, lift, thrown from a boat, heavy metal thing/object, down, put in, bottom of sea, get out, dig in ground, under water, goes in sand, bring down (2) stops the boat from moving, keeps ship still, holds the boat, stops it going anywhere, makes it stay, slows it down, weighs down, not float away, keeps it in place	Made of metal Hangs from/off a boat Has a rope	Hook On/ part of a boat/ ship Heavy In the water Goes into the sea Save people

Lexical Characteristics of Taught Vocabulary

Important variables affecting word learning are specified for the taught words in each teaching block. AoA norms were taken from Kuperman et al. (2012), since it provided a large word list. The MRC Psycholinguistic Database (Wilson, 1988) was consulted for written frequency, word type, word length (number of syllables) and imageability ratings. The lemma (root word) was used in cases when the derived word was not available. In the MRC database, written frequency is taken from Kucera and Francis (1967), ranging from 0-69971 frequency of occurrence. Imageability data are derived from an amalgamation of three sets of ratings (Pavio et al., 1968; Toglia & Battig, 1978; Gilhooly & Logie, 1980), ranging in difficulty from 100-700.

Biphone probability was calculated through the Phonotactic Probability Calculator described in Vitevitch and Luce (2004). Words were first transcribed into International Phonetic Alphabet (IPA) and then into Klattese to insert in the calculator. Whilst no biphone probability norms are available, earlier acquired words, e.g. cat (biphone probability=0.0091), would generally contain more common patterns than later acquired words, e.g. glimpse (biphone probability=0.0118).

AoA, Frequency, word length, imageability norms taken from: https://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.html

Biphone probabilities taken from: http://www.people.ku.edu/~mvitevit/PhonoProbHome.html

Trial block: Naughty Bus

Vocabulary Item	*Lemma if Applicable*	*AoA*	*Written Frequency*	*Word Type*	*Number of Syllables*	*Imageability*	*Biphone Probability*	*Klattese*	*IPA Transcription*
traffic		6.22	68	noun	2	--	.0379	tr@fiK	tɹæfɪk
passenger		6.65	14	noun	3	529	.0295	p@sxnJx	pæsənʤə
tight		5.25	28	adjective	1	495	.0079	tYt	taɪt
reflection	reflect	5.50	32	noun	3	567	.0060	rxflEkS^n	rəflɛkʃʌn
handsome		6.50	40	adjective	2	--	.0313	h@ns^m	hænsʌm
lonely	lone	5.80	25	adjective	2	443	.0070	lxUnli	ləʊnli
building	build	6.16	160	noun	2	--	.0197	bildiG	bɪldɪŋ
hook		6.26	5	noun	1	541	.0039	hUk	hʊk
tuck		5.79	2	verb	1	416	.0063	t^k	tʌk

Block 1: Augustus and His Smile

Vocabulary Item	*Lemma if Applicable*	*AoA*	*Written Frequency*	*Word Type*	*Number of Syllables*	*Imageability*	*Biphone Probability*	*Klattese*	*IPA Transcription*
crept	creep	8.00	10	verb	1	--	.0211	krEpt	kɹɛpt
cluster		10.00	13	noun	2	--	.0251	kl^stx	klʌstə
searched	search	8.72	66	verb	1	402	.0029	sRCt	sɜʧt
scaled	scale	8.85	60	verb	1	463	.0164	skeIld	skeɪld
swirled	swirl	6.67	2	verb	1	--	.0069	swRld	swɜld
frost		5.21	6	noun	1	595	.0256	frast	fɹɒst
patterns	pattern	6.89	113	noun	2	453	.0287	p@txnz	pætənz
paraded	parade	6.53	25	verb	3	578	.0037	PxreId	pəɹeɪd
realised	realise	8.23	69	verb	3	--	.0070	rixlYzd	ɹiəlaizd

Block 2: How to Babysit Grandad

Vocabulary Item	*Lemma if Applicable*	*AoA*	*Written Frequency*	*Word Type*	*Number of Syllables*	*Imageability*	*Biphone Probability*	*Klattese*	*IPA Transcription*
illustrate		8.32	17	verb	3	--	.0175	IlxstreIt	ɪləstɹeɪt
wiggle		3.95	1	verb	2	491	.0162	wIgxl	wɪgəl
statue		7.55	17	noun	2	562	.0281	st@tCu	stæʧu
spy		8.00	9	noun	1	--	.0122	spY	spaɪ
remind		8.15	15	verb	2	--	.0123	rxmYnd	ɹəmaɪnd
pavement	pave	6.30	11	noun	2	--	.0423	peIvmxnt	peɪvmənt
entertain		7.50	14	verb	3	435	.0172	EntxteIn	ɛntəteɪn
somersault		5.63	2	noun	3	562	.0260	samxsalt	sʌməsɒlt
muscles	muscle	8.45	42	noun	2	553	.0151	m^sxlz	mʌsəlz

Block 3: How to Catch Santa

Vocabulary Item	*Lemma if Applicable*	*AoA*	*Written Frequency*	*Word Type*	*Number of Syllables*	*Imageability*	*Biphone Probability*	*Klattese*	*IPA Transcription*
fuel		7.17	17	noun	1	521	.0012	fjul	fjul
invent		8.16	7	verb	2	--	.0159	InvEnt	ɪnvɛnt
headlamp		6.79	--	noun	2	--	.0132	hEdl@mp	hɛdlæmp
distracting	distract	8.72	2	verb	3	--	.0610	dIstr@ktIG	dɪstɹæktɪŋ
luring	lure	9.00	7	verb	2	389	.0142	lyUrIG	ljʊɹɪŋ
glimpse		9.24	16	noun	1	422	.0118	glImps	glɪmps
scatter		7.95	2	verb	2	--	.0185	sk@tx	skætə
whinnying	whinny	7.35	1	verb	3	--	.0323	wIniyIG	wɪnijɪŋ
preparations	preparation	9.33	54	noun	4	313	.0195	prEpxreIS^nz	pɹɛpəɹeɪʃʌnz

Block 4: Don't Spill the Milk

Vocabulary Item	*Lemma if Applicable*	*AoA*	*Written Frequency*	*Word Type*	*Number of Syllables*	*Imageability*	*Biphone Probability*	*Klattese*	*IPA Transcription*
season		6.05	105	noun	2	495	.0203	sizxn	sizən
steady		6.70	41	adjective	2	378	.0304	stRdi	stɛdi
wobble		6.56	3	verb	2	--	.0133	wabxl	wɒbəl
dunes	dune	9.65	1	noun	1	--	.0053	junz	ʤunz
flock		7.18	10	noun	1	516	.0102	flak	flɒk
desert		8.35	21	noun	2	--	.0087	dEzxt	dɛzət
shiver		7.50	4	verb	2	578	.0070	SIvx	ʃɪvə
stalk		8.22	7	verb	1	440	.0249	stck	stɔk
wail		9.41	3	verb	1	452	.0040	weIl	weɪl

Block 5: How to Hide a Lion at School

Vocabulary Item	*Lemma if Applicable*	*AoA*	*Written Frequency*	*Word Type*	*Number of Syllables*	*Imageability*	*Biphone Probability*	*Klattese*	*IPA Transcription*
sneak		5.83	2	verb	1	--	.0084	snik	snik
noticed	notice	7.35	59	verb	2	467	.0017	nxUtIs	nəʊtɪs
museum		8.55	32	noun	3	--	.0101	myuziyxm	mjuzijəm
ancient		8.26	69	adjective	2	451	.0581	eIntCxnt	eɪnʧənt
spotted	spot	5.39	57	verb	2	507	.0205	spatId	spɒtɪd
fetched	fetch	6.61	6	verb	1	--	.0031	fECt	fɛʧt
hitched	hitch	9.39	5	verb	1	--	.0084	hICt	hɪʧt
screeched	screech	8.17	1	verb	1	--	.0267	skriCt	skɹiʧt
anxiously	anxious	10.21	29	adverb	3	376	.0109	@GkSxsli	æŋkʃəsli

Block 6: Could a Penguin Ride a Bike?

Vocabulary Item	*Lemma if Applicable*	*AoA*	*Written Frequency*	*Word Type*	*Number of Syllables*	*Imageability*	*Biphone Probability*	*Klattese*	*IPA Transcription*
imagine		6.06	61	verb	3	--	.0037	Im@JIn	ɪmæʤɪn
hatches	hatch	7.00	5	verb	2	--	.0110	h@CIz	hæʧɪz
champion		8.61	23	noun	2	508	.0397	C@mpixn	ʧæmpiən
waddle		7.79	--	verb	2	--	.0137	wadxl	wɒdəl
webbed	web	8.61	6	adjective	1	--	.0051	wEbd	wɛbd
choir		6.53	8	noun	2	567	.0061	kwYx	kwaiə
squawk		9.26	1	verb	1	--	.0158	skwck	skwɔk
huddle		9.40	4	verb	2	--	.0105	h^dxl	hʌdəl
beak		5.42	143 bill	noun	1	574	.0063	bik	bik

Block 7: The Day Louis got Eaten

Vocabulary item	*Lemma if Applicable*	*AoA*	*Written frequency*	*Word type*	*Number of syllables*	*Imageability*	*Biphone probability*	*Klattese*	*IPA transcription*
gulp		7.29	2	verb	1	--	.0044	g^lp	gʌlp
grab		5.37	16	verb	1	--	.0219	gr@b	gɹæb
snatch		8.18	4	verb	1	--	.0054	sn@C	snæʧ
guzzle		11.72	1	verb	2	--	.0089	g^zxl	gʌzəl
managed (to)	manage	7.67	20	verb	2	303	.0280	m@nIJ	mænɪʤ
track		6.94	38	verb	1	499	.0250	tr@k	tɹæk
lair		11.63	--	noun	1	454	.0033	leIx	leiə
wriggle		6.47	--	verb	2	--	.0165	rIgxl	ɹɪgəl
fled	flee	8.33	1	verb	1	431	.0119	flEd	flɛd

Block 8: Previously

Vocabulary Item	*Lemma if Applicable*	*AoA*	*Written Frequency*	*Word Type*	*Number of Syllables*	*Imageability*	*Biphone Probability*	*Klattese*	*IPA Transcription*
bothered	bother	6.50	22	adjective	2	369	.0083	baDxd	bɒðəd
previously	previous	8.44	86	adverb	4	276	.0257	privixsli	priviəsli
desperate		9.22	26	adjective	3	--	.0188	dEspxrxt	dɛspəɹət
tumbling	tumble	7.89	3	verb	2	461	.0233	t^mblIG	tʌmblɪŋ
stroll		7.83	4	verb	1	--	.0365	strxUl	stɹəʊl
ploughed	plough	7.11	--	verb	1	--	.0103	plaUd	plaʊd
sowed	sow	9.67	3	verb	1	479	.0009	sxUd	səʊd
plank		7.84	7	noun	1	598	.0221	pl@Gk	plæŋk
ballroom		7.37	--	noun	2	--	.0129	bclrum	bɔlɹum

Block 9: Wanted the Perfect Pet

Vocabulary Item	*Lemma if Applicable*	*AoA*	*Written Frequency*	*Word Type*	*Number of Syllables*	*Imageability*	*Biphone Probability*	*Klattese*	*IPA Transcription*
expenses	expense	8.00	50	noun	2	--	.0243	EkspEnsIz	ɛkspɛnsɪz
peace		6.32	198	noun	1	446	.0060	pis	pis
varieties	variety	7.89	85	noun	4	372	.0085	vxralxtiz	vəɹalətɪz
stern		9.47	23	adjective	1	424	.0254	stRn	stɜn
personality		8.68	48	noun	5	405	.0347	pRsxn@lxti	pɜsənæləti
common		6.94	223	adjective	2	276	.0479	kamxn	kɒmən
disguise		8.67	5	verb	2	--	.0213	dIsgaIz	dɪsgaɪz
plopped	plop	7.15	--	verb	1	--	.0099	plap	plɒp
admit		7.56	37	verb	2	--	.0018	@dmIt	ædmɪt
disappointed	disappoint	7.00	--	adjective	4	351	.0620	dIsxpcIntid	dɪsəpɔɪntəd

Block 10: Traction Man is Here

Vocabulary Item	*Lemma if Applicable*	*AoA*	*Written Frequency*	*Word Type*	*Number of Syllables*	*Imageability*	*Biphone Probability*	*Klattese*	*IPA Transcription*
helmet		5.71	1	noun	2	620	.0202	hElmIt	hɛlmɪt
guarding	guard	6.25	48	verb	2	530	.0166	gadIG	gadɪŋ
volunteered	volunteer	6.89	9	verb	3	--	.0366	val^ntixd	vɒlʌntiəd
dive		7.21	23	verb	1	586	.0054	daIv	daɪv
camouflage		7.53	3	verb	3	513	.0257	k@mxflaJ	kæməflaʒ
waist		6.42	11	noun	1	530	.0179	weIst	weɪst
mysterious	mystery	9.00	26	adjective	4	472	.0606	mIstixrixs	mɪstiəɹiəs
rescue		7.17	15	verb	2	456	.0208	rEskyu	ɹɛskju
cliff		7.28	11	noun	1	599	.0132	klIf	klɪf

Block 11: Mrs Armitage on Wheels

Vocabulary Item	*Lemma if Applicable*	*AoA*	*Written Frequency*	*Word Type*	*Number of Syllables*	*Imageability*	*Biphone Probability*	*Klattese*	*IPA Transcription*
greasy	grease	6.39	8	adjective	2	--	.0225	grisi	grisi
tool		5.37	40	noun	1	538	.0056	tul	tul
spanner		12.50	--	noun	2	--	.0190	sp@nx	spænə
screwdriver		6.00	--	noun	3	617	.0226	skrudraIvx	skɹudɹaɪvə
hammer		5.42	9	noun	2	618	.0183	h@mx	hæmə
panting	pant	5.94	--	verb	1	--	.0384	p@nt	pænt
faithful	faith	8.42	12	adjective	2	--	.0030	feITfxl	feɪθfəl
pedal	pedal	6.50	4	noun	2	556	.0149	pEdxl	pɛdəl
exhausted	exhaust	9.19	7	adjective	3	520	.0109	Ekzcxstid	ɛgzɔstəd
anchor		5.72	15	noun	2	561	.0039	@Gkx	æŋkə

Block 12: The Sand Horse

Vocabulary Item	*Lemma if Applicable*	*AoA*	*Written Frequency*	*Word Type*	*Number of Syllables*	*Imageability*	*Biphone Probability*	*Klattese*	*IPA Transcription*
plunge		8.61	5	verb	1	548	.0129	pl^nJ	plʌnʤ
gallop		8.32	4	verb	2	--	.0129	g@lxp	gæləp
mane		5.60	--	noun	1	528	.0076	meIn	meɪn
rippling	ripple	6.58	5	verb	2	--	.0238	rIpxlIG	ɹɪpəlɪŋ
admire		7.42	10	verb	2	390	.0013	@dmYx	ædmaɪə
faint		7.53	25	adjective	1	466	.0119	feInt	feɪnt
prance		8.95	--	verb	1	--	.0211	prants	pɹants
swoop		6.11	2	verb	1	--	.0057	swup	swup
drench		9.35	--	verb	1	--	.0231	drEnC	dɹɛnʧ

Appendix P

Teaching Manual for the Vocabulary Programme

Contents

The above resources will be used during explicit daily vocabulary teaching using the STAR approach.

Select:

- Vocabulary teaching blocks overview
- Teacher planning and recording sheets

Teach:

- Instructional cue card
- Pupil-friendly definitions
- Symbolised vocabulary cards (with thanks to WidgitOnline)

Activate/
Apply:

- Games in wallets (drawing upon Wordaware, Parsons & Branaghan 2014)
- Masters

Review:

- Word wall resources
- Vocabulary self-rating scale
- Word spy

1. Vocabulary Teaching Protocol

The following standard approach has been agreed with participating Year One teachers.

Setting Up the Word Wall

Please set up your intervention word wall on a display board using the resources and template if you wish. Please include all elements on the sheet entitled 'Setting up a word wall'. Children's work linked to the taught vocabulary should also be displayed here to highlight application of the words in written work. Ideally, no other resources would be placed on this display in order to highlight the taught words.

Vocabulary Self-Rating Scale

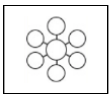
Prior to and after each storybook teaching block, children can complete a self-evaluation of their knowledge of the taught vocabulary. This is provided on a double-sided sheet, one side for the start and the other side for the end of the block. Please read each word aloud and ask the pupils to put a cross or a tick in the √ or **X** column to indicate whether or not they know what the word means. Alternatively, the teacher may wish to complete this activity as a whole class to identify which words they feel that they know/don't know. You will need to explain what you mean by 'know', e.g. can use it in a sentence, can explain what it means.

Daily Explicit Vocabulary Teaching

A 10-12 minute vocabulary teaching session takes place each day linked to the selected fiction text. The *Planning and Recording Sheet* uses the STAR approach – select, teach, activate, review. The approximate timings should be as follows: teaching the word for 4/5 minutes, pupil practice for 4/5 minutes in a game format, review previously learned words for 2 minutes. The steps in the STAR approach are outlined below:

Select

The tier two word for each day is shown on the *Vocabulary Teaching Blocks overview* and is also written on the *Teacher Planning and Recording Sheet*.



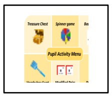
***Teach** (4/5 minutes):*

The required resources for the teaching input are (1) the vocabulary teaching cue card ideally on the whiteboard so it is visible to all, (2) the word of the day picture symbol and (3) the pupil-friendly definition.

Day 1: On the first day, you will need to read the story and preteach the nine target words. Pupils can be asked what they think the words mean, or they can match the word to the definition, etc. The vocabulary self-rating scale should be completed. These will need to be photocopied double-sided with the pre- and post-test for each child.

Days 2-10: Point to the symbol and say the word. Work through each cue on the teaching card, eliciting responses from the class, and providing models where needed. The aim is to provide 8-10 exposures to the word , since this will enable most learners to store the vocabulary item. Suggested wording is given alongside the cue card, and a simple standard definition is provided on the planning sheet. Teaching can take the format of question and answer and/or talk partners. On day 10, pupils can complete the vocabulary self-rating scale again, instead of a game.

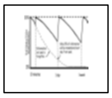
In the case of bank holidays, training days, teacher absence or special events, some suggestions for catch-up include: (1) using your TA to deliver the lesson (although not generally) or (2) doubling up on taught words that week .



Activate/Apply *(4/5 minutes):*

Pupils now have the opportunity to practise their word learning in an interactive game. The teaching cue card should be left on the whiteboard during activity time for easy reference. A range of five activities is provided.

Pupils can play the games in pairs or small groups. At the start, they will need more structure to learn the instructions and routines. Here are a few options suggested by other teachers. Option one: the class can be split in half, with the teacher and TA each supervising half the class. Option two: the groups can rotate through the activities over the week (requiring two copies of each game). Games have been provided in zip wallets to minimise organisational time. It is a good idea to have a quick system for passing out and collecting in the games, such as a games monitor/helper. A traffic light image is available on the word wall to indicate that quiet voice levels are appropriate during game time, so that pupils can clearly hear the language input.



Review *(2 minutes)*

After the activities are tidied away, a brief review session takes place with the whole class. Review words are provided on the *Teacher Planning and Recording Sheet*. The word of the day is reviewed, the previous day's word and the word taught one week previously. Any appropriate review activity could be used, e.g. the children can be given the definition and asked to say the relevant word.

The *Word Spy Record* can be placed on the word wall for staff or children to tick when they use the target word orally or in writing outside the lesson.

Resources provided:

Vocabulary teaching protocol
Teacher planning and review sheet
Instructional cue card + wording
Symbolised vocabulary cards
Vocabulary teaching blocks overview
Pupil self-rating scale
Pupil-friendly definitions
Games in wallets, masters
Word spy record
Word wall resources

2. Example Teacher Planning and Recording Sheet

Select: word of the day Start date:

Teach: vocabulary teaching card (5 minutes)

Activate: pupil activity (5 minutes)

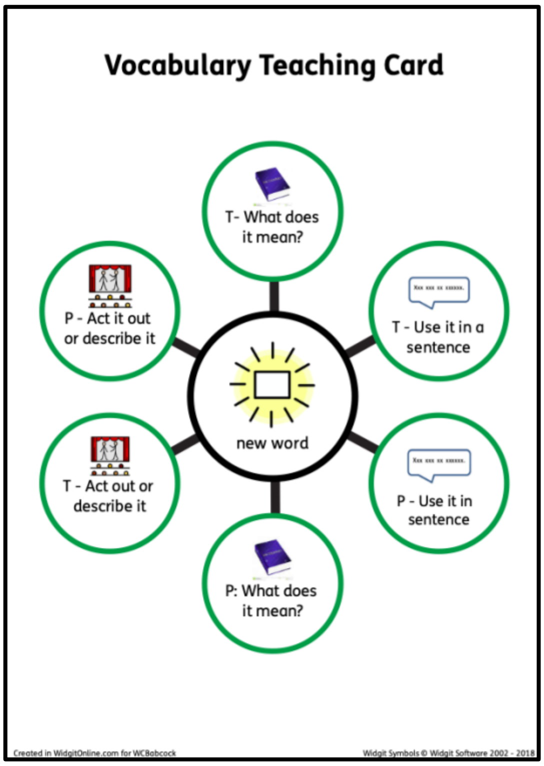
Review: today's word, yesterday's word, last week's word (2 minutes)

Augustus and his Smile

	Week One	**Week two**
Monday	Preteach vocabulary Read story	***swirled*** (swirl): to move around in a circle quickly Review: swirled, scaled, tuck
Tuesday	***crept*** (creep): to move quietly and slowly Review: crept, tuck, handsome	***frost***: a thin, light covering of ice, covered in ice crystals Review: frost, swirled, crept
Wednesday	***cluster***: a small group of things that are similar and close together Review: cluster, crept, lonely	***patterns*** (pattern): something that is repeated again and again, such as shapes, letters, numbers, or colours Review: patterns, frost, cluster
Thursday	***searched*** (search): to look carefully for something or someone Review: searched, cluster, building	***paraded*** (parade): to move down a public street together to celebrate something, usually with people watching Review: paraded, patterns, searched
Friday	***scaled*** (scale): to climb up a steep path or mountain; a hard climb Review: scaled, searched, hook	***realised*** (realise): when you suddenly understand something Review: realised, paraded, scaled
Evaluation		

1. <u>Teaching Cue Cards (A3 Size and Digital Format) with Suggested Wording</u>

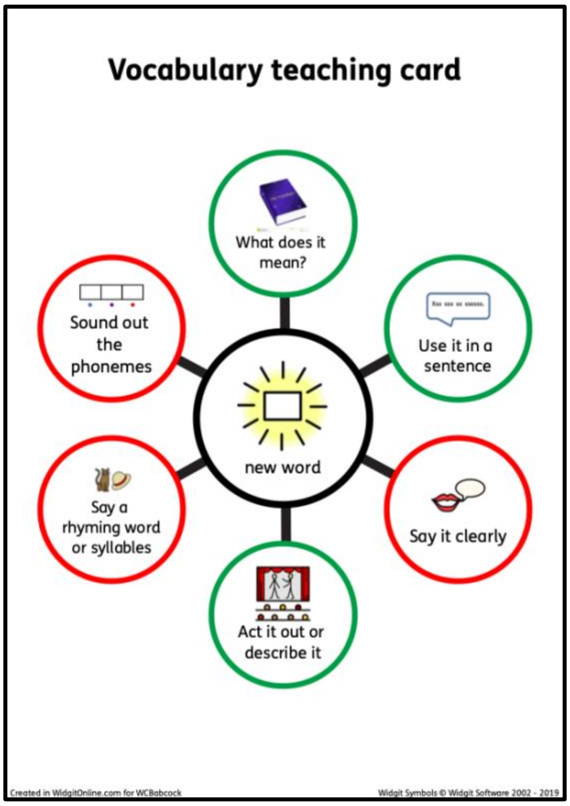
Semantic



Suggested wording (example)

Our word today is **pattern**. A **pattern** is a shape that happens again and again. I can use the word **pattern** in a sentence: We learn about **patterns** in maths. Can you use **pattern** in a sentence? *Teacher to repeat a pupil sentence containing* **pattern**. Can you say the meaning of **pattern**? I can act out/ describe what I know about **pattern** ..... Can you act out or describe to what you know about a **pattern**?

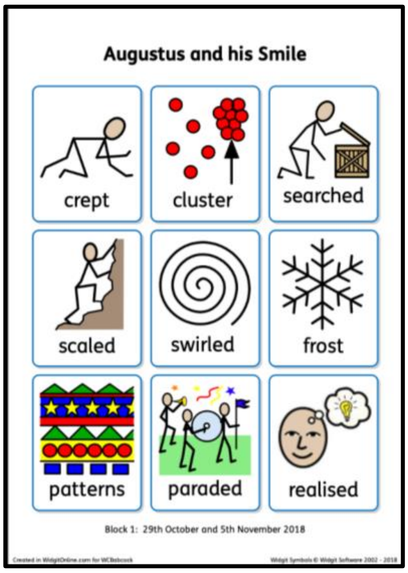
Combined

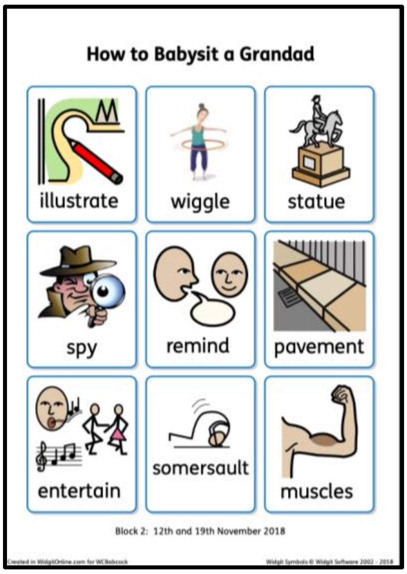


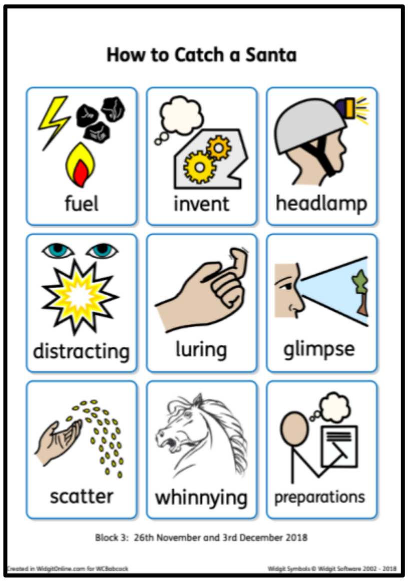
Suggested wording (example)

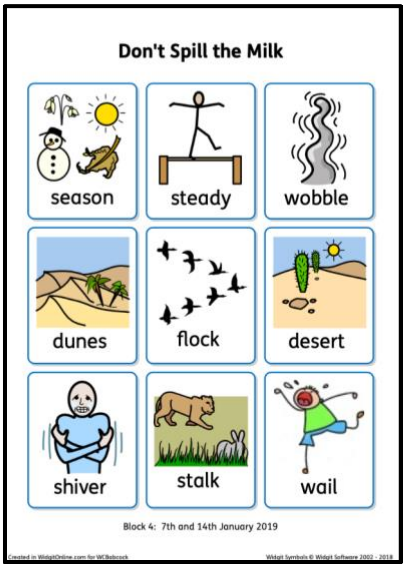
Our word today is **pattern**. A **pattern** is a shape that happens again and again. I can use the word **pattern** in a sentence: We learn about **patterns** in maths. Can you make up a sentence with **pattern**? Can you say the word **pattern** nice and clearly? Can you act out or describe what you know about **pattern**? Let's say a rhyming word / syllables for **pattern**. Let's say all the sounds we hear in **pattern**.

2. Symbolised Vocabulary Cards (A4)

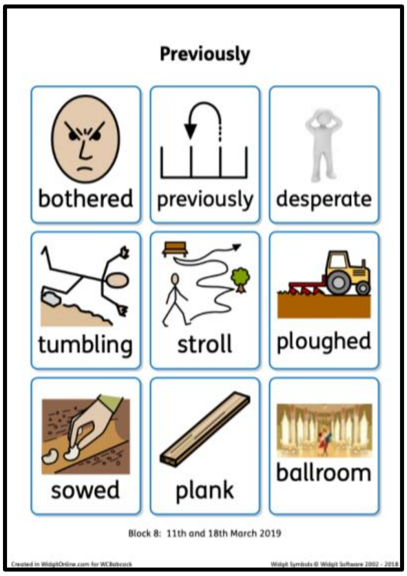








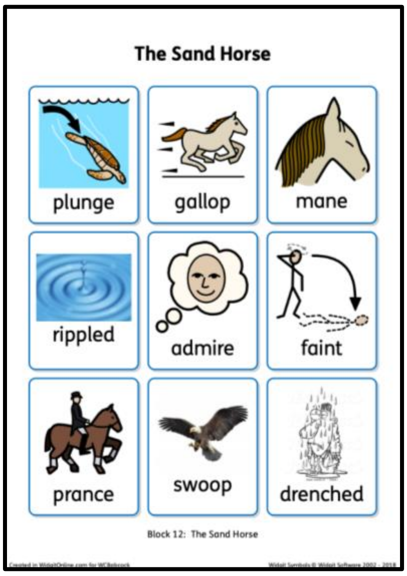

The Day Louis Got Eaten
gulp
grab
snatch
guzzle
manage
track
lair
wriggle
fled
Block 7: 25th February and 4th March 2019
Created in WidgitOnline.com for WCBabcock
Widgit Symbols © Widgit Software 2002 - 2018


Previously
bothered
previously
desperate
tumbling
stroll
ploughed
sowed
plank
ballroom
Block 8: 11th and 18th March 2019
Created in WidgitOnline.com for WCBabcock
Widgit Symbols © Widgit Software 2002 - 2018


Wanted the Perfect Pet
expenses
peace
varieties
personality
common
disguise
plopped
admit
disappointed
Block 9: 25th March and 1st April 2019
Created in WidgitOnline.com for WCBabcock
Widgit Symbols © Widgit Software 2002 - 2018


Traction Man is Here
helmet
guard
volunteered
dive
camouflage
waist
mysterious
rescue
cliff
Block 10: 22nd and 29th April 2019
Created in WidgitOnline.com for WCBabcock
Widgit Symbols © Widgit Software 2002 - 2018


Mrs Armitage on Wheels
OIL
greasy
tool
screwdriver
hammer
panting
faithful
pedal
exhausted
anchor
Block 11: 6th and 13th May 2019
Created in WidgitOnline.com for WCBabcock
Widgit Symbols © Widgit Software 2002 - 2018


The Sand Horse
plunge
gallop
mane
rippled
admire
faint
prance
swoop
drenched
Block 12: The Sand Horse
Created in WidgitOnline.com for WCBabcock
Widgit Symbols © Widgit Software 2002 - 2018

3. Vocabulary Blocks

Autumn Term 1	
Trial block: Naughty Bus	week of 8th and 15th October 2018
Preteach and self-rating scale	6. handsome
1. traffic	7. lonely
2. passenger	8. rescue
3. tight	9. hook
4. reflection	10. tuck + self -rating scale
Autumn Term 2	
Text 1: Augustus and His Smile	week of 29th October and 5th November 2018
Preteach and self-rating scale	6. swirled
1. crept	7. frost
2. cluster	8. patterns
3. searched	9. paraded
4. scaled	10. realised + self -rating scale
Text 2: How to Babysit a Grandad	week of 12th and 19th November 2018
Preteach and self-rating scale	6. remind
1. illustrate	7. pavement
2. wiggle	8. entertain
3. statue	9. somersault
4. spy	10. muscles + self -rating scale
Text 3: How to Catch a Santa	week of 26th November & 3rd December 2018
Preteach and self-rating scale	6. luring
1. fuel	7. glimpse
2. invent	8. scatter
3. headlamp	9. whinnying
4. distracting	10. preparations + self -rating scale
Catch up week	week of 10th December
Spring term 1	
Text 4: Don't Spill the Milk	week of 7th and 14th January 2019
Preteach and self-rating scale	5. flock
1. season	6. desert
2. steady	7. shiver
3. wobble	8. stalk
4. dunes	9. wail + self -rating scale
Text 5: How to Hide a Lion at School	week of 21st and 28th January 2019
Preteach and self-rating scale	6. spotted
1. sneak	7. fetched
2. noticed	8. hitched
3. museum	9. screeched
4. ancient	10. anxiously + self -rating scale

Text 6: Could a Penguin Ride a Bike?	week of 4th and 11th February 2019
Preteach and self-rating scale	6. webbed
1. imagine	7. choir
2. hatches	8. squawk
3. champion	9. huddle
4. waddle	10. beak + self -rating scale
Spring term 2	
Text 7: The Day Louis Got Eaten	week of 25th February and 4th March 2019
Preteach and self-rating scale	6. managed (to)
1. gulp	7. track
2. grab	8. lair
3. snatch	9. wriggle
4. guzzle	10. fled + self -rating scale
Text 8: Previously	week of 11th and 18th March 2019
Preteach and self-rating scale	5. stroll
1. bothered	6. ploughed
2. previously	7. sowed
3. desperate	8. plank
4. tumbling	9. ballroom + self -rating scale
Text 9: Wanted the Perfect Pet	week of 25th March and 1st April 2019
Preteach and self-rating scale	6. common
1. expenses	7. disguise
2. peace	8. plopped
3. varieties	9. admit
4. personality	10. disappointed + self -rating scale
Summer term 1	
Text 10: Traction Man is Here	week of 22nd and 29th April 2019
Preteach and self-rating scale	5. camouflage
1. helmet	6. waist
2. guarding	7. mysterious
3. volunteered	8. rescue
4. dive	9. cliff + self -rating scale
Text 11: Mrs Armitage on Wheels	week of 6th and 13th May 2019
Preteach and self-rating scale	6. panting
1. greasy	7. faithful
2. tool	8. pedal
3. screwdriver	9. exhausted
4. hammer	10. anchor + self -rating scale
Catch up week	week of 20th May
Summer term 2	
Text 12: The Sand Horse	week of 3rd and 10th June 2019
Preteach and self-rating scale	5. admire
1. plunge	6. faint
2. gallop	7. prance
3. mane	8. swoop
4. rippling	9. drench + self -rating scale

6. Vocabulary Self-Rating Scale

1-pre

Name:

Date:

Do you know what the word means?

Word	√ **Yes**	X **No**
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		

2-post

Do you know what the word means?

Word	√ Yes	X No
1.		
2.		
3.		
4.		
5.		
6.		
7.		
8.		
9.		

7. Pupil-Friendly Definitions

Block 1: Augustus and His Smile

crept (creep): to move quietly and slowly

cluster: a small group of things that are similar and close together

searched (search): to look carefully for something or someone

scaled (scale): to climb up a steep path or mountain; a hard climb

swirled (swirl): to move around in a circle quickly

frost: a thin, light covering of ice, covered in ice crystals

patterns (pattern): something that is repeated again and again, such as shapes, letters, numbers, or colours

paraded (parade): to move down a public street together to celebrate something, usually with people watching

realised (realise): when you suddenly understand something

Block 2: How to Babysit a Grandad

illustrate: to draw pictures to go with some writing, usually a book

wiggle: to move quickly from side to side

statue: a large piece of art, usually made of stone or metal. Statues are often in the shape of humans or animals

spy: to notice something **OR** to watch someone secretly **OR** a person who tries to find out secret information

remind: to help a person to remember

pavement: a footpath with a hard surface, usually next to a road

entertain: to amuse or keep someone interested

somersault : a roly poly; when your body rolls all the way over forward or backward

muscles: part of the body that makes our bones move and makes us strong

Block 3: How to Catch a Santa

fuel: what we burn for heat and energy, for example petrol or wood

invent: to think of an idea and make something that has never been made before

headlamp: a bright light, like a torch, worn on the head to help you see in the dark

distracting (distract): to take a person's attention away from what they are doing

luring (lure): to trap someone in to doing something for a reward

glimpse: a quick look

scatter: to throw something around in different directions, such as seeds **OR** when people move quickly in different directions

whinnying (whinny): making a long, high noise like a horse

preparations: getting everything ready, like for a trip

Block 4: Don't Spill the Milk

season: the four main parts of the year with different weather: spring, summer, autumn and winter

steady: balanced, not moving or shaking about

wobble: to move or tip from side to side in an unsteady way; to be out of balance

dunes (dune): a mound or hill of sand created by the wind

flock: a group of animals that stays together, such as birds, sheep or goats

desert: a hot, dry place with little rain and not many plants

shiver: to shake because you are cold or frightened

stalk: to follow quietly in a way that is like hunting **OR** a plant's main stem

wail: a long, loud cry of pain or sadness

Block 5: How to Hide a Lion at School

sneak: to move in a secret or creeping way **OR** to take something secretly

noticed (notice): to see or observe something

museum: a building where important and valuable items are kept, so they can be studied and to allow people to see them

ancient: very old; from a long time ago

spotted (spot): to see or notice someone **OR** marked with spots

fetched (fetch): to pick up something and bring it back

hitched (hitch): to ask for a ride in someone's car **OR** to fasten two things together

screeched *(*screech): to make a harsh, high sound like an owl

anxiously (anxious): feeling worried or nervous

Block 6: Could a Penguin Ride a Bike

imagine: to create a picture in your mind

hatches (hatch): when a baby bird comes out of its shell by breaking the egg from the inside

champion: a person that wins first place in a contest or game; the winner

waddle: to walk with short steps rocking from side to side, like a penguin or duck

webbed: when an animal's fingers or toes are joined together by a thin piece of skin to help them swim

choir: a group of people who sing together

squawk: to give a loud, high scream like a parrot

huddle: to stay close together in a small group

beak: the hard part of a bird's mouth

<u>Block 7: The Day Louis Got Eaten</u>

gulp: to eat, drink or breathe in quickly, so you swallow a lot at one time

grab: to pick something up quickly and roughly

snatch: to take/pull something away quickly or suddenly; grab

guzzle: to eat or drink a lot and very quickly

managed (to): to succeed in doing something difficult

track: to follow the footprints of a person or animal

lair: a wild animal's shelter or den

wriggle: to twist and turn your body quickly from side to side; wiggle

fled (flee): to run away or escape

<u>Block 8: Previously</u>

bothered (bother): to be annoyed or worried about something **OR** to make an effort

previously (previous): before

desperate: when you need something badly

tumbling (tumble): to roll over head first **OR** when something is falling

stroll: to go for a slow walk

ploughed (plough): to turn over the soil by using a plough, usually before seeds are planted

sowed (sow): to plant or scatter seeds over the ground

plank: a thick board of wood

ballroom: a large room used for dancing

<u>Block 9: Wanted the Perfect Pet</u>

expenses: money for bills and the things you need to live

peace: quiet, calm **OR** not at war

varieties (variety): different types of the same thing

personality: what you are like as a person; your character

common: found in large numbers **OR** when things are similar to each other

disguise: to change the way you look so you cannot be recognised **OR** a costume, camouflage or fancy dress

plopped (plop): to drop or fall in a heavy way

admit: to let someone enter **OR** to tell the truth

disappointed (disappoint): when you are unhappy because what you wished for did not happen

Block 10: Traction Man is Here

helmet: a hard hat worn to protect the head so it does not get hurt

guarding (guard): to protect something in order to keep it safe

volunteered (volunteer): to offer to do something for no pay

dive: to go deep underneath the water **OR** to jump down head first from a high place

camouflage: when the colour looks like the surroundings

waist: the middle narrower part of the body between the chest and hips

mysterious: something that is strange and can't be explained

rescue: to help someone out of danger and make them safe again

cliff: high land with a very steep side, usually next to the sea

Block 11: Mrs Armitage on Wheels

greasy: covered with grease; oily

tool: something people use to fix and repair items

screwdriver: a tool for turning a screw. A screwdriver has a handle for turning and a long metal piece that fits the head of the screw

hammer: a tool with a heavy metal head on a handle. A hammer is used to hit things such as nails

panting (pant): to breathe in and out quickly with short breaths

faithful: when you are loyal to a person, religion, or idea

pedal: a part of a machine which you push down with your foot to move

exhausted: worn out and tired, when we have used up nearly all our energy

anchor: a heavy metal object that is attached to a long rope. It can be dropped from a boat or ship to keep it from moving

Block 12: The Sand Horse

plunge: to dive or jump into something soft or liquid like water

gallop: to move or run quickly, like a horse

mane: the long hair on the neck of a horse, lion or other animal

swoop: to dive down suddenly from above (in the air)

rippling (ripple): to move or flow in small waves

admire: to respect and really like someone

faint: hardly noticeable **OR** when you pass out

prance: to move or with quick high steps so people will notice

drench: to soak or make completely wet

7. Pupil Games

(Images taken from Widgit Symbols © Widgit Software 2002 – 2018 and Parsons & Branaghan, 2014)

Instructions	Semantic	Combined

Beetle game

Resources: Foam dice, whiteboard, whiteboard pens.

Instructions: Player one rolls the dice, completes the relevant task on the cue card and draws the indicated part of the beetle on the whiteboard. Play continues until the beetle is complete.

Semantic

Die number	Task to complete	Body part to draw
1	What does the word mean?	Eyes
2	Use the word in a sentence.	Antennae
3	Act out or describe the word.	Legs
4	What does the word mean?	Wings
5	Use the word in a sentence.	Head
6	Act out or describe the word.	Body

Combined

Dice number	Vocabulary cue	Body part to draw
1	Say it aloud.	Eyes
2	Say a rhyming word or syllables.	Antennae
3	Sound it out. (phonemes)	Legs
4	What does the word mean?	Wings
5	Use the word in a sentence.	Head
6	Act out or describe the word.	Body

Dice game

Resources: Foam dice.

Instructions: The first person rolls the dice and carries out the corresponding instruction on the game board.

Semantic

1 Use it in a sentence	2 What does it mean?	3 Act out or describe the word
4 Act out or describe the word	5 Use it in a sentence	6 What does it mean?

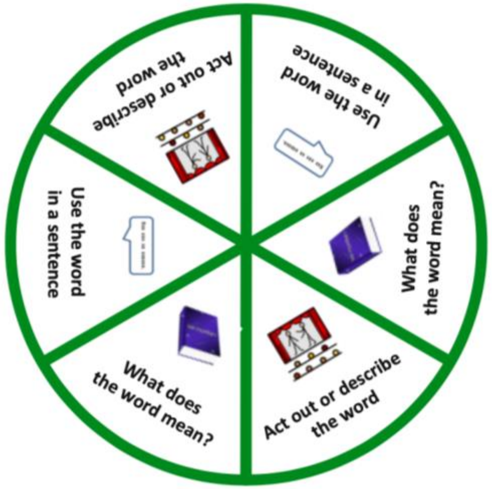
Combined

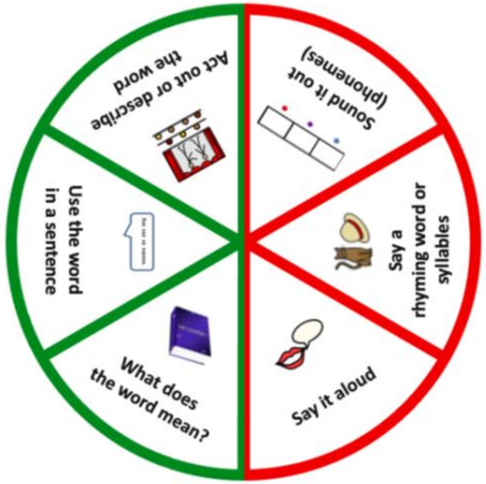
1 Sound it out (phonemes)	2 Say it aloud	3 Say a rhyming word or syllables
4 Act out or describe the word	5 Use the word in a sentence	6 What does the word mean?

Spinner game

Resources: Vocabulary cue spinner

Instructions: The first child spins and says the corresponding cue for the word of the day. Children take it in turns to spin the wheel and say the cue.

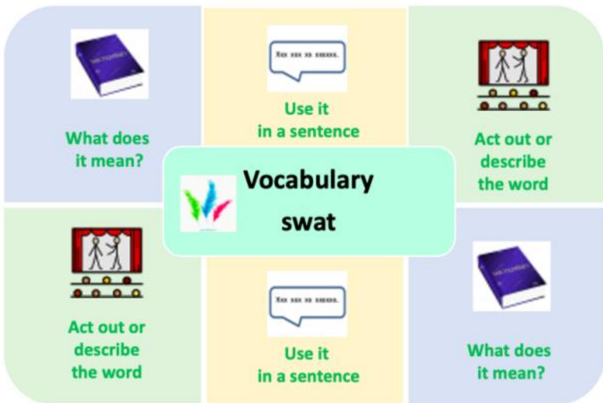


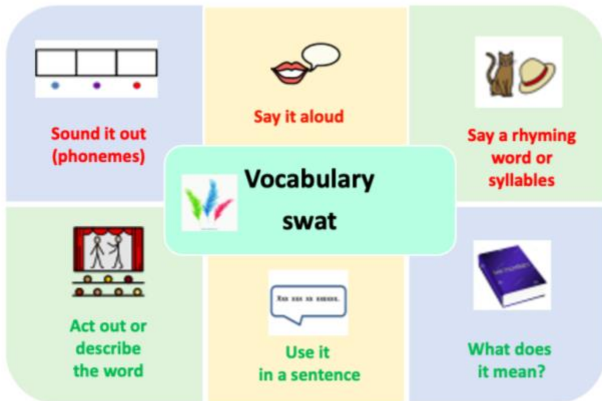


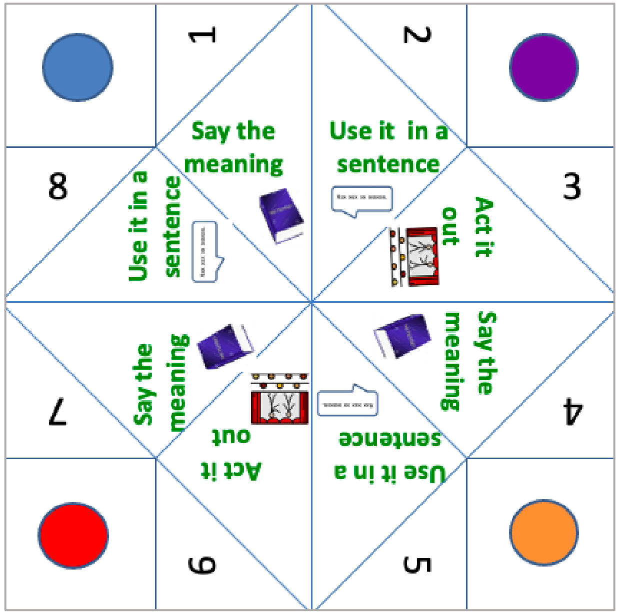
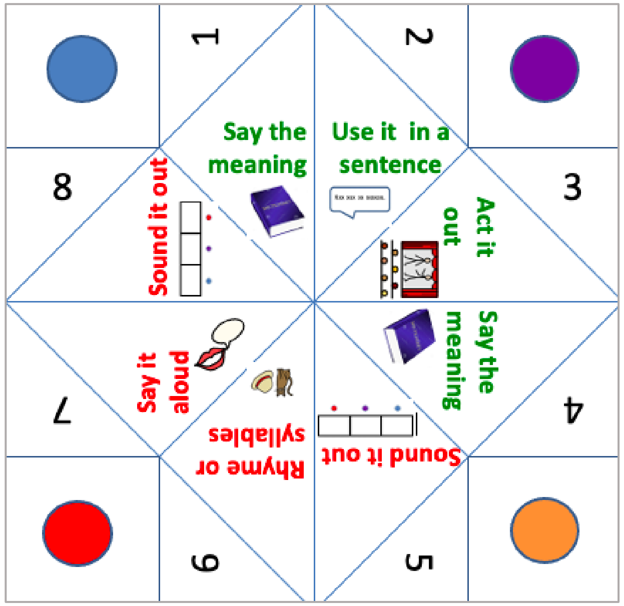
Vocabulary swat

Resources: Feather swatters

Instructions: Pupil 1 closes his/her eyes while the others turn the game card around. When ready, pupil 1 swats the card with the feather and carries out the cue.



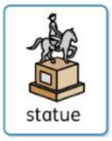


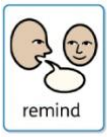
Fortune teller *Resources:* None *Instructions:* Pupil 1 holds the fortune teller and asks pupil 2 to choose a colour and a number. Pupil 1 moves the fortune teller that number of times and reveals the cue below. Pupil 2 carries out the cue.	 

9. Example Word Spy Record

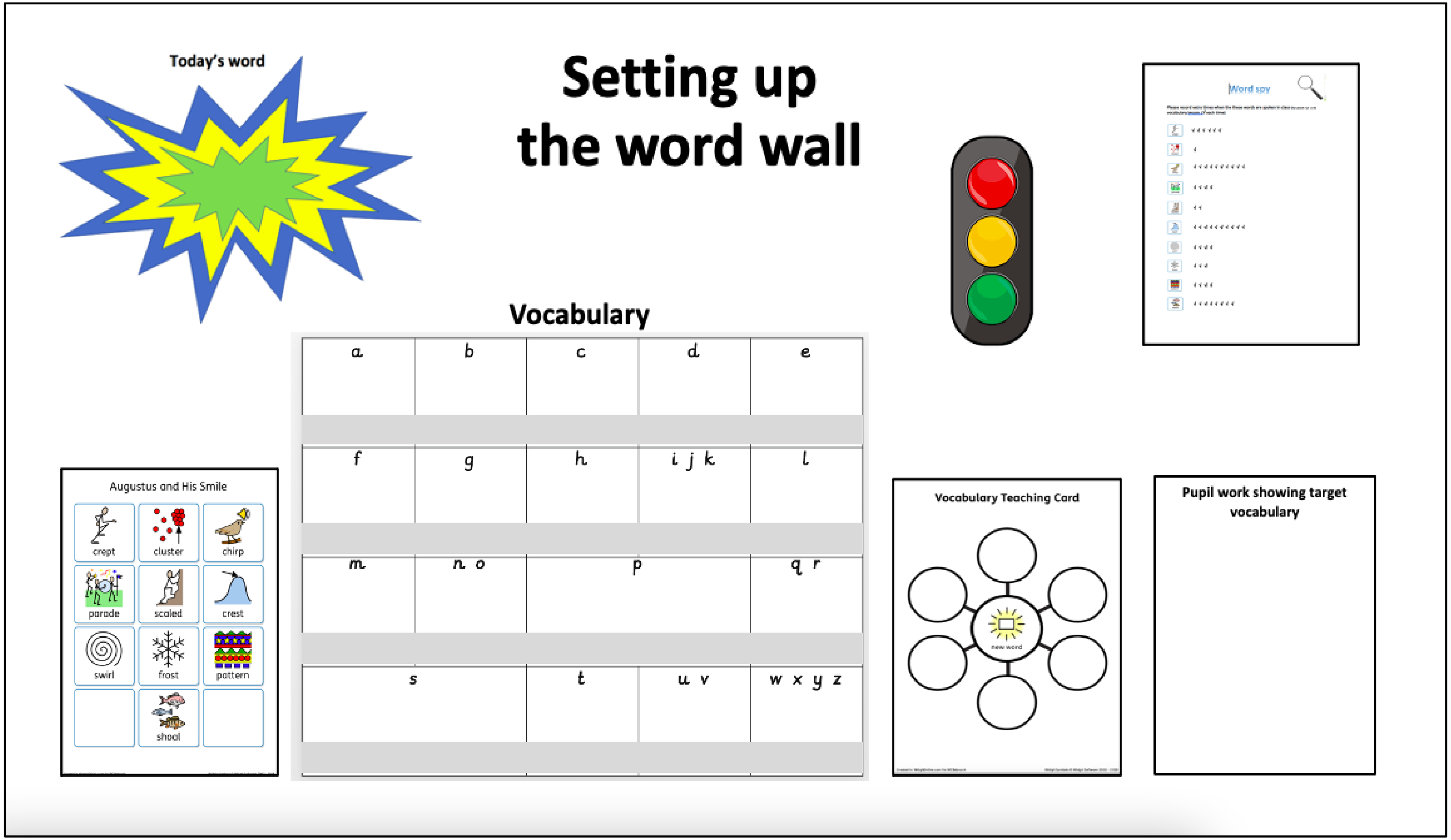
Block 2: How to Babysit a Grandad

Please tick each time these words are spoken in class outside of the vocabulary lesson





10. Word Wall Example and Resources



(Images taken from Widgit Symbols © Widgit Software 2002 – 2018)

Appendix Q

Games Checklists for Technique Feature Analysis

(Involvement Load; Nation & Webb, 2011)

Semantic Condition **Criteria**	Spinner Game	Beetle Game	Dice Game	Modified Pairs	Vocabulary Swat
Motivation					
Is there a clear vocabulary learning goal?	1	1	1	1	1
Does the activity motivate learning?	1	1	0	1	1
Do the learners select the words?	0	0	0	1	1
Noticing					
Does the activity focus attention on the target words	1	1	1	1	1
Does the activity raise awareness of new vocabulary learning?	1	1	1	1	1
Does the activity involve negotiation?	0	1	0	1	0
Retrieval					
Does the activity involve retrieval of the word?	0	0	0	0	0
Is it productive retrieval?	0	0	0	0	0
Is it recall?	0	0	0	0	0
Are there multiple retrievals of each word?	0	0	0	0	0
Is there spacing between retrievals?	0	0	0	0	0
Generation (novel use)					
Does the activity involve generative use?	1	1	1	1	1
Is it productive?	1	1	1	1	1
Is there a marked change that involves the use of other words?	1	1	1	1	1
Retention					
Does the activity ensure successful linking of form and meaning?	1	1	1	1	1
Does the activity involve instantiation? (visually present)	1	1	1	1	1
Does the activity involve imaging?	0	0	0	0	0
Does the activity avoid interference? (same category)	1	1	1	1	1
Total score (maximum 18)	10	11	9	12	11

Semantic Condition (updates)	Fortune Teller	Vocab Swat
Criteria		
Motivation		
Is there a clear vocabulary learning goal?	0	1
Does the activity motivate learning?	1	1
Do the learners select the words?	0	0
Noticing		
Does the activity focus attention on the target words	1	1
Does the activity raise awareness of new vocabulary learning?	1	1
Does the activity involve negotiation?	1	1
Retrieval		
Does the activity involve retrieval of the word?	1	1
Is it productive retrieval?	1	1
Is it recall?	1	1
Are there multiple retrievals of each word?	0	0
Is there spacing between retrievals?	1	1
Generation (novel use)		
Does the activity involve generative use?	1	1
Is it productive?	1	1
Is there a marked change that involves the use of other words?	0	0
Retention		
Does the activity ensure successful linking of form and meaning?	1	1
Does the activity involve instantiation? (visually present)	1	1
Does the activity involve imaging?	1	0
Does the activity avoid interference? (same category)	1	1
Total score (maximum 18)	13	14

Combined Condition	Spinner Game	Beetle Game	Dice Game	Modified Pairs	Vocab Swat
Criteria					
Motivation					
Is there a clear vocabulary learning goal?	1	1	1	1	1
Does the activity motivate learning?	1	1	0	1	1
Do the learners select the words?	0	0	0	1	1
Noticing					
Does the activity focus attention on the target words	1	1	1	1	1
Does the activity raise awareness of new vocabulary learning?	1	1	1	1	1
Does the activity involve negotiation?	0	1	0	1	0
Retrieval					
Does the activity involve retrieval of the word?	1	1	1	1	1
Is it productive retrieval?	0	0	0	1	1
Is it recall?	1	1	1	1	1
Are there multiple retrievals of each word?	1	1	1	1	1
Is there spacing between retrievals?	0	0	0	1	0
Generation (novel use)					
Does the activity involve generative use?	0	0	0	0	0
Is it productive?	0	0	0	0	0
Is there a marked change that involves the use of other words?	0	0	0	0	0
Retention					
Does the activity ensure successful linking of form and meaning?	1	1	1	1	1
Does the activity involve instantiation? (visually present)	1	1	1	1	1
Does the activity involve imaging?	0	0	0	0	0
Does the activity avoid interference? (same category)	1	1	1	1	1
Total score (maximum 18)	10	11	9	14	12

Combined Condition (updates)	Fortune Teller	Vocab Swat
Criteria		
Motivation		
Is there a clear vocabulary learning goal?	0	1
Does the activity motivate learning?	1	1
Do the learners select the words?	0	0
Noticing		
Does the activity focus attention on the target words	1	1
Does the activity raise awareness of new vocabulary learning?	1	1
Does the activity involve negotiation?	1	1
Retrieval		
Does the activity involve retrieval of the word?	1	1
Is it productive retrieval?	1	1
Is it recall?	1	1
Are there multiple retrievals of each word?	0	0
Is there spacing between retrievals?	1	1
Generation (novel use)		
Does the activity involve generative use?	1	1
Is it productive?	1	1
Is there a marked change that involves the use of other words?	0	0
Retention		
Does the activity ensure successful linking of form and meaning?	1	1
Does the activity involve instantiation? (visually present)	1	1
Does the activity involve imaging?	1	0
Does the activity avoid interference? (same category)	1	1
Total score (maximum 18)	13	14

Appendix R

Vocabulary Training Session Presentation

Year One Vocabulary Project
Teacher training session

Rose Brooks
PhD research study
October 2019

The University Of Sheffield.

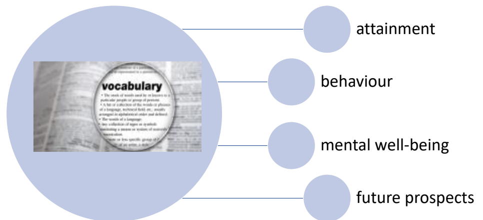
Overview

- Project overview
- Classroom organisation /setup
- The STAR approach
  - Selecting vocabulary
  - Teach (direct instruction)
  - Activate (pupil activities)
  - Review

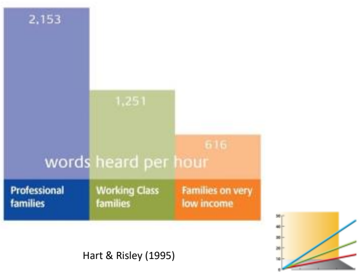
Project overview

- The importance of vocabulary
- What is involved
- Timeline and projected outcomes

The importance of vocabulary

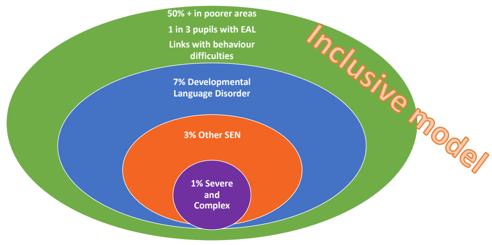


The vocabulary gap



Hart & Risley (1995)

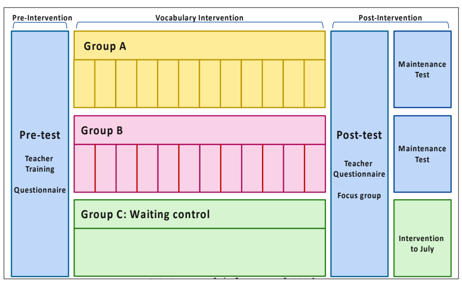
Many pupils may benefit from explicit vocabulary instruction



What is involved

- Setting up a word wall in the classroom
- Teaching a word a day linked to a high quality text for 10-12 minutes
- 12 teaching blocks of around two weeks each
- Email or phone contact to answer questions and share your thoughts
- All resources are provided, apart from the texts
- The teacher learns the planning and teaching approach over the course of a year, and then helps to cascade to other staff

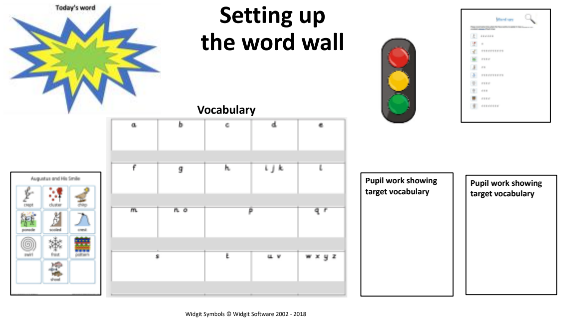
Project outcomes



Classroom organization/setup

- Setting up the word wall
  - Purpose: review, application
  - How and when to use
  - Voice levels
- Teaching the games
- Class routines

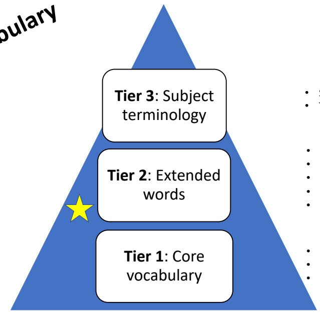
Setting up the word wall



The STAR approach

- Select: tier 2 vocabulary
- Teach: use vocabulary cue card
- Activate: menu of pupil activities
- Review: words from today, yesterday, last week

Select vocabulary



- Subject specific words
- Topic vocabulary

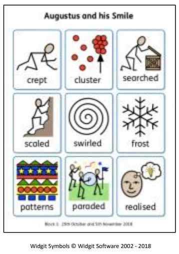
- Book vocabulary
- Wow words
- Thinking verbs
- Alternatives / synonyms
- Detail

- Everyday vocabulary
- Conversational
- Common categories

(Beck, 2002)

Widgit Symbols © Widgit Software 2002 - 2018

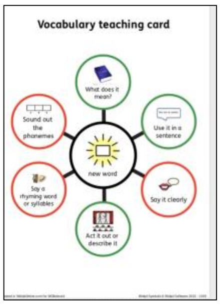
Item	Lemma	Age of acquisition Kuperman et al. (2012)	Written frequency Kucera & Francis (1967)	Word type	Word length: No. of syllables
crept	creep	8.00	10	verb	1
cluster	cluster	10.00	13	adj	2
searched	search	8.72	66	verb	1
scaled	scale	8.85	60	verb	1
swirled	swirl	6.67	2	verb	1
frost	frost	5.21	6	noun	1
patterns	pattern	6.89	113	noun	2
paraded	parade	6.53	25	verb	3
realised	realise	8.23	69	verb	3



Widgit Symbols © Widgit Software 2002 - 2018

Texts

1. **Augustus and his Smile** C. Rayner
2. **How to Babysit a Grandad** J. Reagan
3. **How to Catch Santa** Reagan & Wildish
4. **Don't Spill the Milk** Davies & Corr
5. **How to Hide a Lion at School** H. Stephens
6. **Could a Penguin Ride a Bike?** Bitskoff & Bedoyere
7. **The Day Louis Got Eaten** J. Fardell
8. **Previously** A. Ahlberg
9. **Wanted the Perfect Pet** F. Robertson
10. **Traction Man is Here** M. Grey
11. **Mrs Armitage on Wheels** Q. Blake
12. **The Sand Horse** A. Turnbull



Teach: 5 minutes

Day 1: Preteach vocabulary and read story

Day 2-10: Teach word, activity, review

Optional: day 1 and 10 Pupil Self Rating Scale

8-10 repetitions

Can elicit responses from pupils and / or use response partners

Teaching cue card and word of the day visible

Activate

Spinner game	*Vocabulary Activity Menu*	*Beetle game*
Vocabulary Swat	*Modified Pairs*	*Dice Lotto*

Review

- Review schedule - to support memory
- Vocabulary self-rating scale(optional)
- Word wall to support application - spoken and written
- Word spy (optional) to support application

Vocabulary programme resources

- Standardised teaching protocol
- Vocabulary lists
- Vocabulary planning and recording sheets
- Vocabulary teaching cue card
- Definitions
- Symbolised target vocabulary cards
- Pupil activities and instructions
- Word wall resources

Optional: Word spy record
Optional: Pupil self-rating scale

Appendix S

Teacher Questionnaires

Pre-intervention Questionnaire for Teachers - September 2018

Name: ***School:***

The purpose of this questionnaire is to gather information about teachers' views and professional development prior to taking part in the vocabulary research project. A similar questionnaire will be provided in July.

<table>
<tr><td></td><td colspan="5">Please circle the number 1-5 to represent your response</td></tr>
<tr><td></td><td>Not important</td><td></td><td>Average importance</td><td></td><td>Very important</td></tr>
<tr><td>1. How important is vocabulary teaching?</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td colspan="6">2. Please give reasons for your answer above.</td></tr>
<tr><td></td><td>Not confident</td><td></td><td>Moderately confident</td><td></td><td>Very confident</td></tr>
<tr><td>3. How confident do you feel about teaching new vocabulary?</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td></tr>
<tr><td colspan="6">4. Can you describe the vocabulary teaching and activities you have used, including how frequently you use them?</td></tr>
<tr><td colspan="6">5. What are the main barriers to classroom vocabulary instruction in your view?</td></tr>
<tr><td colspan="6">How many years have you been teaching, and in which year groups?</td></tr>
<tr><td colspan="6">6. What training opportunities have you received to support the teaching of speech and language or vocabulary?</td></tr>
<tr><td colspan="6">Thank you so much for your time in completing this questionnaire.</td></tr>
</table>

Post-intervention Questionnaire for Teachers - July 2019

Name: ***School:***

The purpose of this questionnaire is to gather information about teachers' views and professional development after taking part in the vocabulary research project.

Please circle the number 1-5 to represent your response.

	Not *confident*	*Slightly confident*	*Moderately confident*	*Quite confident*	***Very*** *confident*
1. How *confident* do you feel about teaching new vocabulary?	1	2	3	4	5

2. How *effective* was the programme for improving these aspects of the taught vocabulary?

	Not *effective*	*Slightly effective*	*Moderately effective*	*Quite effective*	***Very*** *effective*
Overall	1	2	3	4	5
Vocabulary understanding	1	2	3	4	5
Vocabulary in talk	1	2	3	4	5
Vocabulary in writing	1	2	3	4	5
Comments:					

3. How *effective* were these strategies for improving children's taught vocabulary?

	Not *effective*	*Slightly effective*	*Moderately effective*	*Quite effective*	***Very*** *effective*
Fiction story books	1	2	3	4	5
The daily lesson – STAR approach	1	2	3	4	5
Picture symbols	1	2	3	4	5
Teaching cue card	1	2	3	4	5
Pupil-friendly definitions	1	2	3	4	5
Teacher planning sheet	1	2	3	4	5
Pupil games	1	2	3	4	5
Pupil self-rating scale	1	2	3	4	5
Word spy record	1	2	3	4	5
Word wall	1	2	3	4	5
Comments:					

4. How likely is it that you will *use* these strategies again in classroom vocabulary teaching?

	Not *likely*	*Slightly likely*	*Moderately likely*	*Quite likely*	***Very*** *likely*
Fiction story books	1	2	3	4	5
The daily lesson – STAR approach	1	2	3	4	5
Picture symbols	1	2	3	4	5
Teaching cue card	1	2	3	4	5
Pupil-friendly definitions	1	2	3	4	5
Teacher planning sheet	1	2	3	4	5
Pupil games	1	2	3	4	5
Pupil self-rating scale	1	2	3	4	5
Word spy record	1	2	3	4	5
Word wall	1	2	3	4	5
Comments:					

	Not *enjoyed*	*Slightly enjoyed*	*Moderately enjoyed*	*Quite enjoyed*	***Very*** *much enjoyed*
5. **To what extent did you *enjoy* delivering the Year 1 vocabulary programme?**	1	2	3	4	5
Comments:					

	Not *changed*	*Slightly changed*	*Moderately changed*	*Quite changed*	***Very*** *much changed*
6. To what extent has participation in the programme *changed your* vocabulary teaching *practice?*	1	2	3	4	5
Comments:					

Please comment on any challenges or suggested changes to the programme.

Any other comments?

Thank you so much for completing the questionnaire and the Year 1 vocabulary project.

Responses - Pre-intervention Teacher Questionnaire

	How important is vocabulary teaching? (1=low;-5=high)	*How confident do you feel about vocabulary teaching? (1=low; 5=high)*	*Describe the vocabulary teaching activities you have used*	*How many years have you been teaching?*	*Which age groups?*	*What vocabulary training opportunities have you had?*
Combined group						
School A Class 1	5	4	Incidental – stories and subject lessons	5	Year 1 & 2	Part of a school INSET session
School A Class 2	5	2	Incidental – stories and subject lessons	9	All	Part of a school INSET session
School E Class 7	4	3	Incidental - stories	14	Year 1 & 2	None
School G Class 9	4	4	Incidental - stories Word wall	7	Year 1-6	Tier 2 vocabulary
Semantic group						
School B Class 3	4	3	Visual imagery	3	Year 1 & 4	Not much
School C Class 4	5	4	Incidental - stories	4	Year 1	Not much
School C Class 5	4	2	Word wall Vocabulary banks Talk time	1	Year 1	Oracy training
School D Class 6	5	2	Word of the week Act out words	1	Year 1	Talk for writing
School F Class 8	5	3	Incidental - stories Act out words	1	Year 1	None
Control group[3]						
School H (5 classes)	4	3.5 (3-4)	Incidental – stories and subject lessons	2, 4, 5, 6, 10 *M*=5.4	All	School INSET session
School I (2 classes)	5	3	Incidental – stories Preteaching groups	3, 5 *M*=4	Key Stage 1	School INSET session

[3] The headteacher of the two control schools completed the questionnaire as a summary of participating class teachers.

Appendix T

Consistency Checklist for Lesson Observations

Lesson observation date:

	SELECT
☐	1. The targeted word is taught daily, with notations made on planning sheet if lessons are missed
	TEACH
☐	2. Cues on vocabulary teaching card are included in lesson
☐	3. Approximate timings are followed: 5 minutes teaching, 5 minutes game, 2 minutes review
☐	4. Simple definitions are used, although these can be expanded
	ACTIVATE
☐	5. Games are updated fortnightly with the new vocabulary cards
☐	6. Noise levels are monitored during activity time
☐	7. Pupils have access to the full range of games over the week / fortnight
	REVIEW
☐	8. Words are reviewed according to the planning sheet
☐	9. Pupil Self Rating Scales are completed at the start and end of each fortnightly teaching block
☐	10. The Word Wall is updated daily with the new target word and fortnightly for the new block

TOTAL SCORE: /10

COMMENTS: