Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Which skills influence pre-school children's repetition of words, non-words and sentences?

Hannah Jane Hockey

October 2014

City University London
Department of Language and Communication Science
School of Health Sciences

THE FOLLOWING PARTS OF THIS THESIS HAVE BEEN REDACTED FOR COPYRIGHT REASONS:

pp.67-69: Photographs in Table 2-7

p.89: Photographs in Table 2-10

p.91: Photographs in Table 2-12

p.95: Photograph in Figures 2-1 & 2-2

p.98: Photographs in Figure 2-3

p.264: Photographs and contact details in information sheet in Appendix D

p.266: Photograph and contact details in information sheet in Appendix E

Table of Contents

Chapter 1: Introduction

Chapter 2: Methods: Selecting the stimuli; scoring the tasks and procedure for collecting the data

List of Tables

Table 4-2 Effects of item length on known words and non-words at T2

List of Figures

List of Appendices

Acknowledgements

The journey from setting out on this project, through to its completion has been long and not without complication. I’ve been very lucky to have had the company, guidance and support of many *teachers*, friends and companions along the way, whose influences have enabled me to develop the skills I’ve needed to navigate the various challenges.

First, I would like to thank all the children who took part in the study. Also their parents who agreed for their participation and who welcomed me into their homes with cups of tea and chats. I have learnt so much from the children, who are astonishingly wise for having lived so few years, and whose enthusiasm and quirkiness brought cheer to the days of data collection and analysis.

Next I would like to thank my PhD supervisors and brilliant role models, Chloë Marshall and Lucy Dipper. Having spent a year without my project and feeling quite bruised and injured, Chloë and Lucy bravely and kindly took on the role of PhD supervisors midway through my PhD. I have learnt so much from them both and am very grateful for their support, positivity, knowledge and outward calmness. Chloë seems to know exactly when to provide support and encouragement, and when to *crack the whip*. Her high research standards and attention to detail have helped me to maintain a belief in research. Lucy is a great role model in balance and has an uncanny ability to see the bigger picture and to hear the underlying message. She has helped me to recognise that intuition and insight also have an important place in scientific thinking. Thanks particularly to them during the final 2-3 months of the write-up, during which I’ve expressed to them just about every emotion, and to which they have responded with kindness, calm and centeredness.

I would also like to thank my original project supervisors, Penny Roy and Shula Chiat who were very generous with their time in the initial stages of planning the project, and with their thoughts throughout the project. I am also grateful for their support throughout the many funding applications.

I would like to thank my parents, Jane and Roger Hockey for their love, encouragement and support throughout my life and particularly along the PhD path. And thank you to my little brother, Will Hockey who together with my parents literally rescued me twice from total confusion and despair during some terrifying moments in the PhD journey. Also thanks to my big sister, Beth Kuipers von Lande and her cuties, Arwen and Ewan, for their help in the study and for providing sunny, warm and peaceful mid-winter retreats.

I am also very grateful to Gila Falkus, my first speech therapy supervisor and manager, who encouraged me several years ago now to embark on the PhD path. Gila has given me emotional and practical support throughout the journey, particularly during the first half. She is probably the most honest and fairest person I know. She is certainly the best manager that I have had, and whenever I am unsure how to handle a situation at work I think about how Gila would respond.

Tim Pring has also provided support and encouragement to me throughout my PhD, also initially encouraging me to set out on the PhD journey. Especially thanks to Tim and also Nikki Botting for helping me to gain funding for my project.

I am very grateful to Sarah May, formerly of the Waterloo Foundation, for her ‘adapted’ and informal funding interview strategy that included a large pot of tea and plate of biscuits. And thank you so much to the Waterloo Foundation for funding the final 3 ½ years of the project. Sarah also provided welcome guidance and encouragement while she was working in the Foundation.

I am grateful as well to the various PhD students with whom I have worked over the past few years. Particularly I would like to thank Becky Moss, Maddie Pritchard, Anne Maine and Robert Stocker. And thank you too to the speech therapy BSc students who gave their time to transcribe some of the videos for the purposes of reliability.

CLCH was very generous in funding my PhD fees in the first two years. Thank you too to my various speech therapy teams in CLCH who have been so helpful in identifying possible participants and in encouraging me to keep going with the project. Special mention to Clara Morris, Stephanie Lynham, Severine Laftaly, Tashia Pillay and Ciara Tilley. And also to Claire Withey, who has met with me several times to provide her insight and wisdom on many levels.

I would like to thank Tracy Rowson, Jill Francis and Carol Cox who patiently listened and responded to my correspondence during my role as PhD rep. They have also been a great source of encouragement with their calm and thoughtful influences.

Extra special thanks to Penny Cooper, who floated in and out from the ether in true Mary Poppins-style. Penny worked intensively with me for 6 months during a break in my PhD, acting as a mentor and confidante. She helped me to develop my confidence in finding my voice and in presenting my research. Through her kindness, openness and reflection, she also helped unlock some of my awareness of the power and awesome mysticism of the unconscious mind. Hopefully through her guidance I have become a slightly wiser monkey.

Thank you so much to Connor Maloney, Zoë Wharfe and David Glyn who helped me through some tricky navigations as well.

Thank you also to the Careers Service at St Andrews who suggested Speech Therapy as a career. I’d never heard of it, but speech therapy matched exactly the criteria that I set out for a career. Also to my Psychology final year project supervisor Arlene Astell, who initially suggested the PhD path, but who I ignored, in favour of a career in speech therapy. I know too that my research approach has been greatly influenced by Gerry Quinn, who had an inspiring faith in his students’ motivation, ability to learn and curious minds. At City, I would like to thank Jane Marshall who guided me through my MSc project and who also encouraged me to begin a PhD.

I would also like to thank my lovely friends who have distracted me from my PhD with much needed fun, relaxation and adventures along the way. They have rescued me from insanity and kept me mostly socially acceptable. Especially to Kate Bunting, Katherine Scantlebury, Jayne Nicholls, Kathryn Harvey, Alex Jeyaseelen, Kate Law, Jo Kiddell, Penny Eskell and Anna Smyk. And I owe a huge debt of gratitude to my very loving, patient and kind boyfriend, Chris Lewis who has prised me away from my laptop at various points to encourage me to interact with the world outside and to remind me of the truly important things.

And finally, I would like to thank Mary Poppins, who is my absolute idol and true inspiration. I aspire to be like her: a model of calmness, kindness and assertiveness; someone who has managed skilfully to balance rational thought, imagination, her magical powers and mysticism. And it is Mary who should receive all the credit for the invention of non-words, with her practically perfect one: supercalifragilisticexpialidocious.

Abstract

This study explores the role of existing language knowledge and phonological short-term memory (PSTM) on pre-school children's non-word, word and sentence repetition (NWR, WR and SR). Previous studies have revealed that children with language difficulties find these tasks difficult, but there is debate about which skills are measured. This study aimed to contribute to this understanding. Identification of the underlying skills would enable speech therapists to plan targeted therapy to support the children's difficulties.

Data was collected at two time points: at time one from fifty-four participants, aged 3-3 ½ years old; and at time two from fifty-two of the original sample (aged 4 -4 ½ years).

The study is split into four parts. First it explores three influences on the children's WR and NWR: knowledge of the words, speech sound skills and PSTM, at both time-points. The second part divides the group into children with and without identified speech and language difficulties. It explores differences in performance by the two groups. Part three explores the influence of grammar (morphology) and PSTM on sentence repetition. Part four investigates relationships between children's NWR and WR at both time points with their SR at the second time-point.

There was evidence at both time-points that children draw on long-term word knowledge during WR and no evidence of them using PSTM in this task. There was a clear influence of PSTM on their NWR. The children's speech affected both NWR and WR. The clinical group repeated both known words and non-words less accurately than the non-clinical group. They showed a similar pattern of performance in their repetition of non-words, but achieved lower scores across all syllable lengths.

Children aged 4 years used existing grammatical skills when repeating sentences. There was limited evidence of the influence of PSTM. A correlation was found between children's NWR and later SR. The relationship was due to the influence of language knowledge and PSTM on both tasks.

Results from the study suggest that for both NWR and SR language knowledge and PTSM interact in their effect on accuracy. The tasks are however useful clinically because children's scores are influenced by their existing language knowledge.

List of Abbreviations

(Alphabetical order)

BAS	British Ability Scales
BPVS	British Picture Vocabulary Scale
C	Consonant
CELF	Clinical Evaluation of Language Fundamentals
CNREP	Children’s Test of Nonword Repetition
CEYS	Community Early Years Service
DfES	Department for Education and Skills
DLS	Derbyshire Language Scheme
ERB	Early Repetition Battery
GAPS	Grammar and Phonology Screener
NHS	National Health Service
NRT	Nonword Repetition Test
NS	Non-significant
NWR	Non-word repetition
PLS	Pre-School Language Scales
PSREP	Pre-School Repetition Test
PSTM	Phonological short-term memory
SD	Standard Deviation
SIT	Sentence Imitation Test
SLI	Specific Language Impairment
SLT	Speech and Language Therapy
SR	Sentence repetition
T1	Time 1 (data collection time-point)
T2	Time 2 (data collection time-point)
V	Vowel
WR	Word repetition
WSD	Weak Syllable Deletion

Chapter 1

Introduction

1.1. Spoken language disorders

A spoken language disorder, commonly referred to as a Specific Language Impairment (hereafter: SLI), and sometimes as a Primary Language Impairment, a Developmental Language Disorder or a Language Learning Impairment, is estimated to affect between 3 and 7% of the population (Tomblin, Records, Buckwalter, Zhang, Smith and O'Brien, 1997). Estimates of prevalence vary however according to different classifications (see recent reviews about the label of SLI by Bishop, 2014; Reilly, Bishop and Tomblin, 2014 and Reilly, Tomblin, Law, McKean, Mensah, Morgan, Goldfeld, Nicholson and Wake, 2014). Where receptive language is affected, people who have a SLI typically find it difficult to understand spoken questions, to follow spoken directions and to learn new words. Difficulties encountered by these individuals in the domain of expressive language include retrieving the correct word when talking, constructing grammatically correct sentences, using the correct grammatical morphemes, and with narrating stories. People with these disorders sometimes also have difficulty pronouncing words, or they have a history of pronunciation difficulties. In spite of the language difficulties, the term SLI has generally been applied where there is no concern about other cognitive skills and where there is a discrepancy between the two types of skills in favour of non-verbal ability (International Classification of Diseases 10 -World Health Organisation, 2010). Language difficulties can therefore affect people across the spectrum of 'normal' non-verbal intelligence.

Spoken language disorders are often not detected (e.g. McCool and Stevens, 2011; Spencer, Clegg and Stackhouse, 2012), and individuals with these difficulties may be labelled as quiet, shy, disruptive or generally less able. Incidence of literacy difficulties, social difficulties, unemployment and mental health problems is higher in this population than in the general population (e.g. Bishop and Snowling, 2004; Clegg, Hollis, Mawhood and Rutter, 2005; Conti-Ramsden, Mok, Pickles and Durkin, 2013). Historically language disorders have been difficult to identify early. Early identification is clearly important to ensure that children receive the correct support for their difficulties throughout their development, and so that they learn strategies to overcome barriers that their language processing difficulties create.

1.2. Repetition tasks and language disorders

One form of assessment that has received considerable attention in the research literature is children's immediate repetition of language. This has taken two main forms: the repetition of single 'words', whether real or made-up ("non-words"), and the repetition of sentences. These language repetition tasks have been put forward by researchers as potential clinical markers of SLI (e.g. Conti-Ramsden, Botting and Faragher, 2001). Advantages of this type of assessment include the simplicity of the task and its speed of administration. Furthermore, it can be used with a range of different client groups, successfully identifying those individuals who have a current or history of language processing difficulties (e.g. Bishop, North and Donlan, 1996; Botting and Conti-Ramsden, 2001; Chiat and Roy, 2007; 2008; Conti-Ramsden et al., 2001; Dollaghan and Campbell, 1998; Gray, 2003; Stothard, Snowling, Bishop, Chipchase, and Kaplan, 1998).

While repetition tasks have emerged as potential indicators of language impairment, there is debate about what the tasks measure. No research study that advocates using repetition tasks to identify language difficulties is suggesting that these difficulties might resolve through the practice of repeating non-words and sentences. Instead, it is through understanding the mechanisms that underlie the repetition performance that enables targeted therapy programmes to be devised.

1.3. Focus of the present study

Repetition tasks and their relationship to language development and language disorders are the focus of the present study. There are four parts to the study. The first part investigated the influence of three factors in children's repetition of words and non-words: pronunciation skills, word knowledge and phonological short-term memory (hereafter PSTM) in a heterogeneous sample of children aged 3 years, and again at 4 years. The second part of the study split the sample into those who were known to speech and language therapy (hereafter SLT) and those who were not previously identified as having any speech or language difficulties. It compared performance by the two groups to explore qualitative differences in their performance. The third part of the study investigated the children's performance on a sentence repetition task and it explored relationships between performance on this task with independent measures of grammar (morphology) and PSTM. The fourth part of the study examined relationships between children's repetition of single items (words and non-words) and their repetition of sentences both concurrently and predictively.

In its investigation of young children's repetition accuracy, the present study sought to clarify contributions of existing language knowledge and PSTM to children's performance. This had both theoretical and clinical motivations. Increased understanding of the processes underlying language repetition tasks: i) enables the development of theoretical models that serve to explain children's language processing; ii) enables researchers to devise intervention studies that test these models empirically; and iii) enables therapists to plan intervention that supports children who present with language difficulties in clinic.

1.4. Theoretical models of language and memory

The study employed a number of theoretical cognitive models to conceptualise the mechanisms involved in children's repetition. Being cognitive models, their purpose was to simplify the potentially very complex, multi-factorial influences that contribute to performance on a task, rather than to explain the minutiae of individual differences in performance. The theoretical models used by the study are described in the following account. First, two well-documented models are presented, followed by a proposed novel theoretical model for explaining children's performance on word, non-word, sentence and list recall tasks.

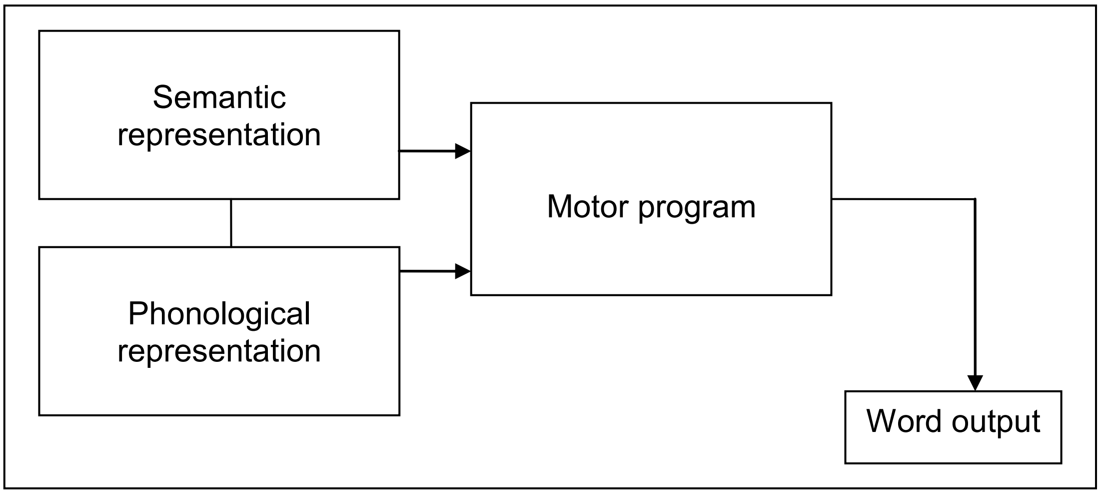
1.4.1. Psycholinguistic model of single word processing (Stackhouse and Wells, 1997)

The first model considered by the study is a theoretical model of word naming and single word (or non-word) processing. The model is based on the psycholinguistic model advanced by Stackhouse and Wells (1997). Their model is built upon findings from a series of single case studies or groups of children presenting with different speech and/or language difficulties. It gives an explanation of possible mechanisms involved during naming and repetition of familiar and unfamiliar words.

1.4.1.1. Naming

The psycholinguistic model proposes that during a naming task, a long-term lexical representation of the word is elicited. Elicitation of the lexical representation is two-pronged, incorporating both semantic and phonological information about the item to be named. As a stored phonological representation is activated, any errors in the word's existing stored phonology will affect accurate production. A motor program for the word is further activated, enabling articulation of the word. This is illustrated in the diagram below (figure 1-1).

Figure 1-1 **Processing during naming** (based on Stackhouse and Wells, 1997)



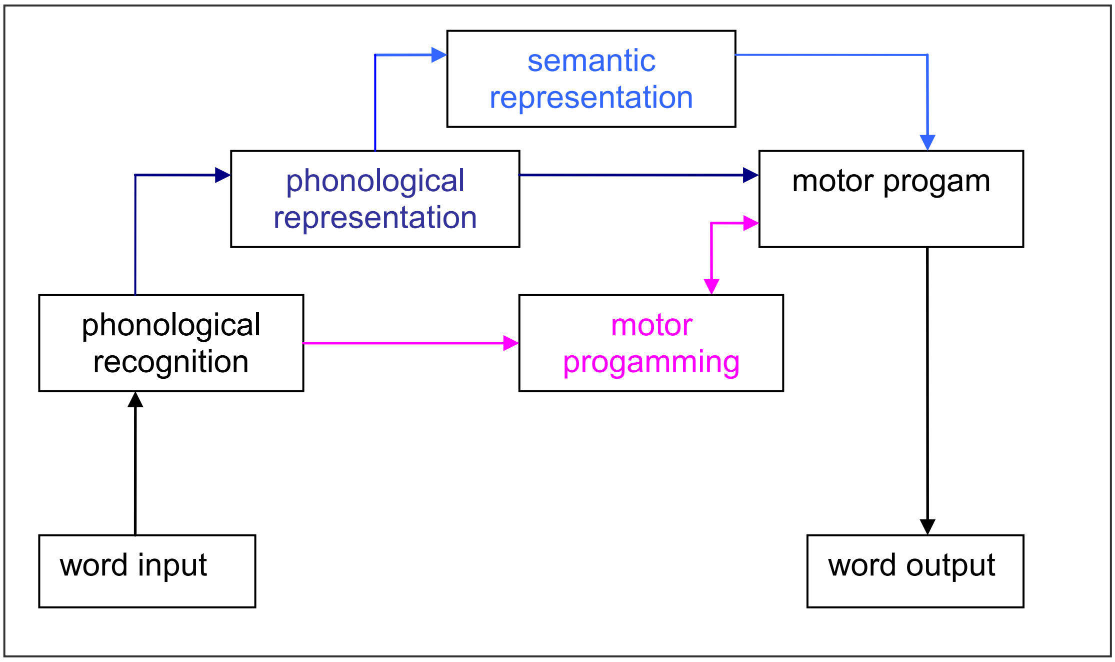
1.4.1.2. Word Repetition

The processing involved during naming may or may not be the same as that applied during word repetition. During repetition of a known lexical item, children may recognise the item and therefore draw on their stored lexical knowledge. It is thought however that repetition can bypass this long-term storage, and dissociations have been found between children's production of words during repetition compared to their spontaneous productions (e.g. Chiat and Hunt, 1993). The dissociations can be explained in terms of 'depths of processing' (Craik and Lockhart, 1972). Having heard the word, children may process it 'deeply', accessing their stored lexical representation (phonological and semantic) of this item. As previously discussed for the naming task, errors would occur in their production if they access a representation that has been stored in long-term memory incorrectly. The incorrectly stored long-term phonological representation might override the temporary representation stored in short-term memory, leading to inaccurate repetition. However it is possible during repetition to process the word only at a ‘shallow’ level, bypassing any long-term representations and therefore relying entirely on accurate perception and temporary storage of the item. This could happen where the child does not have the word in their existing lexicon (as for a novel word or non-word), or where he/she has a ‘fuzzy’ phonological and/or semantic representation of the word. Alternatively it could occur where the child has difficulty recognising the known lexical item as familiar when this is heard out of context (such as during a repetition task). Finally, ‘shallow’ level processing may occur where the child does not perceive the sounds correctly and does not connect the sounds he/she hears with the stored representation.

The depth at which an individual lexical item is processed is known to affect immediate recall for words that are presented in a list. For example, lists of concrete nouns are more easily remembered than abstract nouns that are matched for word length (Walker and Hulme, 1999). Concrete nouns are thought to be processed more deeply because of their richer semantic content, leading to stronger activation. There is also evidence that the lexical status of an item (whether a word or a non-word), and therefore the depth at which it can be processed, influences young children’s repetition of single items. Repetition of words is consistently found to be more accurate than repetition of non-words (e.g. Casalini, Brizzolara, Chilosi, Cipriani, Marcolini, Pecini, Roncoli and Burani, 2007; Chiat and Roy, 2007, 2008; Dispaldro, Benelli, Marcolini and Stella, 2009; Gathercole and Adams, 1993; Roy and Chiat, 2004; Vance, Stackhouse and Wells, 2005).

A diagram showing the two possible processing routes involved during repetition of known words is illustrated in figure 1-2. The pink route illustrates the processing involved in repeating novel words or non-words, while the blue route illustrates the processing involved in the repetition of known words (although known words could also pass through the pink route). It is possible that the child could be familiar with the phonology of a word, without understanding its meaning, and this is discussed further in the account that follows. In the diagram, the lighter blue colour in the diagram indicates that this 'deeper' semantic level of processing may or may not be activated.

Figure 1-2 Processing during repetition of known or unknown lexical items (based on Stackhouse and Wells, 1997)



Levels of 'known-ness'

So far it has been discussed that during repetition, known words likely benefit from semantic and phonological knowledge, while unknown words and non-words rely more on some sort of 'bottom-up' processing, encompassing accurate perception, temporary storage and articulatory planning and execution. However there are different levels at which a child might 'know' a word. One way of tapping word knowledge is to use a naming task. If the child is able to label a picture of the item, then it can be concluded that the child knows the word. One widely used test of vocabulary, however, assesses recognition rather than naming of words. The British Picture Vocabulary Scale, currently in its third edition (Dunn, Dunn, Styles and Sewell, 2009), requires the participant to point to the correct picture from a choice of four when the label is given. A test of receptive vocabulary obviously taps a different level of familiarity with an item compared to a naming task. Conversely, a child might be familiar with the phonology of a word due to hearing it often, but he/she might not yet have an understanding of the meaning. In this case, the familiar phonology might aid repetition, but the word is not considered to be 'known' as tested by either a naming or a recognition task.

1.4.1.3. Non-word repetition

In addition to the different levels of processing available during repetition of familiar lexical material, there is also considerable evidence in support of non-words benefiting from stored long-term lexical knowledge. For example, studies have compared different types of non-words: those that are more- and less word-like (e.g. Archibald and Gathercole, 2006). These non-words varied in terms of a number of different lexical properties that are explored further in section 1.5. Items that were more word-like were repeated more accurately than those which were less word-like. The model in figure 1-2 cannot easily account for this finding, but implicit, probabilistic learning can explain why this is the case. Non-words that more closely resemble real words have the benefit of fitting existing phonological templates that a child's language-learning system has created. These templates or 'word-recipes' are described by Velleman and Vihman (2002) as resulting from an interaction between the language heard by the child in their early development and the child's own articulatory experiences. The authors suggest that the child develops a perceptuo-motor filter that drives the formation of word templates. These templates influence the child's perception and production of subsequent words. The templates described above are regarded here as implicit word knowledge and use of this knowledge during repetition is considered as 'sub-lexical processing'.

1.4.1.4. Summary

In summary, previous studies have shown that the degree of familiarity with the item-to-be-remembered (word or non-word) aids repetition. Semantic and phonological properties of the item influence performance. A spectrum of familiarity is presented in table 1-1 together with cognitive/linguistic resources that are proposed to be available to aid repetition performance.

Table 1-1 Type of item (word or non-word) and proposed available resources during repetition

	Item type					
Available resources to aid repetition accuracy	Known words *(items that children can name)*	Less known words *(items that children are unable to name, but can recognise)*	Familiar words *(words that children recognise, but do not understand)*	Word-like non-words	Unfamiliar words	Unword-like non-words
Semantic representations (word specific)	✓	(✓)				
Phonological representations (word specific)	✓	(✓)	✓			
Word-general phonological templates	(✓)	(✓)	(✓)	✓	(✓)	
Temporary activation in short-term memory	✓	✓	✓	✓	✓	✓

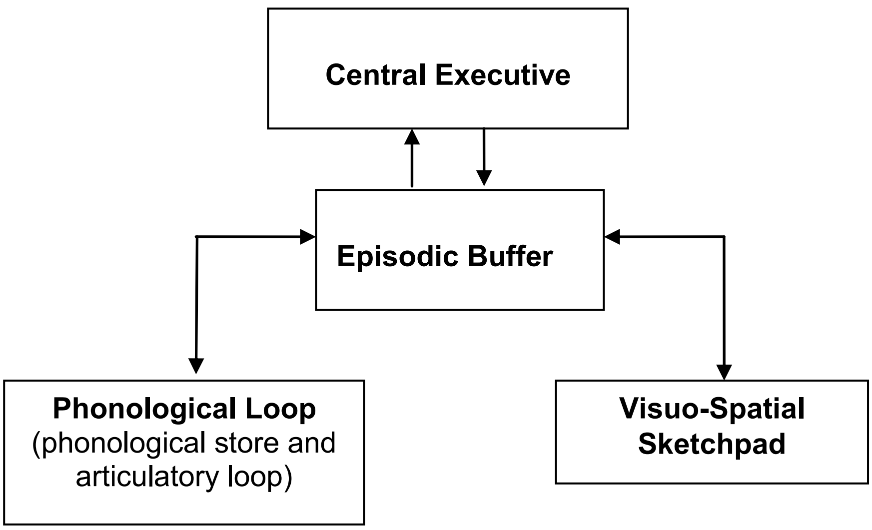
So far discussion has focussed on lexical and sub-lexical processing that might occur during a repetition task. Stackhouse and Wells's (1997) psycholinguistic model serves to clarify this processing to some extent. It does not however explain some of the observed findings. First it does not explain why non-words that are more word-like are easier to repeat. Second, it is well documented that increasing the amount of information to be recalled impairs performance: when a non-word is long, containing several syllables (e.g. Gathercole and Baddeley, 1989), or when there are multiple items to remember like in the recall of word lists (e.g. Hulme, Thomson, Muir and Lawrence, 1984; Miller, 1956). To account for these latter findings, the need for a model of short-term memory is indicated: verbal information needs to be stored for a short-period of time before being articulated. The following account focuses on one such very influential model of short-term memory: Baddeley's model of working memory.

1.4.2. Working Memory

Baddeley and Hitch (1974) proposed a tripartite system of Working Memory (WM) that included two 'slave systems': the Visuo-Spatial Sketchpad and the Phonological Loop; and a governing system: the Central Executive. The model has been extensively researched, developed and documented, and in later versions Baddeley (2000) added a fourth component, the 'episodic buffer'.

The WM system is illustrated in the diagram below (Figure 1-3) and a brief account of each sub-system follows.

Figure 1-3 Simplified model of Working Memory, based on Baddeley et al., 2009



The **Central Executive** is a system responsible for a range of functions. It is thought to control the direction of attention to relevant sensory information and to inhibit response to distracting perceptual input. It is also considered to be responsible for coordinating the two slave systems. Additionally it provides supplementary resources when the slave systems become overloaded, e.g. when they are required to perform multiple tasks at once.

Baddeley, Hitch and Allen (2009) describe the **episodic buffer** as an attention-dependent “limited capacity store in which information from the short-term stores and long-term memory can be integrated into episodic chunks” (p439). The system has the role of integrating temporary representations from other cognitive systems, including perception, as well as components of WM (Baddeley, 2000). It binds the information from these different components together. Important to the present study is their claim that it accounts for the advantage of sentence recall over word lists (see chapter 6). However, since the introduction of this new component fourteen years ago, its exact structure and the methods of testing the system have remained vague.

The **visuo-spatial sketchpad** manages temporary storage of visual and spatial information and as it is unlikely to be involved in the processing of verbal linguistic information, this component of the WM system will not be discussed further.

Of most relevance to the present study is the **phonological loop** component of the WM system, which is often referred to in the literature as Phonological Short-Term Memory (PSTM). This part of the WM system is thought to store incoming phonological information for brief periods to enable repetition or encoding for long-term storage. The phonological loop comprises a limited capacity store (*phonological store*) and a rehearsal mechanism (*the articulatory loop*) which prevents information in the store from decaying through its continual sub-vocal repetition. The phonological loop is a limited capacity system and is typically measured by requiring participants to repeat a series of randomly selected digits or single syllable words in sequence. The number of items stored in the correct sequence (digit or word span) increases with age, generally reaching a maximum of 7 +/- 2 items (Miller, 1956), and this grows further where 'chunking' occurs (see summary in figure 1-4, page 33).

Evidence in support of the separate phonological store and articulatory loop comes from studies investigating disruption to recall. The main pieces of evidence in support of the phonological store are the phonological similarity effect (Conrad and Hull, 1964) and the irrelevant speech effect (Salamé and Baddeley, 1982). The main pieces of evidence in support of the articulatory loop are the word length effect (Baddeley, Thomson and Buchanan, 1975) and the articulatory suppression effect (described in Gathercole and Baddeley, 1995). Non-word repetition is thought to tap the phonological loop (see chapters 4 and 5) and some studies suggest that sentence repetition also taps this system (chapter 6). It was important for the present study to consider the properties of the store and the loop in the design of the measure of PSTM adopted in the study. Further discussion about this is provided in chapter 2, section 1.

1.4.3. Alternatives to the Working Memory Model

While Baddeley's Working Memory model is widely influential, it has also been subject to much criticism. Among other theoretical accounts, Cowan (2008), for example, suggests that memory is not split into two separate short-term and long-term stores. Instead, Cowan (2008) argues that short-term memory represents temporary activations of representations in long-term memory, mediated by the focus of attention. The representations are activated to different degrees depending on the recency of the activation and the frequency with which they are activated. More frequent and more recent activations become more available or more sensitive to reactivation. The focus of attention is limited, explaining why several pieces of information cannot be held at once.

Chunking

PSTM is measured by requiring participants to repeat a series of digits or semantically unrelated single syllable words in sequence. The number of items stored in the correct sequence (digit or word span) is considered to indicate the capacity of the system. However, we may use strategies to facilitate our recall. One strategy adopted is ‘chunking’. McLean and Gregg (1967) provide a description of ‘chunks’ in verbal recall as “groups of items recited together quickly”. They outline three ways in which chunks can be created. First, some of the information to be remembered may already form a group with which the participant is familiar, forming a natural ‘chunk’. Second, “external punctuation of the stimuli may serve to create groupings of the individual elements”. An example of this is ‘prosodic chunking’. Third, the participant may “monitor his own performance and impose structure by selective attention, rehearsal, or other means”, e.g. conscious reorganisation of the material into meaningful chunks. These methods of chunking will be illustrated below with reference to lists of digits and phonemes in words and non-words.

During presentation of the following number sequence:

3 7 1 8 5 6 2 (three, seven, one, eight, five, six, two)

It is possible that part of the digit sequence has some personal relevance to a participant. For example, he/she may recognise that a series of the digits within the sequence is identical to his/her bank pin code or the last three digits of his/her phone number. Because part of the sequence is familiar, this would form a chunk, and would therefore assist recall.

An alternative method of ‘chunking’ involves a reorganisation of the information prosodically. For example, digits presented one at a time at regular intervals without prosody are more difficult to retain than digits that are grouped. The example above might instead be recalled as:

3-7-1 8-5-6-2 (three-seven-one, eight-five-six-two).

Prosodic organisation like this is common when, for example, dictating a telephone number.
We can further facilitate recall by re-organising the digits into larger chunks. So, the above sequence of digits can be linguistically recoded into larger units, or ‘chunked’ as:

37 18 56 2 (thirty-seven, eighteen, fifty-six, two)

These ‘chunks’ reduce the load on the PSTM, as seven pieces of information are reduced to four. However, if this last method is used, the system must then reorganise the 'chunked’ information back to its original form for recall.

While chunking might occur internally by the individual recalling the sequence (as in the examples given above), the information might also be presented in pre-existing chunks. For example, the sequence of numbers above might instead be presented as:

1-2-3 8-7-6-5

Here the recall of information presumably draws only minimally, if at all, on phonological short-term memory (assuming basic counting skills by the person recalling). Instead the individual can use his/her long-term knowledge to recall the two familiar sequences of numbers.

Further examples of the use of pre-existing 'chunks' of linguistic information during recall might be assumed during the recall of phonemes and words. The following sequence of phonemes is difficult to remember in sequence:

/p/ /k/ /l/ /ɪ/ /ə/ /æ/ /t/ /ə/

However, if the sequence is re-ordered and presented as a meaningful known word, there is no difficulty recalling the sounds:

/kætəpɪlə/

Accurate recall of this familiar sequence of phonemes is likely facilitated by existing knowledge of the word and its phonology.

However, even when the phonemes are rearranged to form a novel 'word', language unimpaired adults have sufficient knowledge of words and language for 'chunking' to occur and for the load on phonological short-term memory to be reduced, leading to simple recall of the sequence of sounds:

/tɪkəlæpə/

Recall of this unfamiliar sequence of phonemes is facilitated by several factors outlined in section 1.5.2, e.g. similarity to known words; the stress pattern of the word, the likelihood of phonemes occurring together ('phonotactic probability').

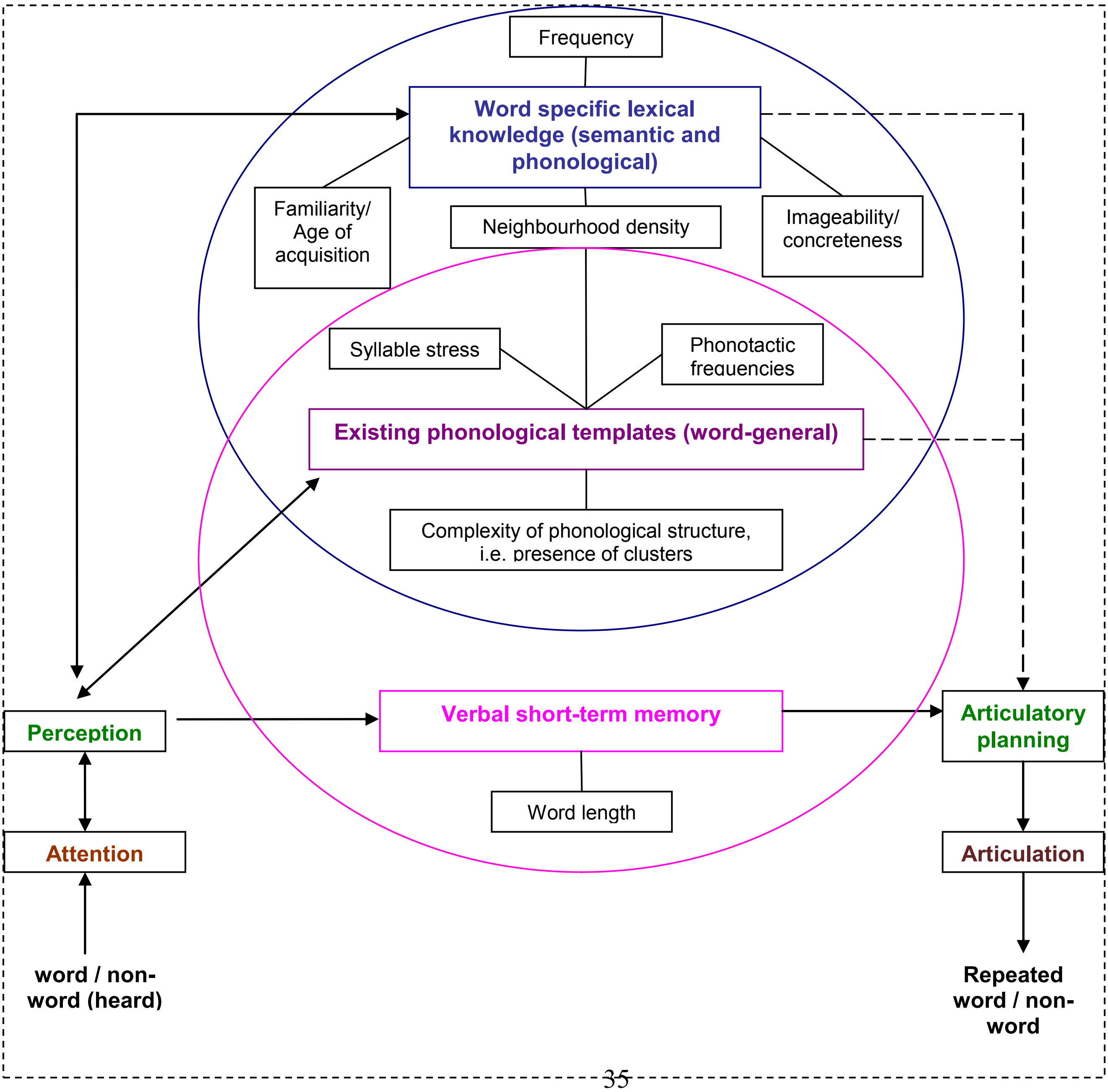
Similarly, in the case of word lists (rather than digits), recall is enhanced by semantic, phonological, prosodic and syntactic relationships (see section 1.6.), which can also be considered as 'chunking' of linguistic information. This explains the advantage for sentence recall over lists of unrelated words (see also chapter 6).

Whether short-term memory represents a separate system or whether it is intrinsically part of long-term memory is not the main focus of the present study. However, what is of interest is how phenomena typically associated with PSTM (e.g. capacity constraints, phonological similarity effects), and which are present in single item repetition tasks, interact with phenomena associated with existing language knowledge to influence repetition. The next section explores some of the evidence in support of the influence of these PSTM and linguistic factors and it proposes a model that specifies these additional factors.

1.5. The influence of language and memory on repetition tasks: model of single 'word' repetition

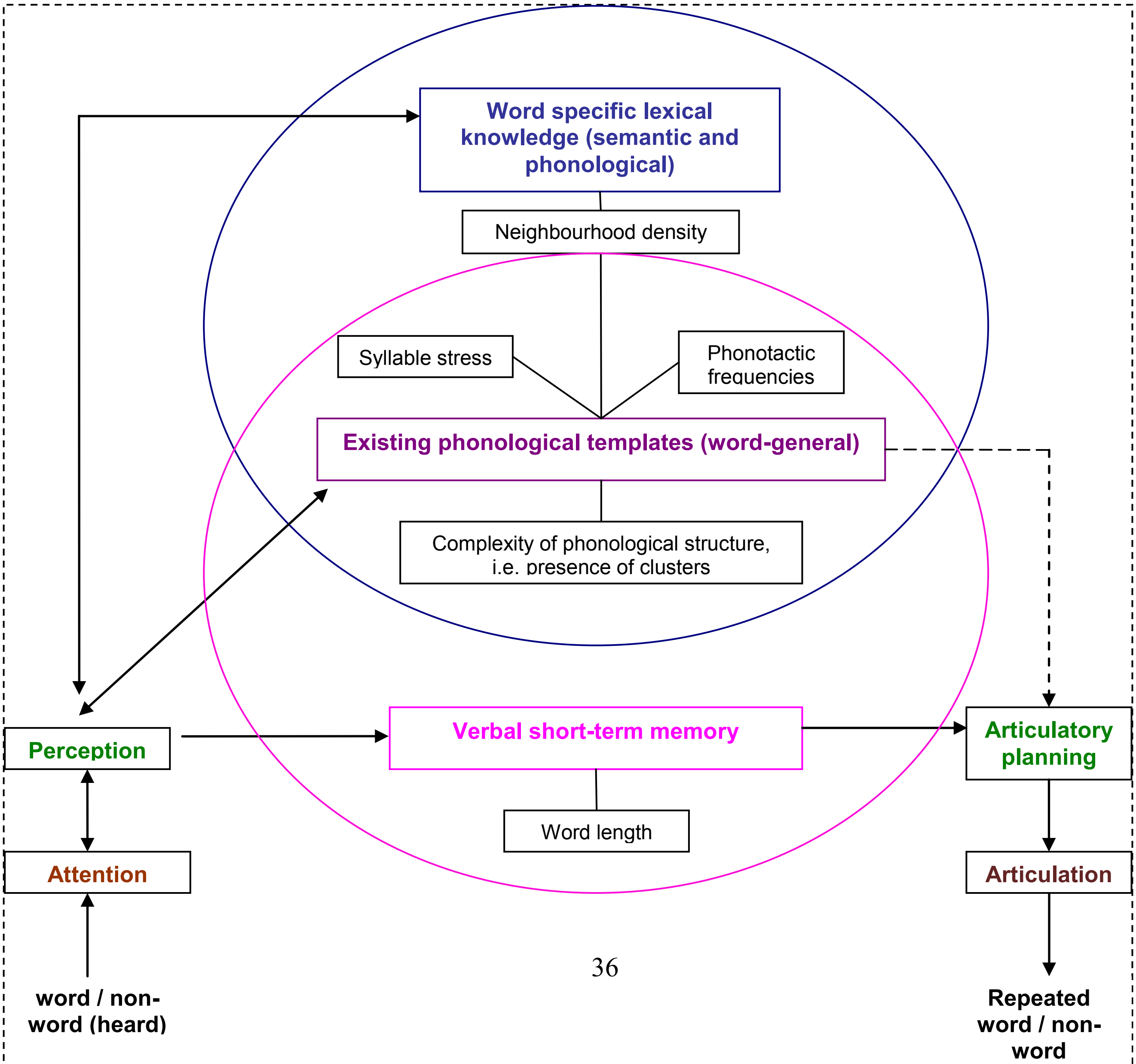
An assumption made by this study is that during repetition of linguistic material, a child's cognitive/linguistic system will employ whatever resources are available to enable the most accurate repetition of the target. The proposed model that follows has been created for this study using information from existing research findings into NWR and word list recall (e.g. Coady and Evans, 2008; Gathercole, Pickering, Hall and Peaker, 2001; Luce and Pisoni, 1998; Metsala and Chisholm, 2010), as well as from theories of 'redintegration', e.g. Hulme, Roodenrys, Schweikert, Brown Martin and Stuart (1997). It also draws on psychological models of implicit and explicit memory that Velleman and Vihman (2002) also describe in their explanations of children's word learning and use. The model presents the interaction between three layers of memory relevant to the language system: i) explicit word knowledge (i.e. information about the meaning of specific words and the specific phonological representation for these words, stored explicitly in long-term memory); ii) implicit word knowledge (i.e. sensitivity to phonological patterns of words that are typical to the language); and iii) capacity to store sequences of new phonological information (PSTM). A model illustrating the system is shown in the diagram (figure 1-5). Information associated with explicit word knowledge is encompassed in the blue region of the diagram; while information relating to verbal short term memory is shown in pink. Implicit word knowledge may be drawn upon to the extent that the phoneme sequence to be recalled (the non-word or word) is familiar or prototypical for the language. It is proposed therefore that this level of processing interacts both with verbal short-term memory and with explicit word knowledge. For this reason it is encompassed by the overlapping blue and pink regions of the diagram and is coloured in purple.

Figure 1-5 Levels of processing for repetition of words and non-words



The model presents a mutually dependent set of systems: activation of a system encompassed within the pink or blue area triggers activation of another system within the same area, thus knowledge stored at the level of phonological templates interacts both with verbal short-term memory and also with existing word knowledge. The model assumes that in focussing attention on a novel sequence of phonemes (a non-word), these phonemes, together with their prosodic patterning are filtered through the perceptual system and stored temporarily, in verbal short-term memory. In order for accurate repetition to occur, the verbal short term memory system accesses existing phonological templates which reduce the load on the memory system by enabling a form of ‘chunking’ to occur (described in the box on page 32-33).

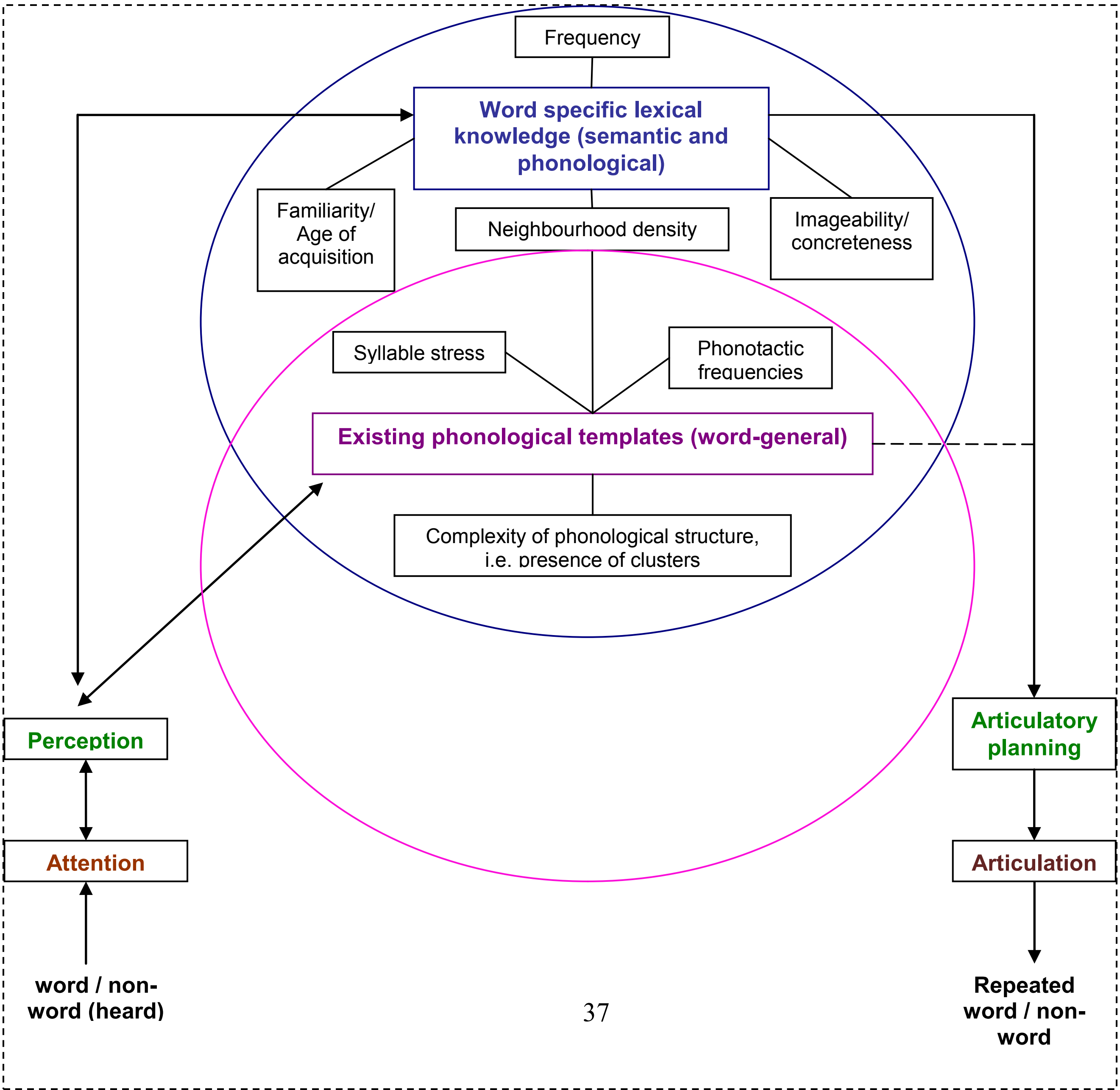
Figure 1-6 Influences during non-word repetition



The parts of the model likely activated during the repetition of non-words are illustrated in figure 1-6.

In the case of a familiar sequence of phonemes (a known word), the same journey through the system might be followed, with additional stronger activation of the word specific lexical templates. It is possible however that verbal short term memory is bypassed (figure 1-7), instead activating directly the stored information about the word (word specific lexical knowledge), This information would then be used in the articulation of the word. This possibility is discussed further in chapter 4.

Figure 1-7 Influences during word repetition



The main layers of the model and rationale for the inclusion of each level of processing are described next.

1.5.1. Word Specific Lexical Knowledge

The **word specific lexical knowledge** relates to the long-term store in the form of semantic and phonological representations for a specific word. There is debate in the literature about which factors affect the semantic and phonological strength of a word and the extent to which they impact on repetition. The model includes the following factors that are associated with specific word knowledge: age of acquisition (e.g. Turner, Henry and Smith, 2000; Turner, Henry, Smith and Brown, 2004), familiarity (Stuart and Hulme, 2000), imageability (Majerus and van der Linden, 2003), level of concreteness (e.g. Walker and Hulme, 1998) and frequency (Coady, Mainela-Arnold and Evans, 2013; Turner et al., 2000, 2004). The strength with which these representations are activated depends on their phonological relationship to other words (‘phonological neighbourhood’) (e.g. Chen and Mirman, 2013). Furthermore, some of these individual lexical factors are known to interact with each other. The interactions have not been included in the proposed model however, for ease of presentation. These influences will be discussed in turn.

Evidence in support of there being word-specific influences from long-term memory during verbal repetition tasks comes mainly from research investigating recall of lists of words (e.g. Hulme et al., 1997; Turner et al, 2000, 2004; Walker and Hulme, 1999), rather than repetition of single items. It should be noted therefore that it is questionable whether it is appropriate to generalise these findings to support the above model, which is primarily a model of single word (or non-word) recall. However the consistent finding that single words are repeated more accurately than non-words in pre-school children (e.g. Chiat and Roy, 2007; Casalini et al., 2007, Dispaldro et al., 2009) provides some evidence for this level of processing being important in word and non-word repetition too.

1.5.1.1. Concreteness and Frequency

Two factors which affect span recall in children and adults, but which have not been explored in children's single word repetition, are levels of concreteness and word frequency. Lists of abstract words result in shorter spans for adults than lists of concrete words (Walker and Hulme, 1999). Frequency also affects adults' span recall (Hulme et al., 1997; Roodenrys, Hulme, Alban, Ellis and Brown, 1994) as well as children's span recall (Coady et al., 2013; Turner et al., 2000, 2004), with high frequency words being more easily recalled than low frequency words. It is not known whether word frequency also affects children's single word repetition.

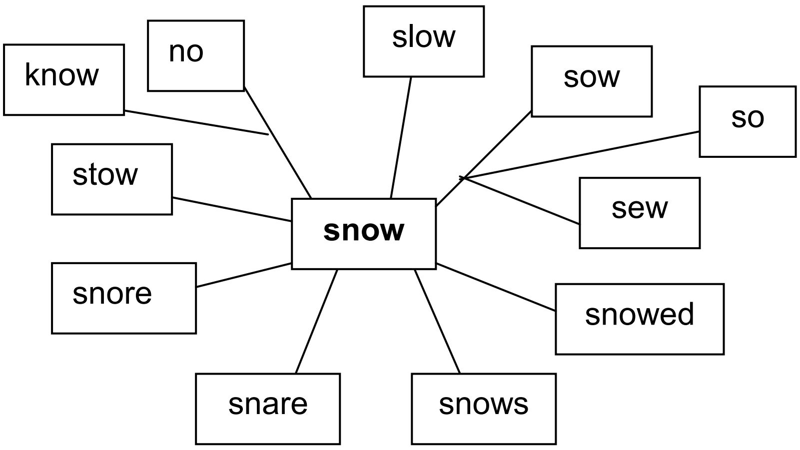
1.5.1.2. Age of acquisition/familiarity

There is evidence that age of acquisition affects children's repetition of sequences of words (their word span) (Turner et al., 2000, 2004). It is unclear however whether estimates of the age at which a word was acquired influence performance on a word repetition task. A study by Dispaldro et al. (2009) reported that Italian children aged 3-4 years did not show an advantage for early acquired words compared to later acquired words. However, age of acquisition data is derived based on adults' estimates at which they acquired the words. Given such large vocabularies, it seems unlikely that adults have an accurate knowledge of when they learnt any given word. Furthermore, as language is an ever-evolving system, it seems likely that age of acquisition for many words differs across generations. The present study was interested in whether children had actually acquired the words, rather than the figures attached to words given as an estimate of when they are acquired. Therefore, in spite of the findings to the contrary (Dispaldro et al., 2009), the present study hypothesised that children will repeat words that they know (and have therefore ‘acquired’) more accurately than words that they have not yet acquired. This is discussed further in chapters 4 and 5.

1.5.1.3. Neighbourhood Density

Neighbourhood density is defined as the number of lexical items that are connected to a given word by their phonological similarity (e.g. Metsala and Chisholm, 2010). The most commonly used method of calculating neighbourhood density is to determine the number of words that can be made by changing, adding or omitting one phoneme from the word (e.g. Goldinger, Luce, and Pisoni, 1989; Luce and Pisoni, 1998). Using this method of calculation, some of the words in the neighbourhood of the item ‘snow’ are given in the diagram below (figure 1-8).

Figure 1-8 Examples of the lexical neighbourhood for the word ‘snow’



During spoken word recognition, known words with sparse neighbourhoods are recognised more quickly and easily than known words with dense neighbourhoods (Luce and Pisoni, 1998; Vitevitch and Luce, 1999). This can be explained in terms of levels of activation. If lexical items are stored in long-term memory in networks, with phonologically similar items closely linked to the target item, when a word is heard, all those items that are linked phonologically to the target word will be activated. The item with the strongest similarity to the heard item will reach its threshold of activation quickest, and will be selected. Where there are several phonological neighbours, competition increases. Potential problems arise where stored representations are poorly specified or ‘fuzzy’. Here, it may take longer for an item to reach its threshold of activation, or several stored representations may become activated and compete for selection, leading to inaccuracies in production. Alternatively, if the word is perceived incorrectly, this will lead to problems with retrieval of the stored lexical item.

There is evidence for the influence of neighbourhood density in the repetition of word lists. Children aged 8-9 years performed better where words came from a dense neighbourhood (Thomson, Richardson and Goswami, 2005) compared to a sparse neighbourhood. This was the cases for children whose language was developing typically and also for children with dyslexia.

During non-word repetition neighbourhood density also has an influence (Metsala and Chisholm, 2010; Thomson et al., 2005). Thomson et al. (2005) found that children aged 8-9 years repeated non-words more accurately when these were from dense neighbourhoods. In younger children, Metsala and Chisholm (2010) demonstrated that 3 and 4 syllable non-words with dense (real word) neighbourhoods were repeated more accurately than those from sparse neighbourhoods in children aged 3 - 7 years. This was not however the case for 2-syllable words. Their findings might however be confounded by the children’s limited vocabularies. It is important to consider that due to the age of their participants and also those in the present study, it is possible that the children would have much sparser neighbourhood networks than an adult would and this is discussed in the study’s methods for selecting the stimuli (chapter 2, section 1). It might be that the effect of neighbourhood density is less pronounced for children’s repetition (particularly children with language difficulties) because their overall vocabularies are small.

As non-word recall is influenced by neighbourhood density, existing word knowledge must be drawn upon to some extent during non-word repetition. This highlights the interacting levels of the model and neighbourhood density acts as a linking factor between the levels of lexical processing (phonological templates and word specific lexical knowledge) in the model. For this reason this factor has been illustrated as lying between word-specific lexical knowledge and phonological templates in the model (see figures 1-5, 1-6, 1-7).

1.5.2. Existing Phonological Templates (sub-lexical processing)

As described in section 1.4.1.3, Vihman (e.g. Velleman and Vihman, 2002) has proposed that, in the early stages of word-learning, children acquire ‘phonological templates’ based on experience of the phonological patterns of the language (both prosodic and phonemic). The templates aid subsequent word learning through focussing attention to incoming phonological patterns that conform to the templates. Some researchers refer to this level of word learning as sub-lexical processing or implicit language knowledge. The templates are not static, but instead become increasingly specified through exposure to language and growth in vocabulary. This can however lead to incorrect over-generalisation. As language contains some non-conformist words (i.e. those with atypical phoneme sequences or stress patterns), there would need to be sufficient flexibility within the template system to support these atypical rebels.

1.5.2.1. Syllable Stress

Word/syllable stress has been shown to affect children’s repetition (e.g. Gerken, 1994; 1996; Carter and Gerken, 2003; McGregor and Leonard, 1994). In English, words typically follow a strong-weak (trochaic) foot structure, where the first syllable is stressed more than the second, for example, “**a**pple”, “**ta**ble”. A less common structure in English is the iambic structure, where the second syllable takes the greater stress, for example, “gi**raffe**”, “gui**tar**”. Gerken (1994) noted children’s tendency to prefer trochaic, rather than iambic structured words and sentences in their expressive language. She reported that when repeating words and phrases, children in the early stages of learning language tend to omit the initial syllable when this is weakly stressed (Gerken, 1994, 1996). Children with language difficulties also show this tendency. For example, a study by Sahlen, Reuterskioeld-Wagner, Nettelbladt, and Radeborg (1999) revealed that Swedish children with SLI omitted unstressed syllables from words and non-words more often when the stress occurred in an iambic foot structure compared to a trochaic structure. In younger English-speaking children, Chiat and Roy (2007) found that 2 ½ - 3 ½ year old children omitted more initial syllables that were weakly-stressed. Furthermore they found that children who had been referred to speech and language therapy omitted more first syllables that were weakly stressed than children whose language was developing typically.

A further study (Archibald and Gathercole, 2007) showed that non-word repetition is facilitated by prosody when compared with maintaining an even stress on syllables in a non-word. Until recently, non-word repetition tests have not taken word stress into consideration in their design, which has confounded interpretation of the findings. Drawing on findings from the studies that highlight word stress as an important factor in word (and non-word) repetition by children with language disorders (e.g. Chiat and Roy, 2007), it was considered important that the present study’s model included word/syllable stress as a factor in the design for the stimuli.

1.5.2.2. Phoneme composition

Another factor affecting children's repetition of words and non-words is the phonological composition of the words, which might be affected, for example, by the children's speech sound acquisition. For example where words contain later developing phonemes, e.g. 'r', articulation difficulties will prevent fully accurate repetition. Cluster reduction is another speech process that persists in children's speech late in development. Where WR and NWR tasks do not adequately control for the range of speech errors a child might make, this might lead the child to score more poorly on a repetition task due to their pronunciation difficulties (see Chapter 4 and see also Bishop et al. 1996). However, interestingly a study published by Marshall and van der Lely (2009) found that children with language disorders have more difficulty repeating word-medial clusters than their typically developing peers, and this difficulty was not explained by pronunciation errors. This latter finding suggests that the cluster increases difficulty for a language impaired child for reasons other than its increased articulatory complexity, perhaps due to the increase in phonological information to recall.

1.5.2.3. Phonotactic Probability

Phonotactic probability refers to the frequency with which individual phonemes appear in particular positions within words and the frequency with which sequences of phonemes appear in combination in these positions. In non-word span recall, Gathercole et al. (1999) found that 7 - 8 year-old children's recall of non-words with high phonotactic probability was more accurate than with non-words with low phonotactic probability. However a study by Roodenrys and Hinton (2002) showed that the reported phonotactic probability effects were confounded by the effects of neighbourhood density (described previously). They designed non-word stimuli that could be manipulated either for phonotactic frequency or for neighbourhood and they found that the latter, but not the former variable, affected repetition accuracy in adults.

In single item non-word repetition however, studies have shown that non-words with higher phonotactic probability values are more easily repeated (e.g. Munson, Edwards and Beckman, 2005). This can be explained by enhanced perception of phonemes that occur commonly in words, due to a form of 'priming' by the existing phonological templates.

1.5.3. Verbal Short-Term Memory

The **verbal short-term memory** system, as discussed previously, is a limited capacity system that has the role of storing novel sequences of phonemes or syllables for a limited time. The storage time is estimated to be approximately 2 seconds (Baddeley, Thomson and Buchanan, 1975), though there is evidence that for children over 5 years (see Henry, 1991), the memory trace can be refreshed through a process of sub-vocal rehearsal (Baddeley et al., 1975). Due to the restricted capacity of the system, when there is a lot of incoming phonological material or when that material is phonologically complex, the system can become saturated. During repetition of unfamiliar sequences of sounds (non-words), this 'overloading' would lead to errors.

1.5.3.1. Word Length

Word length could refer to the number of phonemes, or to the number of syllables in the word. However, for ease of explanation, the following account will consider 'word length' as being reflected by the number of syllables in the word. Consideration about number of phonemes has already been given in relation to clusters of phonemes in the description of the influence of phoneme composition (section 1.5.2.2.).

In multi-word repetition, i.e. span tasks, word length effects are documented widely in the memory literature. A consistent finding is that short-term memory span for longer words is smaller than for shorter words (Baddeley et al., 1975). As discussed previously, this has been interpreted as reflecting the limited capacity of PSTM, and the fact that longer words take longer to rehearse by the articulatory loop.

It is also widely documented in the literature that repetition is more accurate for shorter non-words than longer non-words reflecting application of PSTM (e.g. Gathercole and Baddeley, 1989). In support of this interpretation is the high correlation found between digit span and non-word repetition ability (Gathercole, 2006). A PSTM account explains this in terms of longer words demanding more resources from the temporary storage system. The system, being limited capacity, reaches saturation point and cannot hold the whole sequence of consonants long enough for accurate recall.

An alternative interpretation for the word length effect highlights the role of phonological perception, segmentation or articulation and also of typical versus atypical phonological representations (Snowling, Chiat and Hulme, 1991; Gallon, Harris and van der Lely, 2007). This position is supported by the finding that children's real word repetition also shows a word length effect. For example, Chiat and Roy (2007) found that children referred to speech and language therapy (but not typically developing children) repeated longer real words less accurately than shorter words, particularly where the real words had an atypical (iambic) stress pattern. Repetition of known words should not require application of PSTM, since stored representations from long-term memory are likely to be accessed for these lexical items (see chapters 4 and 5 for an alternative explanation).

Drawing both of the above interpretations together, while findings so far have suggested that the strength of the word length effect is moderated by position of the stressed syllable, the consistent demonstration that children with language disorders repeat longer unfamiliar sequences of phonemes less accurately than their typically developing peers provides the rationale for including word length as a factor in the model.

1.5.4. Perception and attention

The model above incorporates **perception** of the phonemes. Perception is influenced by both top-down (or ‘conceptually-driven’) and bottom-up (or ‘data-driven’) processing. The top-down processing takes the form of existing knowledge, and in the case of hearing non-words or novel words, it is the implicit/sub-lexical knowledge. As discussed previously in relation to the paper by Velleman and Vihman (1980), this ‘knowledge’ affects the child’s expectation of the sequence of sounds he ‘hears’ and the stress patterns of the words. In turn, the expectation guides interpretation of the series of sounds and leads the child to ignore information that is inconsistent with his expected perceptions. This happens throughout our perceptual experiences (Warren, 1970).

Perception is also mediated by bottom-up processing, i.e. **attention** (Cherry, 1953). If we are able to attend to details within the sequence of sounds and to discrepancies between the actual data and our preconceived expectations then repetition accuracy should be enhanced (assuming satisfactory subsequent temporary storage). The perceptual system strives for a coherent whole. Where there are contradictions between the incoming data and our existing knowledge (i.e. between top-down and bottom-up processing), the perceptual system has the role of reconciling these differences (Rock, 1983).

1.5.5. Articulatory planning and Articulation

Articulatory planning refers in this model to the planning of movements needed to articulate the sounds in the word. Articulation is the execution of the movements in order to produce the word (or non-word). These components, which are likely to be influenced by the length and complexity of the word or non-word (Maner, Smith and Grayson, 2000) were not the focus of the present study, but might have played a role in children's repetition performance. However, the possible impact of these factors were minimised by matching words and non-words phonologically and for length, and by scoring for proportion of phonemes correct (see chapter 2).

1.6. Sentence repetition

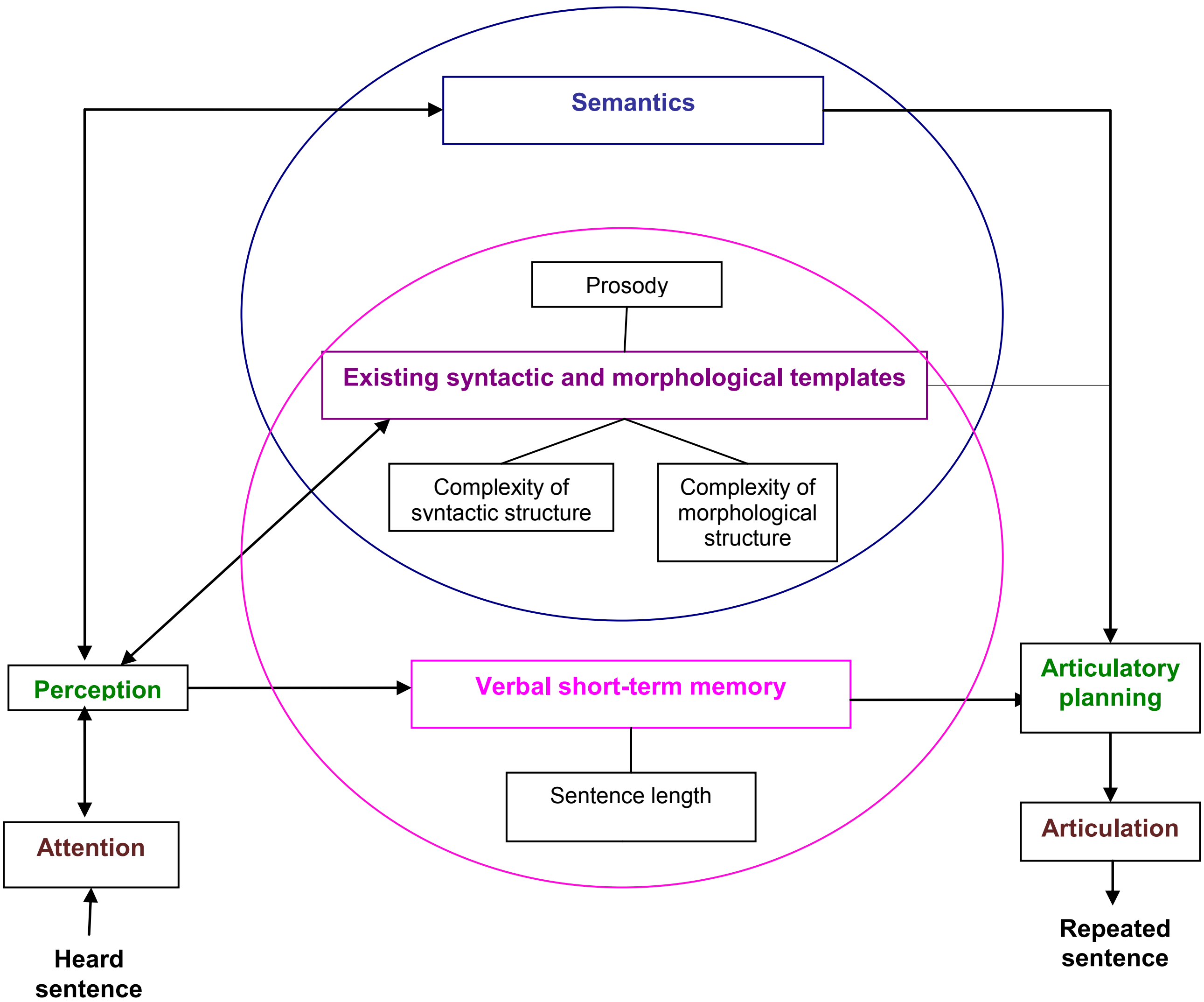
So far a model has been proposed (figure 1-5) to account for single-item repetition (NWR and WR). It does not however account for the pattern of performance observed in the repetition of sentences.

Like non-word repetition, sentence repetition (hereafter SR) has been linked to PSTM (e.g. Alloway and Gathercole, 2005; Willis and Gathercole, 2001). Further discussion about the details of these studies and links with PSTM are saved for chapters 6 and 7. However what is evident is that, while SR and span tasks (that are traditionally used as measures of PSTM) are both immediate recall tasks, the information contained in a sentence enjoys much more support from long-term memory. Repetition of words in sentences far exceeds repetition of lists of familiar single syllable words. These long-term memory influences have been reported in a study by Polišenská (2011) that systematically controlled for different long-term memory influences. She found that there was a large advantage for sentences that respected syntactic rules and that maintained semantic coherence. A significant, but a less powerful influence found in her study was the preservation of typical prosodic patterns. This study is discussed further in chapter 6.

A limitation that the author acknowledges in the study, in terms of understanding the contribution of PSTM to sentence repetition, was that it did not make any direct comparison between repetition of the experimental stimuli and more traditional methods of measuring PSTM (e.g. digit span or unrelated word span). However, the study provides useful evidence in support of the long-term influences described in the model that follows (figure 1-9).

Figure 1-9 is an adapted version of the first model, to account for performance in SR tasks. Like the first model (figure 1-5), the influences of verbal short-term memory in this model are encapsulated in the pink region of the diagram. The region of the diagram that is coloured in blue manages semantics. Here, this might include the meaning of individual words as well as the overall meaning of the sentence. The model assumes that this level of processing is explicit knowledge, i.e. that the individual repeating the sentence is aware of the meaning. As was the case for the model outlined in figure 1-5, the model of sentence processing (figure 1-9) also includes a level of implicit knowledge or processing. Grammatical knowledge (syntax and morphology) is included here, together with the sentence's prosody.

Figure 1-9 Levels of processing for repetition of sentences



The model proposes that during a sentence repetition task, attention is focussed on the incoming sequence of words. The words are filtered through the perceptual system, which is influenced by long-term knowledge about the individual word meanings and the overall meaning of the sentence, which itself is influenced by syntactic and prosodic knowledge.

Further discussion about the evidence in support of the components included in the model is reserved for chapters 6 and 7.

1.7. Aims and structure of the study

The present study aimed to investigate specific components of the models introduced here, while attempting to control for other factors. It thereby aimed to contribute to the knowledge base about factors influencing repetition tasks. Specifically it sought to clarify the shared underlying skills involved in the different repetition tasks and the contributions of long-term knowledge and PSTM on repetition accuracy. The specific questions, rationale and hypotheses for each part of the larger study are given in the individual relevant chapters (chapters 4, 5, 6 and 7).

The study used a longitudinal design, assessing children at two time-points. At the first time-point (T1) children were aged 3 – 3 ½ years and at the second time point (T2) they were aged 4 – 4 ½ years. Further information about the design for the study follows in the next chapter, and further information about the participants is found in chapter 3. The four parts of the study (chapters 4 - 7) emerged from data collected from the same population sample over two time points. These have been written as stand-alone studies, each incorporating their distinct aims, rationales, methods, results and conclusions. Taken together, the separate parts of the study collectively addressed a wider research question about the skills involved in language repetition tasks.

Chapter 8 brings the findings from the distinct parts of the study together and clarifies which skills have been indicated as important in children's repetition of language. It considers the models offered in this chapter (chapter 1) and discusses the extent to which the findings support the models. It suggests openings for future research that might clarify the underlying skills further.

Chapter 2

Methods: selecting the stimuli; scoring the tasks and procedure for collecting the data

2.1. Chapter overview

This chapter discusses the study's methodology. First it discusses briefly the tasks that were included to provide a context for the following sections. Then it discusses the methods for designing the stimuli used in the study and the methods for scoring the children's responses. Finally it reports the procedures adopted when collecting the data.

2.2. Methods for selecting the stimuli

The main tasks under investigation by the study were WR, NWR and SR tasks. In addition a naming task and a recognition task were relevant to the main research questions, as use of the same word stimuli would indicated the child's knowledge of the words they were to repeat. A further task, a word recall (or span) task was relevant to the study's questions, providing a measure of PSTM. These were all novel tasks, so that they could be carefully manipulated for variables outlined in chapter 1 and which will be discussed further in the following account. Further tasks used in the study were drawn from existing standardised assessments. The next section discusses the methods for designing the novel tasks used in the study's tasks.

2.2.1. Existing non-word (and word) repetition tasks

During the consideration of the design for the tasks for the study, existing tests of WR and NWR were explored. Existing tests are presented in table 2-1. The table shows tests available to measure NWR ability. It shows similarities and differences between these tests. Information is given about the target age range of the tests and the number of stimuli. As discussed in the introduction (section 1.5.), the length of the stimulus (number of syllables), the structure of the stimulus and the stress pattern are all known to affect children's repetition accuracy. Therefore, information is given about these dimensions.

Table 2-1 Published NWR tests that are used clinically

Authors	Test	Age range	Number of non-words	Syllable range	Phonemic composition and syllabic structure	Stress
Gathercole, Willis, Baddeley and Emslie (1994)	Children's Non-word Repetition Test (CNREP)	4-8 years	40	2 – 5	Half the items contain consonant clusters in at least one position	A mixture of typical and atypical stress patterns
Dollaghan and Campbell (1998)	Non-word Repetition Test (NRT)	5 ¾ -12yrs	16	1 – 4	-only early developing phonemes (11 consonants, 9 vowels) -no consonant clusters	No unstressed vowels (unlike English stress pattern)
Chiat and Roy (2004)	PreSchool Repetition Test (PSREP)	2yrs-6 years	18 (and 18 real words)	1 – 3	A mixture of single consonants and consonant clusters (based on real words)	A mixture of typical and atypical stress patterns (based on English words)
van der Lely, Gardner, Froud and McClelland (2007)	The Grammar and Phonology Screening Test (GAPS)	3 ½ -6 ½ yrs	10	1 -3	Consonant clusters included	A mixture of typical and atypical stress patterns

As indicated in table 2-1, some of the existing tests contain a variety of prosodic structures and some contain earlier developing or later developing phonemes. The present study could not use the existing NWR tests as it sought to carefully match non-words and words on several dimensions. Furthermore, the present study required words in the repetition task to also be imageable so that they could be used in a naming task. The children's responses in the naming task would enable the children's knowledge of the word to be established (see 2.2.2).

2.2.2. Word stimuli for the repetition, naming and picture recognition tasks

Novel experimental stimuli for the word and non-word tasks were required for the study so that the same word stimuli could be used across WR, naming and recognition tasks, and so that non-words could be matched to the words phonologically. As previously mentioned (2.2.1), a key requirement was that the word stimuli were imageable. The following additional factors were considered in the task design and stimulus selection: word length, word stress, phoneme composition, familiarity, neighbourhood density, phonotactic frequencies. These factors were considered as a result of previous studies' findings that they influence performance on repetition tasks. Evidence in support of the influence of the factors has been discussed in the introduction (Chapter 1). The discussion below focuses on how the factors were considered in the design of the stimuli.

2.2.2.1. Imageability, length, stress and phoneme composition

First a list of imageable words was generated that had a length of between 1 and 4 syllables. It was important that 3- and 4- syllable words would be included in the stimuli, as previous studies (e.g. Gathercole and Baddeley, 1990) had shown non-words of this length to be the most discriminating between groups of children with language difficulties and those without language difficulties. Consideration was first given to two, three and four syllable words. Not only did these words need to be familiar to 3- and 4- year old children, but they were also required to reflect different word stress patterns. The aim was to include an equal number of differently stressed items. For two-syllable items, the main stress could fall either on the first (typical) or second (atypical) syllable. For three- and four- syllable items, the main stress could fall on the first (typical), second (atypical) or third (atypical) syllable. As previously discussed (chapter 1), a balance of typical and atypical stress patterns was required because Chiat and Roy (2007) had found that the stress pattern of their stimuli influenced the likelihood of accurate repetition, particularly for the clinical group of children.

A further factor to consider was the inclusion or non-inclusion of later-developing phonemes and consonant clusters. These can be problematic in the scoring of repetition tasks where the tasks are used to detect difficulties in the repetition by young children or children with phonological delay or disorder, as it can be unclear whether errors in responses reflect phonological or repetition difficulties. However, a study by Marshall and van der Lely (2009) found that older children with SLI showed particular difficulty repeating word-medial clusters of consonants. Indeed this difficulty might differentiate children with language difficulties from their peers. In addition, the presence of clusters might reduce the presence of ceiling effects by all the children. Therefore it was decided that consonant clusters would be included and, where possible, there would be examples of these across the different syllable numbers and stress patterns. This was achieved for 1, 2 and 3 syllable words. It was not possible to find early developing imageable 4 syllable words containing clusters.

2.2.2.2. Familiarity

To measure the extent to which children draw on their long-term stored phonological representations during repetition of known and made-up words, it was necessary first to establish whether they knew the vocabulary and to assess how they produced the words during a naming task. This would allow comparison of naming and repetition productions of the same words. Selection of stimulus words that children would be expected to be able to name at age 3 -3 ½ years was therefore required.

One challenge to this constraint was this study's requirement to include a range of word lengths, to enable measurement of the 'word length effect', as discussed previously. Children's early vocabularies tend to be built around one and two syllable words, making identification of suitable longer words difficult. The aim, as far as possible, was to match words across the different word lengths by a measure of probable word knowledge.

Data on age of acquisition was first investigated. Section 1.5.1.2 described that most commonly used method of collecting values for age of acquisition is to ask adults how old they believe they were when they learnt each word (Gilhooly and Logie, 1980). There are however potential problems with this method. First, people tend not to be accurate in recalling their acquisition of knowledge temporally (e.g. Loftus and Loftus, 1980). Second, language is constantly evolving and words move in and out of popularity and usage. Thus assuming an accurate estimation by adults of the age at which they acquired a word would not necessarily correspond to the age at which children in a different generation acquired a word.

Other values were considered. For example, there are known high correlations between the frequency of a word and its reported age of acquisition (Bird, Franklin and Howard, 2001). Therefore by using word frequency data it may be possible to assume that the words are also acquired early. However, this is not always the case. For example, function words, while occurring frequently in language, emerge later than some less abstract content words (nouns, verbs).

For the purposes of the present study, a measure that was considered a more precise gauge of children's word knowledge was to obtain a confidence rating from nursery practitioners. Six experienced nursery practitioners from different nurseries and children's centres in the socio-geographical area of the planned data collection were given a list of forty-four concrete nouns from 1-4 syllables with varying stress patterns. They were given the following instructions:

"Please rate on a scale of 1-7 how confident you are that a child aged three years old would be able to name the following objects (where 1 is not at all confident and 7 is very confident). It is not important whether the child would pronounce these words correctly.

Please assume the child is developing typically and speaks only English."

Initially it was planned that the ratings would be used to inform selection of the stimuli, so that words would be matched for familiarity across syllable numbers. However this was not possible due to the very small number of 3 and 4 syllable imageable nouns with suitable stress patterns. Where there was a choice of words for a given syllable number and stress pattern, the ratings were, however, used to choose between stimuli. For example for 3 syllable words with the main stress on the final syllable, 'trampoline' (mean confidence rating =5.3) was selected in place of 'tambourine' (mean confidence rating =3.8).

Following collection of these pilot data, 28 stimulus words (10 single syllable, and 6 at each of the multi-syllabic lengths) were selected for use in the tasks for the main data collection. Ten (rather than six) 1 syllable words were selected so that these could also be used as the closed set of words in the span task. This allowed cross-referencing to be made regarding the children's pronunciation of the single syllable words. The words are listed in table 2-2, together with the mean ratings obtained from the six nursery practitioners. Overall means for each length are also given.

Table 2-2 Selected words with familiarity ratings

Word	Mean Rating	Word	Mean Rating
Toe	6.7	Dinosaur	6.3
Ear	7	Kangaroo	4.3
Glove	6.2	Banana	7
Egg	5.7	Trampoline	5.3
Tree	7	Umbrella	5.8
Door	6.8	Elephant	5.8
Car	7	**Mean 3 syllable**	**5.8**
Cat	7	Caterpillar	6.2
Ball	7	Helicopter	5.7
Sand	6.8	Binoculars	2.7
Mean 1 syllable	**6.7**	Avocado	3.2
Tiger	6	Macaroni	3
Tractor	5.2	Harmonica	1
Princess	6.2	**Mean 4 syllable**	**3.6**
Giraffe	5.3		
Rabbit	5.7		
Guitar	4.7		
Mean 2 syllable	**5.5**		

As is evident in table 2-2, familiarity ratings were lower for 4 syllable words compared to the other word lengths, potentially resulting in a confounding variable when comparisons were made across words of different lengths. The potential confound was minimised by the methods of scoring and analysing the data: any words that were unknown by each child (as judged by a naming task) were removed from the selection of known words for that child. However it could be argued that the variable continued to exert some level of influence where children were able to name a relatively unfamiliar item. For example, a child who could name 'binoculars' would likely use this word less frequently in his/her spontaneous talking than e.g. the word 'cat'. This might mean that the word was less stable phonologically and therefore more prone to errors during repetition.

2.2.3. Non-word stimuli for the repetition task

To enable accurate comparison between repetition of word and non-word stimuli, the non-words needed to be as closely matched to words on ‘word’ length, phoneme composition and stress pattern. Obviously non-words cannot match words on items such as familiarity and imageability, but attempts were made to match them to the words on phonotactic probability. The design of the non-words is discussed in the following account with respect to these factors.

2.2.3.1. Phoneme composition and stress

Initially it was decided to create the non-words by altering the voicing of the consonant phonemes and replacing the vowel sounds (except schwa) from the word stimuli. New vowel sounds would be matched to the originals by length. For example, ‘tiger’ (/tɑɪgə/) would become /deɪkə/. However, using this method to create the stimuli resulted in two problems. The first problem pertained to the phonotactic probability of the non-words. Phonotactic probabilities were calculated for the words and non-words designed using this method, and there was found to be a significant difference between these. The second problem in creating the non-word stimuli in this way was that one of the experimental factors, namely ability to produce the phoneme, was affected. Developmentally it is common for children to go through a period of ‘voicing’ voiceless consonants. Therefore by altering the voicing of the consonants in the non-words, this might confound the results.

It was therefore decided to break down the words into their component syllables, maintaining phonological structure and stress of the syllable, and then to reassemble the syllables into new words, matching syllable number and stress pattern to the original words.

To illustrate, the non-word /bəbrɛtə/ was created using the first (weak) syllable of /bənɑnə/, the second (strong) syllable of /ʌmbrɛlə/ and the final (weak) syllable of /træktə/. To match non-words to real words in terms of articulatory complexity, where the onset consonant of a syllable agreed with the coda consonant of the previous syllable in the real word, this was altered in the non-words. For example, the first syllable of the word /træmpəlin/ was changed to /træŋkəru/, to facilitate articulation of the velar plosive /k/, as is typical for English words. Individual non-words were not matched exactly to the individual word-equivalent stimuli in terms of phonological complexity. Therefore, while some non-words were phonologically more complex than their word-equivalent, others were simpler. This was not considered to be a problem due to the methods of analysing the data. Indeed, the method adopted to create the non-words aimed to match the groups of words with the groups of non-words for each syllable number (i.e. group of 1 syllable words compared to group of 1 syllable non-words etc) on this dimension and this reflected the methods for analysis. Single syllable words were created using the consonant sounds of the real words and altering the vowel sound (maintaining original vowel length).

The final set of non-words are presented in the table 2-3.

Table 2-3 Final word and non-word stimuli set (syllables that are underlined represent the main stress in the word/non-word; primary and secondary stress is marked /ˈ/ and /ˌ/ respectively for the non-words)

Word	Phonetic Transcription	Word Structure	Non-word (phonetic transcription)	Non-word Structure
toe	təʊ	cv	teɪ	cv
ear	ɪə	cv	əʊə	v
door	dɔ	cv	dɑ	cv
car	kɑ	cv	kɜ	cv
cat	kæt	cvc	kɛt	cvc
ball	bɔl	cvc	bɜl	cvc
sand	sænd	cvcc	sɒnd	cvcc
tree	tri	ccv	trɜ	ccv
glove	glʌv	ccvc	glɪv	ccvc
egg	ɛg	vc	ɪg	vc
giraffe	dʒɪ.<u>rɑf</u>	cv.cvc	dʒɪ.ˈ<u>lin</u>	cv.cvc
rabbit	<u>ræ</u>.bɪt	cv.cvc	ˈ<u>ræ</u>.ləz	cv.cvc
guitar	gɪ.<u>tɑ</u>	cv.cv	gɪ.ˈ<u>sɛs</u>	cv.cvc
tiger	<u>taɪ</u>.gə	cv.cv	ˈ<u>taɪ</u>.kə	cv.cv
tractor	<u>træk</u>.tə	ccvc.cv	ˈ<u>træk</u>.lə	ccvc.cv

princess	prɪn.sɛs	ccvc.cvc	prɪn.ˈtɑ	ccvc.cv
dinosaur	daɪ.nə.sɔ	cv.cv.cv	ˈdaɪ.tə.fənt	cv.cv.cvcc
banana	bə.nɑ.nə	cv.cv.cv	bə.ˈbrɛ.tə	cv.ccv.cv
kangaroo	kæŋ.gə.ru	cvc.cv.cv	ˌkæm. pə.ˈrɑf	cvc.cv.cvc
umbrella	ʌm.brɛ.lə	vc.ccv.cv	ʌm.ˈbɑ.dəʊ	cv.cv.cv
trampoline	træm.pə.lin	ccvc.cv.cvc	ˌtræŋ.kə.ˈru	ccvc.cv.cv
elephant	ɛ.lə.fənt	v.cv.cvcc	ˈɛ.və.sɔ	v.cv.cv
caterpillar	kæ.tə.pɪ.lə	cv.cv.cv.cv	ˈkæ.nɪ.ˌkɒp.ni	cv.cv.cvc.cv
helicopter	hɛ.lɪ.kɒp.tə	cv.cv.cvc.cv	ˈhɛ.nə.kjɪ.lə	cv.cv.ccv.cv
binoculars	bɪ.nɒ.kjɪ.ləz	cv.cv.ccv.cvc	bɪ.ˈmɒ.pɪ.tə	cv.cv.cv.cv
harmonica	hɑ.mɒ.nɪ.kə	cv.cv.cv.cv	hɑ.ˈnɒ.gə.nə	cv.cv.cv.cv
avocado	æ.və.kɑ.dəʊ	v.cv.cv.cv	æ.lɪ.ˈrəʊ.gə	cv.cv.cv.cv
macaroni	mæ.kə.rəʊ.ni	cv.cv.cv.cv	mæ.lə.ˈkɑ.bɪt	cv.cv.cv.cvc

2.2.3.2. Neighbourhood density

Two further factors that were taken into consideration were neighbourhood density and phonotactic frequencies of the items.

As discussed in the introduction (section 1.5.1.3), neighbourhood density affects both repetition and naming. While it would not be possible to match words and non-words exactly by neighbourhood density, and not possible to match across word syllable numbers (shorter words necessarily having denser neighbourhoods in English than their longer counterparts), neighbourhood densities were nevertheless calculated.

The following table (table 2-4) presents neighbourhood density values collected from three sources: the lexical database created by Goswami and Cara (2002) (http://portail.unice.fr/jahia/page12414.html); data calculated for the study by McKean (2009); and through using the neighbourhood density calculations from The Irvine Phonotactic Online Dictionary (Vaden, 2009) (http://www.iphod.com/calculator/). It should be noted that the neighbourhood density for each individual may be different depending on the size of their vocabulary. It is unlikely, for example, that children's developing vocabularies have the same number of phonological neighbours as an adult's, from which the norms below are derived. It might be expected therefore that any predicted influences of neighbourhood density are reduced.

Table 2-4 Neighbourhood densities for word and non-word stimuli

Word	Number of Phonological Neighbours	Non-word	Number of Phonological Neighbours
Toe	46	teɪ	57
Ear	32	əʊə	22
Glove	9	glɪv	3
Egg	20	ɪg	23
Tree	42	trɜ	13
Door	44	dɑ	44
Car	45	kɜ	43
Cat	54	kɛt	30
Ball	49	bɜl	43
Sand	11	sɒnd	17
Mean	**35.2**	**Mean**	**25.5**
Tiger	15	taɪ.kə	9
Tractor	4	træk.lə	1
Princess	0	prɪn.tɑ	3
Giraffe	0	dʒɪ.lin	0
Rabbit	9	ræ.ləz	1
Guitar	0	gɪ.ses	0
Mean	**4.67**	**Mean**	**2.33**
Dinosaur	0	daɪ.tə.fənt	0
Kangaroo	0	kæm.pə.rɑf	0
Banana	1	bə.brɛ.tə	0
Trampoline	0	træŋ.kə.ru	0
Umbrella	0	ʌm.bɑ.dəʊ	0
Elephant	2	ɛ.və.sɔ	0
Mean	**0.5**	**Mean**	**0**
Caterpillar	0	kæ.nɪ.kɒp.ni	0
Helicopter	0	hɛ.nə.kjʊ.lə	0
Binoculars	0	bɪ.mɒ.pɪ.tə	0
Avocado	0	æ.lɪ.rəʊ.gə	0
Macaroni	0	mæ.lə.kɑ.bɪt	0
Harmonica	0	hɑ.nɒ.gə.nə	0
Mean	**0**	**Mean**	**0**

Overall, there was no significant difference between words and non-words t(27)=1.770, p=0.088 in terms of neighbourhood densities. As is evident from examining the raw scores and the means above, there is an unavoidable difference in the neighbourhood densities across the different item lengths.

2.2.3.3. Phonotactic frequencies

It was important to consider phonotactic probability in selecting the stimuli, as items with higher phonotactic frequencies have been shown to be repeated more accurately than those with lower frequencies (Gathercole et al., 1999). Ideally phonotactics would not differ significantly across different word lengths, or between word groups (words versus non-words).

Phonotactics can be calculated in different ways, described by Vitevitch and Luce (2004). They designed a calculator that computes phonotactics for American English words. However, since American vowels differ from Standard English vowels, this method of calculation would not have provided accurate values for the words in the present study. Calculations were therefore made using CELEX data and a formula used in the study by McKean (2009). The probabilities of each pair of phonemes arising in each word were collected. Biphoneme probabilities are listed in table 2-5, and a summary of the means and ranges for each word length is given in table 2-6. The mean of each word's (or non-word's) phonotactic probability is also listed.

Table 2-5 Phonotactic probabilities for word and non-word stimuli

Word	Biphoneme probability	Mean probability (4 d.p.)	Non-word	Biphoneme probability	Mean probability (4 d.p.)
toe	0.0089	0.0089	teɪ	0.0340	0.0340
ear	0.0340	0.0340	əʊə	0.1020	0.0102
glove	0.0431 0.0145	0.0231	glɪv	0.0431 0.2089	0.0874

	0.0118			0.0102	
egg	0.00507	0.0051	ɪg	0.0066	0.0066
tree	0.0480 0.0601	0.0540	trɜ	0.0480 0.0000	0.0240
door	0.0063	0.0063	dɑ	0.0038	0.0038
car	0.0303	0.0303	kɜ	0.0072	0.0072
cat	0.0610 0.0085	0.0347	kɛt	0.0041 0.0281	0.0161
ball	0.0107 0.0079	0.0093	bɜl	0.0070 0.0604	0.0337
sand	0.0051 0.0053 0.1546	0.0055	sɒnd	0.0070 0.1165 0.1546	0.0927
tiger	0.0263 0.0005 0.1039	0.0435	taɪ.kə	0.0263 0.0388 0.1376	0.0675
tractor	0.0480 0.0343 0.0610 0.3005 0.0911	0.1069	træk.lə	0.0480 0.0343 0.0610 0.0436 0.0564	0.0487
princess	0.1512 0.2460 0.1244 0.1644 0.0572 0.0877	0.1385	prɪn.tɑ	0.1512 0.2460 0.1244 0.2911 0.0910	0.1808
giraffe	0.1260 0.1576 0.0144 0.1040	0.0804	dʒɪ.lin	0.1260 0.0161 0.0234 0.0447	0.0525
rabbit	0.0343 0.0035 0.1263 0.0671	0.0578	ræ.ləz	0.0343 0.0859 0.0564 0.0331	0.0524
guitar	0.0653 0.1189 0.0034	0.0625	gɪ.ses	0.0653 0.1252 0.2911 0.0910	0.1431
dinosaur	0.0046 0.1167 0.0273 0.0622 0.0033	0.0428	daɪ.tə.fənt	0.0046 0.0837 0.0910 0.0282 0.0364 0.2896 0.0714	0.0864
kangaroo	0.0331 0.0136 0.5149 0.1039 0.1576 0.0397	0.1438	kæm. pə.rɑf	0.0331 0.0154 0.3634 0.1408 0.1576 0.0144	0.1184

				0.1040	
banana	0.1785 0.0801 0.0005 0.0180 0.0273	0.0609	bə.brɛ.tə	0.1785 0.1159 0.0495 0.0761 0.0671 0.0910	0.0967
trampoline	0.0480 0.0343 0.0154 0.3634 0.1408 0.1248 0.0234 0.0447	0.0993	træŋ.kə.ru	0.0480 0.0343 0.0136 0.2800 0.1376 0.1576 0.0397	0.0102
umbrella	0.1532 0.3696 0.0494 0.0761 0.0873 0.0564	0.1320	ʌm.bɑ.dəʊ	0.2056 0.0029 0.0004 0.0313 0.0714	0.0623
elephant	0.0765 0.0564 0.0000 0.2896 0.0714	0.0884	ɛ.və.sɔ	0.1753 0.0273 0.0313 0.0146	0.0621
caterpillar	0.0331 0.1269 0.0910 0.0532 0.0528 0.0765 0.0564	0.1388	kæ.nɪ.kɒp.ni	0.0331 0.1269 0.0497 0.1222 0.0638 0.0094 0.0237 0.0041	0.0541
helicopter	0.0384 0.0706 0.2089 0.1222 0.0638 0.0094 0.3447 0.0910	0.1186	hɛ.nə.kjʊ.lə	0.0384 0.1839 0.1643 0.0261 0.0210 0.0716 0.3275 0.0564	0.1079
binoculars	0.1263 0.0633 0.0259 0.0148 0.0210 0.0716 0.3275 0.0564 0.0331	0.0822	bɪ.mɒ.pɪ.tə	0.1263 0.0616 0.0095 0.0587 0.0528 0.1189 0.0910	0.0714
avocado	0.0582 0.1643	0.0561	æ.lɪ.rəʊ.gə	0.0859 0.2089	0.0804

	0.0261 0.0152 0.0581 0.0146			0.0161 0.0363 0.0312 0.1039	
macaroni	0.0387 0.0422 0.1376 0.1576 0.0363 0.0358 0.0041	0.0646	mæ.lə.kɑ.bɪt	0.0387 0.0859 0.0564 0.0261 0.0152 0.0267 0.1263 0.0671	0.0553
harmonica	0.0275 0.0832 0.0096 0.0267 0.0497 0.1222 0.1376	0.0652	hɑ.nɒ.gə.nə	0.0275 0.0180 0.0259 0.0050 0.1039 0.0801 0.0273	0.0411

There was no significant difference in the phonotactic probabilities of the words and the non-words: t(27)=0.365, p=0.718. For the multi-syllabic items (1 syllable items were not included due to unequal sample sizes), an ANOVA revealed no significant main effect of word type (F(1,10) =1.25, p=0.31) or word length (F(1, 10)=0.118, p=0.89). There was also no significant interaction between word type and length (F(2,10)=0.703, p=0.52). It is evident however that the 1 syllable items have lower phonotactic probabilities than the other items and this could have implications for the repetition of these items.

Table 2-6 Mean phonotactic probabilities for items of different lengths

	Word		**Non-word**	
Length	**Mean**	**Standard deviation**	**Mean**	**Standard deviation**
1 syllable	0.0211	0.0159	0.0316	0.0310
2 syllable	0.0816	0.0323	0.0908	0.0518
3 syllable	0.0945	0.0358	0.0727	0.0341
4 syllable	0.0876	0.0306	0.0684	0.0217

2.2.4. Selecting picture stimuli for the naming task

Following selection of the word stimuli, two sets of pictures representing the words were chosen using an internet search engine. These were agreed by two independent judges to be suitable picture representations of the target words. The independent judges were both female, adult, practising speech and language therapists. They were not otherwise involved with the study. Two sets of pictures were selected so that one set could be used in the naming task, and the other set in the recognition task for each child. These would be counterbalanced (see procedure, section 2.4.3.).

Consideration was given to which type of picture should be used. Many published tests of receptive vocabulary and naming ability use line drawings as their stimuli, e.g. the British Picture Vocabulary Scale (BPVS) (Dunn et al., 1999), the Psycholinguistic Assessment of Language Processing in Aphasia (PALPA) (Kay, Lesser and Coltheart, 1992) and the Renfrew Word Finding Test (Renfrew, 1995). Given the young age of participants in the present study, it was decided however to use colour photograph stimuli for the present task, these being assumed to be more easily recognisable and more motivating to look at. Stimuli pictures are presented in table 2-7.

Table 2-7 Pictures for eliciting naming

Target	Picture Set One	Picture Set Two
Toe	[redacted]	[redacted]
Ear	[redacted]	[redacted]

Glove	[redacted]	[redacted]
Egg	[redacted]	[redacted]
Tree	[redacted]	[redacted]
Door	[redacted]	[redacted]
Car	[redacted]	[redacted]
Cat	[redacted]	[redacted]
Ball	[redacted]	[redacted]
Sand	[redacted]	[redacted]
Tiger	[redacted]	[redacted]
Tractor	[redacted]	[redacted]
Princess	[redacted]	[redacted]
Giraffe	[redacted]	[redacted]
Rabbit	[redacted]	[redacted]

Guitar	[redacted]	[redacted]
Dinosaur	[redacted]	[redacted]
Kangaroo	[redacted]	[redacted]
Banana	[redacted]	[redacted]
Trampoline	[redacted]	[redacted]
Umbrella	[redacted]	[redacted]
Elephant	[redacted]	[redacted]
Caterpillar	[redacted]	[redacted]
Helicopter	[redacted]	[redacted]
Binoculars	[redacted]	[redacted]
Avocado	[redacted]	[redacted]
Macaroni	[redacted]	[redacted]
Harmonica	[redacted]	[redacted]

2.2.5. Selecting the stimuli for the span task

The present study used a word recall (or span) task as its measure of PSTM. 'Words' as opposed to digits were used because it was necessary for the lexical items to be familiar for the children. Three-year old children would not be expected necessarily yet to know numbers, and this lack of knowledge might confound the results. Several additional factors are known to influence span: item length (e.g. Baddeley et al. 1975), phonological similarity (Conrad and Hull, 1964), semantic similarity (e.g. Saint Aubin and Poirier, 1999) and syntactic relationships (Baddeley et al. 2009). Therefore, items included in the span task needed to be single syllable words that did not rhyme, did not have close semantic links and did not relate to each other syntactically.

Some studies have instead used non-word span tasks to measure PSTM (e.g. Gathercole, Pickering, Ambridge and Wearing, 2004; Hulme, Maughan and Brown, 1991; Hulme et al., 1995). They have found that recall of these unfamiliar items is reduced in comparison to words. This can be explained in terms of 'chunking'. Novel words (or non-words) do not have existing lexical templates and therefore each phoneme within the 'word' requires additional resources from the PSTM system. The system therefore reaches saturation point more quickly than it does when storing sequences of familiar phonological material, as in the case for real words. This leads it to store fewer sequenced items successfully. During span for non-words the PSTM system would however presumably draw on existing long-term word-general knowledge, such as sensitivity to phonotactics. It is therefore less clear, during non-word span tasks, what would be defined as a 'unit' of short-term memory. For this reason it was decided not to include span for non-words in the present study.

It was decided that the words used in the span task would be drawn from the stimuli used in the repetition and naming tasks. Inclusion in the naming task would establish whether the words were familiar to the child and therefore the level of processing available to the child during span recall (see chapter 1, sections 1.4, 1.5). Stimuli used in the span task are presented in table 2-8. As it is known that immediate serial recall is better for grouped items that are similar in meaning than for semantically dissimilar words (Poirier and Saint-Aubin, 1995; Saint-Aubin and Poirier, 1999), where words belonged to the same semantic category (i.e. ear and toe), these words were not included adjacent to each other in the same span set.

Table 2-8 Stimuli used in the span task

2-word span	3-word span	4-word span	5-word span
toe–egg	car-glove–toe	egg–toe-tree–car	egg–sand–car–cat–glove
door–sand	sand-cat–tree	door–sand–ear–cat	toe-door-tree-ear-egg
tree-car	egg-door–ear	glove–egg–car–toe	glove-cat-ear-car-tree
ear–cat	toe-sand–door	sand–tree–cat-egg	sand-door-egg-toe-cat
glove-egg	ear-car–cat	ear–door–egg-sand	door-tree-sand-car-toe

2.2.6. Selecting the stimuli for the recognition task

The ‘recognition’ task aimed to identify whether participants had a stored representation (receptive phonological-semantic representation) of the word in their lexicon. It aimed to tap a different level of ‘knowledge’ for the words than the naming task. As discussed in chapter 1 (section 1.4.1.4), for some of the items that the children are not able to name, it is likely that the children have some previous experience of the words, but either the picture stimulus or the reduced familiarity with the word is not sufficient to enable them to retrieve the word when shown a picture representing the item. It is plausible that the children would repeat these familiar, but unnamed items more accurately than items they had never heard before. The picture-pointing recognition task would pick up this level of 'knowledge' or familiarity.

In designing the task, existing language assessments that target measurement of single word comprehension were used as models. Examples of assessments include the British Picture Vocabulary Scale (BPVS) 3rd edition (Dunn et al., 1999) which is an assessment with norm-referenced scores for a population of children aged 3-16 years, designed to measure the child's vocabulary. A further assessment that follows the described design is the single word receptive language component of the PALPA (Kay, et al., 1992). These assessments take a multiple-choice format, i.e. they present the person being assessed with a number of pictures that include the target stimulus and a number of distractors. The assessments require the person being assessed to point to a picture in response to the verbal presentation of the target label by the assessor. Distractors may be related or unrelated semantically, visually or phonologically.

Performance on a task like this is likely be influenced by the types of distractors presented, the degree of target-distractor relatedness, the type and degree of the language difficulty experienced by the individual and an interaction between these factors. For example, an individual whose primary deficit is semantic, might make more errors if distractor stimuli are closely related semantically. By contrast, it would seem reasonable to assume that an individual whose primary deficit is phonological would make more mistakes where distractors are close phonological neighbours. One study that investigated types of distractors on individuals' performance was that by Bishop and Byng (1984). They devised the test of *Lexical Understanding with Visual and Semantic Distractors*. The test was used with adults with aphasia, so findings are of limited relevance to the present study. However the design of the assessment is informative in the design of the present study's task and will therefore be discussed.

Each target in the study by Bishop and Byng (1984), eighty in total, was presented with four or eight distractor pictures. Distractors were reported to be related visually, semantically, semantic-visually or unrelated to the target stimulus. Visual similarity was reported to be "judged subjectively" (Bishop and Byng, 1984, p236), but the study does not give any further information about how this judgement was made. Examples of visually similar items are given, e.g. thistle-shaving brush, butterfly-bow, balloon-magnifying glass, button-coin (Bishop and Byng, 1984). It could however be argued that these items differ from each other in terms of degree of visual, semantic and phonological similarity. This might lead to item specific error effects. A second distractor was related semantically to the target. Semantic similarity was judged by asking 14 "normal subjects" (Bishop and Byng, 1984, p236) to decide whether 320 orthographically presented word pairs were synonymous. The semantic-visual distractors were selected from combining the above two approaches, and random distractors were selected from those that were judged not to be synonymous in the above experiment.

The present task was designed to include four pictures on each trial: the target item together with three distractors. In designing the task, several factors were considered potentially relevant. These were: age of acquisition, imageability, phonological similarity and semantic similarity.

Semantic distractors were identified as items belonging to the same category, e.g. the semantic distractor for ‘toe’ was identified as ‘finger’; the semantic distractor for ‘car’ was identified as ‘lorry’. Phonological distractors were identified either by having identical onset phonemes, e.g. ‘**gl**ove’ and ‘**gl**ue’ or by their identical rime, e.g. ‘egg’ and ‘peg’. Where possible, these items were also matched in terms of word length. In addition, words were identified that were considered by two independent judges (see section 2.2.4) to be acquired early.

As stimuli were being presented in picture form, all stimuli needed to be imageable. No further consideration was made to the degree of imageability and no consideration was made of the distractors frequency or age of acquisition. This was due to inherent difficulties in matching stimuli along all of these dimensions. In addition, as the task served only to identify whether children had some receptive knowledge of the target stimulus and was not the main focus of the present study, it was not considered necessary to match distractor stimuli according to all possible factors.

Each target-item picture was presented together with three distractor pictures, one which was semantically related (e.g. a lorry, related semantically to a car), one was phonologically related (e.g. a cow, related phonologically to a car) and one was considered unrelated (e.g. a sock, arguably unrelated to a car).

Items were judged as semantically related, phonologically related or unrelated by the researcher and by one other judge (an adult female academic linguist, who was one of the supervisors of the project). No objective measures of relatedness were used in the selection of the stimuli and this is a limitation of the test. For example, a tiger may be more semantically related to a lion than a car is to a lorry. This might make the child more likely to erroneously select lion for tiger compared to lorry for car. Furthermore, some stimuli may be considered to be both semantically and phonologically related, e.g. ‘eye’ and ‘ear’. The full list of stimuli used in this task is given in table 2-9.

Table 2-9 Stimuli used in the recognition task

Target	Semantic distractor	Phonological distractor	Unrelated distractor
car	lorry	cow	sock
caterpillar	butterfly	cauliflower	watermelon
toe	finger	tie	key
ear	eye	bear	spoon
tiger	lion	table	violin
ball	bat	bowl	nose
tree	flower	train	castle
macaroni	spaghetti	maraca	calculator
egg	chicken	peg	hand
avocado	pear	anteater	recorder
door	window	doll	road
dinosaur	unicorn	dungarees	hat
guitar	piano	glitter	hippo
tractor	van	trousers	palace
rabbit	guinea pig	rattle	house
giraffe	zebra	grass	sofa
binoculars	goggles	volcano	ladybird
trampoline	bouncy castle	tambourine	crocodile
kangaroo	koala	caravan	broccoli
glove	scarf	glue	toast
banana	pineapple	tomato	camera
umbrella	parachute	ambulance	rhino
elephant	gorilla	telescope	balance bar
helicopter	aeroplane	radiator	hammer
princess	crown	printer	tortoise
harmonica	accordion	thermometer	asparagus
cat	dog	cot	shoe
sand	spade	sun	nest

2.2.7. Standardised Assessments

In addition to the tasks described above, four standardised assessments (or subtests from those assessments) were used in the study. The first was the *Pre-School Language Scales 4* (PLS4) (Zimmerman, Steiner and Pond, 2002). This was selected because it is an omnibus standardised assessment of young children’s language. The test has an internal consistency value (Cronbach's alpha) of α=0.88 for auditory comprehension at both ages 3-3.5 years and 4-4.5 years, α=0.90 for expressive communication at age 3-3.5 years and α=0.87 at age 4-4.5 years (measured by examining inter-correlations between items on the test). Values of 0.7 to 0.8 are generally considered to be acceptable (Field, 2009). The second standardised assessment was the *British Ability Scales II* (BAS II) (Elliot, Smith and McCullough, 1996). Two subtests (*Block Building* and *Picture Similarities*) were selected from this assessment to serve as a measure for the children’s non-verbal cognitive skills. Scores from these subtests form the *Special Non-verbal Composite* score which is considered an appropriate tool for measuring non-verbal ability of children with speech and language difficulties, aged 2.6-3.5 years. The same subtests are also appropriate for older pre-school children, and were used at the second time-point of the study. These were used in favour of the three subtests that make up the *Non-verbal Composite* for this age group, which would have further increased the assessment time for the children, which was already long. The subtests have acceptable internal reliability coefficients at both ages. These values are as follows: Block Building α=0.89 at 3 - 3 1/2 years and α =0.88 at 4 - 4 1/2 years; Picture Similarities α =0.82 at 3 - 3 1/2 years and α=0.86 at 4 - 4 1/2 years. Test-retest reliabilities are not available for the BAS II subtests for pre-school children as an insufficient number of children this age were recruited to the test-retest study. However, test-retest reliabilities are available for the Differential Ability Scales (DAS), which is the US version of the BAS. Based on the US data for this age-group, Block Building task has a reliability value of r=0.67 and Picture Similarities has a reliability value of r=0.56. The third standardised assessment was the *Sentence Imitation Test* from the *Early Repetition Battery* (ERB) (Seeff-Gabriel, Chiat and Roy, 2008). This test has anacceptable internal consistency value of α=0.92 and a test-retest reliability value of α=0.88. Furthermore, the authors report an inter-rater reliability of α=0.98. The fourth test was the *Word Structure* test of the *Clinical Evaluation of Language Fundamentals (*CELF*) 2* (Wiig et al., 2004), which was selected as an independent measure of the children's ability to use morphemes. This test has an acceptable internal consistency value of α=0.79 at age 4 - 4 1/2 years, and it has an overall test-retest reliability value of r=0.81.

2.3. Scoring

Scoring for the standardised tasks in the study followed the methods set out in the manuals (see *PLS4 manual*, Zimmerman et al. (2002); *BAS II manual*, Elliott et al. (1996); *CELF pre-school 2* manual, Wiig et al. (2004); *ERB* manual, Seeff-Gabriel et al. (2007)). The following account describes the method for scoring each of the tasks designed specifically for the present study.

2.3.1. Scoring for the WR and NWR tasks

There are two main methods for scoring NWR tasks reported in the literature. One method scores for whole items correct, thus any deviation from the target is scored as incorrect. A second method scores for percentage phonemes correct. While there is evidence that both methods are discriminating of children with and without language difficulties in English, Graf Estes, Evans and Else-Quest (2007) reported that the difference between the groups using the whole item correct method of scoring was smaller than using the percentage of phonemes correct method (d = 0.48 compared to d = 1.17). For this reason, and because the present study was interested in the effects of phonological errors on repetition accuracy, the latter method of scoring was adopted.

The children's responses during the repetition tasks were transcribed using broad phonetic transcription. Scoring of the tasks was based loosely on the procedure used by Dollaghan and Campbell (1998) and also aligns with one of the methods used by Dispaldro, Leonard and Deevy (2013a).

Where children omitted or added syllables in their response, the response syllable sequence was aligned as closely as possible to the target sequence, using vowels as anchors. In most cases, vowel-anchoring was obvious (as in the first example below). In other cases (as in the second example below), vowel anchoring was established by ensuring the highest number of identical vowels in the target and response without disrupting the vowel order. The phonemes of the repetition response were compared to the phonemes of the target stimulus.

e.g.	target:	/bənɑnə/
	response:	/nɑnə/
e.g. 2	target:	/hɛlɪkɒptə/
	response:	/hɒpəhɪwə/

Phonemes were scored either as correct or incorrect based on whether they matched the corresponding target phoneme. A phoneme that was added by the participant in his/her repetition (as in the second example above) was ignored. This was consistent with Dollaghan and Campbell's procedure (1998), where phoneme additions were not counted as errors. Their rationale was that they were "interested in the extent to which participants were able to represent the target phonemes in memory long enough to repeat them; additions by definition do not reflect a loss of information about the target phonemes." (p. 1139).

Like Dollaghan and Campbell's study, the present study was investigating the children's ability to hold phonological information in short-term memory. However the present study was also interested in determining which other factors might influence children's performance on this task. It attempted to determine the extent to which long-term existing knowledge about words was being applied in repetition of both real words and non-words. It could be argued that during repetition, any distortion (omission, addition or alteration) from the target response represents an inaccurate short-term phonological trace or the application of an inaccurate long-term representation for the word. Errors of phoneme addition might therefore provide useful information about children's storage of words. After consideration, however, the present study proposes that children making errors of addition would also likely make errors of omission and therefore their inaccurate representations would be captured as a loss of phonological information. Furthermore, phoneme additions were rare: at the first time-point, thirty-two (out of 54) of the children made no phoneme additions at all and the overall mean number of phonemes added was 0.56 for words (range between 0 and 5) and 0.39 for non-words (range between 0 and 3). The maximum number of phonemes added in any child's overall repetition (words and non-words) was 5, and this represented 1.52% of the phonemes produced during the child's repetition of words and non-words. Phonemes that were erroneously added were therefore ignored.

There were a total of 26 phonemes in each of the 1 syllable word and non-word items, 31 phonemes in each of the 2 syllable word and non-word items, 42 phonemes in each of the 3-syllable groups and 50 phonemes in each of the 4 syllable groups. Phonemes repeated correctly were totalled for 1-syllable, 2-syllable, 3-syllable and 4-syllable items. Percentage phoneme correct scores were calculated for each of the syllable groups. Overall percentage phoneme correct scores for all items in both of the real-word and non-word groups were also calculated. Individual error analyses for percentage phonemes substituted, omitted or added were not calculated as this was beyond the scope of the present study.

In summary, scoring for the word and non-word repetition task initially involved counting the overall number of phonemes correct. Words and non-words were then divided into items of different lengths, correct phonemes were counted and the percentage of phonemes correct was determined for each item length. Further scoring of the repetition tasks ensued and this is described in section 2.3.2.

2.3.2. Purpose and scoring of the naming task

The naming task had several purposes. The first purpose was to identify which words were known by each child and which words were *unknown.* The naming task thereby provided a means of assessing the child's 'knowledge' of the word. Scoring for this purpose was not simple, particularly for the children with speech and language needs, where phonological systems were delayed or disordered. This made it difficult to determine which words could be considered to be 'known' by the child, i.e. had sufficiently accurate phonological representations and which were not known. It was decided that all known typical phonological processes (e.g. stopping, voicing etc.) would not be counted as errors and therefore words containing these 'errors' were considered to be *known*. Further processes that were disordered, but which appeared as a consistent pattern in each individual child's speech were also allowed. Responses were also required to be consistent across the two naming tasks (see section 2.4.5.1). Where items were difficult to score as 'known' or 'unknown', these items were discussed with the project supervisors until a consensus was reached.

The second purpose was to compare repetition of *known* words with repetition of *unknown* words. The number of phonemes correctly repeated was counted for all the *known* words for each child. This figure was then divided by the total number of phonemes that the child could have scored correctly for his/her sample of repeated *known* words. The same calculation was made for *unknown* words (i.e. those that the child was unable to name, or which were unrecognisable as the target when they were named) that the child repeated.

The third purpose of the naming task was to identify typical and atypical phonological processes that were present in the child's speech. This served to enable analysis of the children's repetition errors. When working with children this age it is often difficult to determine which of the children's repetition errors are errors due to delayed or disordered phonology (Bishop et al. 1996) (see also introduction for chapter 4). Through direct comparison of children's phoneme errors made during the repetition task and those made during the naming task, the study aimed to better determine the cause of apparent repetition errors. For this purpose, scoring involved direct comparison between naming and repetition of the target syllables in naming and in the two repetition conditions (word and non-word). Where a phoneme error occurred in repetition, this was compared with the child's naming attempts. Where the child also consistently made the same phoneme error for the named stimulus (based on their two naming attempts), this was scored as correct for the repetition task.

Some children are known to be inconsistent in their productions of lexical items (see Dodd and Bradford, 2000) during naming; children with inconsistent phonological disorder representing around 9.4% of those presenting with speech sound difficulties (Broomfield and Dodd, 2004). Inconsistent productions might confound the results, as discrepancies in the children's responses on the naming and repetition tasks, thought to reflect differences in skills exploited by the tasks might instead be due to inconsistent productions. The purpose of presenting the naming task for the second time was to minimise this type of scoring error.

In the case of there being evidence that children were omitting the weak syllable in words (and non-words), where scores were used to predict SR, responses were scored both allowing for this error and not allowing for it (see results in chapter 7). This had the rationale that previous studies had identified repetition of weak syllables to be particularly problematic for children with language difficulties, particularly where the weak syllable forms part of an iambic foot structure (see section 1.5.2.1) (e.g. Chiat and Roy, 2007).

Where the child had not correctly named the item during either naming attempts, a phoneme error made during repetition was only scored correct where the error was clearly part of a phonological process that was present for that particular phoneme throughout the other named items.

A fourth purpose of the naming task was to serve as a measure of the children's word knowledge, i.e. an approximation to their vocabulary. For this purposes, the naming task was scored for whole items correct, allowing for phonological errors.

The final purpose of the naming task was to facilitate scoring of the span task. As discussed above in the case of the repetition task, errors on the span task may be due to repetition errors, caused perhaps by decay of the phonological representation in phonological short-term memory, or they may reflect delayed or disordered phonological processes. In the latter case these errors would also be apparent in the naming task. In scoring the span task, item errors were therefore compared to the child's naming of the same item. Where phoneme errors were evident in the child’s naming attempt, these were allowed for the purposes of scoring the span task.

In summary, the naming task served several purposes and was scored differently according to each purpose:

1) Children received a simple score out of 28 for the number of whole items they correctly named. Speech errors were ignored for this purpose.

2) Naming responses were analysed for each child and any items that the child had been unable to name were considered to be *unknown*. The WR task was then rescored, splitting *known* words from *unknown* words, and percentage scores were derived for each word length, based on the maximum score achievable for each individual child.

3) Speech errors on the naming task were analysed and compared to each child's production on the repetition tasks. Repetition responses were then rescored, where speech errors occurred consistently on the naming tasks.

2.3.3. Scoring for the span task

In the consideration of the method of scoring the span task, the existing literature was reviewed. Fallon, Groves and Tehan (1999) describe the different ways in which immediate serial recall can be scored at the item level. They explain that while serial recall tasks are typically scored by totalling the number of items recalled in the correct position, some studies adopt an approach where recalled items are counted regardless of the sequential order. They further describe an approach, where the number of items in the correct position is scored as a proportion of the total number of items recalled (Poirier and Saint-Aubin, 1995).

2.3.3.1. Item level scoring

The present study considered two alternative techniques for scoring the span task based on data collected during a pilot study of the tasks. First it considered scoring at the item level, combining the approaches described by Fallon et al. (1999). Responses would attract points for items correctly recalled, and further points for the correct order of items. For example, for the target “sand – cat – tree”, a response of “cat – sand – tree” would score three points for all three items correct and a further one point for the position of the word “tree”. Therefore the maximum score for 2- word spans was four points, and six points for each 3- word span. The rationale for scoring misplaced targets positively was that children were demonstrating that they had retained information about the target items, i.e. the phonological trace had not decayed from the phonological store.

A common recall error made by children during the pilot study was to repeat only two out of the three possible items. For example, for the target “sand – cat – tree”, a response might be “sand – cat” or “sand - tree”. In such cases, using the first novel method of scoring, children would receive a point for each item that was recalled correctly and additional points for position of the items where these were correct. Where there were ambiguities about the ‘correctness’ of the item’s position, e.g. where the first and final but not the middle item were recalled, responses would be scored positively so that the child received the maximum possible number of points. For example, points would be awarded for first and second items positioned correctly in the first example and first and final items recalled correctly in the second example.

2.3.3.2. Phoneme level scoring

Phoneme level scoring was also considered. The potential value of phoneme level scoring was brought to light following observations made during piloting that children made phoneme migration errors, i.e. they were discovered to interchange onset phonemes of words in the list. For example, for the target span “sand – cat – tree”, the response might have been “cand –sat – tree”. In keeping with the scoring methods above, at the item level, a child giving this response would achieve a score of 2 points (for recalling the correct lexical item for tree in the correct place). No points would be assigned for the other two items in the list. It could be argued however that the child giving this response had retained considerable information from the original sequence in their phonological trace, but that it had become distorted or confused at some level of processing. By scoring at the level of the phoneme, this information would still be captured.

Both of the novel methods of scoring the span task described above were however not pursued. The reason for this was that it was considered that scores obtained through these methods would not reflect an accurate measure of the child’s PSTM, as it is typically understood, i.e. number of items recalled in the correct sequence. Instead, children scored a point for each full sequence of words that they recalled correctly, achieving a possible maximum score of 20 points. A *stopping rule* was adopted to minimise fatigue effects and to ensure that no child was required to attempt tasks that exceeded too greatly their ability. Children had to score at least 2 out of 5 full sequences of words at any level (2-word, 3-word, 4-word, etc.) in order to move onto the next level.

2.3.4. Scoring for the recognition task

The purpose of the picture-label recognition task was to identify which items in the stimuli set the children had sufficient knowledge of to correctly identify the target from a choice of four items. In a similar way to the second purpose of the naming task, this would allow comparison of the children’s accuracy in repeating ‘known’ words with their accuracy in repeating ‘unknown’ words. It would provide a different means of assessing ‘knowingness’ about the items. The intention was to use the children’s responses as a means of scoring the repetition task.

Despite the careful design for this task, its overall correlation with the naming task was only modest (r(52)=0.561) and it yielded results that were in conflict with the naming task. Upon scoring the data for this task it was noted that some of the children correctly named items, but when the same items were included in the recognition task they instead pointed to the distractor picture. This was the case for as many as seven of the total twenty-eight items for one of the participants who had named the items correctly. It was difficult to ascertain why this happened. Three explanations are proposed. The first is that the children’s knowledge for the words was fragile and that the presence of the distractors led them to point to the incorrect picture. The second suggestion is that the children were more motivated by one of the other pictures on the screen and pointed to the one that they liked the most. The third explanation is that the children were exhibiting fatigue effects and that these led them to point to any picture on the screen. This is plausible, as the comprehension task was the last task performed by the children at the first visit (see Procedure, section 2.4.).

Given the unexpected findings from this task, it was decided that the task would not be used in the analysis as had been planned. The task was therefore also abandoned from T2 data collection.

2.4. Procedure

2.4.1. Time 1: Overall procedure

At T1 a total of three visits were made to each child. The first visit was a meeting with the parents to discuss the study, provide the information leaflets (see appendix A, B and C) and to gain informed written consent for the children to participate in the research (see appendix D and E). The purpose of the second and third visits was to collect data for the study. Participants (see chapter 3) were visited twice either in their homes or at nursery. Each visit lasted approximately 40-45 minutes. One child was seen in the SLT clinic, as this was most convenient for his mother.

In all cases, a quiet space was identified, away from the distraction of other children (peers in nursery or siblings). Some parents chose to sit beside, or in the same room as the children, while others chose not to be present. For two of the children seen in nursery, their nursery 'keyworker' opted to sit in and observe the research. This meant that each child did not experience exactly the same conditions for the tests. However, given the young ages of the participants, it was felt that some of the children would benefit from having their parent(s) with them. The reason for not having all the parents sit with their children during testing was either because parents were unable to be present (due, for example, to work commitments), or because they stated that their child would "behave better" without them being there.

In terms of generalisability, testing conditions in the present study were similar though did not completely replicate those typically experienced (in the researcher's experience) by children when they are being assessed for consideration over acceptance onto a SLT caseload. Assessments of this kind may take place in a setting that is familiar for the child, e.g. their home or nursery, or they may take place in an unfamiliar therapy clinic environment. The assessments were led by the researcher who had 6 ½ years' experience working as a therapist with pre-school children at T1 and 8 ½ years' experience by the end of data collection T2.

2.4.2. Maintaining the children's attention and motivation

The researcher drew up a picture timetable for each child to help him/her to understand the order and content of the tasks to be completed. This picture timetable method is often used in speech therapy clinic and is widely believed to help maintain children's attention to the tasks and their motivation. The researcher talked through the timetable with each child, explaining each picture. These line drawings depicted two laptops, described as "the computer games", a tower of blocks, described as "the blocks game", a puppet, described as "the puppet game", a picture book, described as "some pictures" and a pot of bubbles, described as "bubbles". For an example of the time-table, please see appendix G. Each picture was crossed out, either by the researcher or the child (when the child expressed a desire to do this) on completion of the corresponding task.

2.4.3. Counterbalancing the tasks and stimuli

Attempts were made to counterbalance the tasks so that half the children completed the naming task first and the other half completed the repetition task first. This was because testing involved the same word stimuli in the naming (computer game) and the repetition (puppet game) tasks. By counter-balancing the tasks, the study aimed to minimise possible practice effects and fatigue effects.

The naming and repetition tasks were also separated by one of the non-verbal tasks, usually the blocks task, but occasionally the picture task. In addition to their value in contributing to information about the child's non-verbal skills, these tasks served as distractors to minimise practice or familiarity effects of the word stimuli.

Children were assigned either to ‘task order A’ or ‘task order B’, as shown in the table 2-10.

Table 2-10 Different possible order of tasks at the first visit (T1)

T1: Visit 1, Task Order A						
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Description (for the child)	Computer game	Blocks game	Puppet game	Bubbles	Picture game	Another computer game
Test name	Naming task	BAS block building	Repetition tests and span test	Distractor /reward game	BAS picture similarities test	Recognition test
T1: Visit 1, Task Order B						
	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
Description (for the child)	Puppet game	Blocks game	Computer game	Bubbles	Picture game	Another computer game
Test name	Repetition tests and span test	BAS block building	Naming test	Distractor /reward game	BAS picture similarities test	Recognition test

During testing, there were however occasional deviations from the task orders set out above. For example, where a child was initially reluctant to talk, testing began either with the block building task or the picture similarities task. These were selected because the tasks put no pressure on the child to contribute verbally. As the child became more familiar with the researcher, he/she generally appeared more willing and motivated to engage in the repetition and naming tests.

In addition to the counterbalancing of tasks, picture stimuli were counterbalanced. As described before (section 2.2.4), there were two pictures selected for each of the word stimuli, for use in the naming and comprehension tasks. These were arranged into set 1 and set 2 stimuli. To ensure that there were no effects due to the possibility of one set being easier to recognise than the other, half of the participants received set 1 in the naming task and set 2 in the comprehension task, while the other half received the reverse stimuli groupings (i.e. set 2 in the naming task and set 1 in the comprehension task). To further counterbalance, attempts were made for equal numbers of participants receiving order A and order B (described previously) to complete naming set 1 and comprehension set 2, and vice-versa.

Unfortunately this carefully planned counterbalancing was not achieved, as many of the children refused to participate in the repetition task before the naming task. Table 2-11 presents actual figures at T1 for children receiving each order of presentation and each naming/recognition set.

Table 2-11 Task order and stimulus set

	Order A, Naming set 1	**Order A, naming set 2**	**Order B, naming set 1**	**Order B, naming set 2**
Clinical group	9	9	4	5
Non-clinical group	9	8	5	5
Total	**18**	**17**	**9**	**10**

During the second visit, the researcher again drew up a picture timetable list of activities to be completed. This showed a picture book, described as “a picture game”, a laptop, described as “a computer game”, another picture book, described as “another picture game, where you need to do some talking” and bubbles. The order of tasks during the second visit is shown in table 2-12.

Table 2-12 Task order at the second visit

T1: Visit 2, Order of tasks				
	[redacted]	[redacted]	[redacted]	[redacted]
Description for the child	Picture game	Bubbles	Another picture game	Computer game
Test/ task name	Pre-school Language Scales - auditory	Bubbles (distractor/ reward game)	Pre-school Language Scales - expressive	Naming task

2.4.4. Time 2: Overall procedure

Children were visited again a year after the initial data collection visit. They were visited either in their homes or at nursery or school. As with T1, there were two data collection visits. Some parents opted to sit in the same room as their children while others opted not to be present. A quiet space was identified and each visit lasted between 45 and 60 minutes.

Visits resembled those of T1. The first visit was identically structured, but with the addition of the *Word Structure* sub-test from the CELF-Preschool 2 (Wiig et al., 2004) and the removal of the ‘recognition test’ used in T1 (see section 2.3.4). The second visit was identical to the T1 second visit, but additionally included the *Sentence Imitation Test* (SIT) from the ERB (Seeff-Gabriel et al., 2008).

2.4.5. Individual task procedures

The procedures for administering the tasks designed specifically for the present study will be discussed in turn, followed by the procedures for administering the standardised tests.

2.4.5.1. Naming task

The researcher informed the child that he/she was going to play a computer game. The researcher explained that pictures would pop up on the screen and the child should try to name them. The researcher told the child that some of the pictures might be tricky and that it was ok for the child to guess the tricky ones or it was also ok for the child to say that he/she was not sure if the child did not know the object's name.

Twenty-eight pictures were presented to each child in a PowerPoint presentation format on a laptop which had a screen size 22.5cm x 13cm. Pictures were presented one at a time and the child was asked either to answer the question "what is this?" or to complete a sentence, e.g. "they're jumping on a ….". Each picture entered onto the screen with a different animation to maintain the child's interest and to keep his/her attention and motivation. All items were shown to all the children. The task was video recorded for later phonetic transcription. The researcher checked for consent to video from both the parent (written consent) and the child (verbal consent). In one case video consent was not granted by the parent and instead the task was audio-recorded.

Participants completed the naming task for the second time on the second research visit. As discussed (section 2.3.2), the purpose of the second presentation was to assess for phonological consistency. Assessing phonological consistency over two assessment sessions is not typical for tests of phonological consistency (e.g. in the Diagnostic Evaluation of Articulation and Phonology, Dodd, Hua, Crosbie, Holm and Ozanne, 2002), instead multiple productions of the same word are elicited in one session. The reason for the different procedural approach adopted in the present study was to maintain motivation by the very young children taking part in the study.

During testing, the children were given positive feedback. When a child named the item correctly, the researcher responded “good”, “well done”, “uh huh” (with rising intonation) or “that’s right”. When a child gave an incorrect response, the researcher commented, “it does look a bit like that, doesn’t it”. When a child was unable to give a response, the researcher commented “yes, that is a tricky one, isn’t it. We’ll hear what it is called later”. In the instance of a child being unable to give a response, the researcher did not provide the child with the answer on the first presentation, as this might affect their later repetition of the word, their recognition of the word or their second attempt at naming the item. The target label was given to the child on the second presentation, as this would not affect their future performance on any of the tasks. In the event of a child giving a plausible response, but not the target word (e.g. "foot" for *toe*), the researcher acknowledged that the child's response was correct and asked for further specification. For example, for the target *toe,* the researcher pointed to a toe on the picture and said "yes, that's a foot and this is a..."; in the case of the child saying "T-rex" for dinosaur, the researcher said "yes, and a T-rex is a kind of...". If they did not then supply the target word their response was scored as incorrect.

2.4.5.2. Repetition tasks

The word and non-word repetition tasks each involved the presentation of 28 stimulus items, preceded in each case by a practice item. The procedure adhered closely to that of Seeff-Gabriel et al. (2007). Consistent with Seeff-Gabriel et al. (2007), the items were presented in order of increasing length (i.e. one-syllable items first, followed by two-syllable items, etc.) and the complete set of words was administered before the administration of the non-words.

Words were presented one at a time and children were only permitted a repetition where they had clearly been distracted by an outside stimulus during the first presentation. This procedure differed from Seeff-Gabriel et al. (2007) where repetitions are allowed. It is common in short-term memory research not to allow repetitions of the stimuli, as external repetitions may likely serve to update and rehearse a possible fuzzy representation that is already stored in short-term memory from the initial item presentation. This provided the rationale for not allowing repetitions in the present study.

Hand puppets were used in the presentation of the stimuli to increase motivation and compliance: one for presentation of the words and another for the non-words. The children were introduced to a hand puppet with a moveable mouth. They were told that the puppet had some words that he would say and that he wanted to hear the child say the same words after he had said them. It was explained that the puppet squeaks when he is happy and the child could try to make him squeak by saying his words.

Figure 2-1 Puppet used in the word repetition and span tasks

[redacted]uitar”

The non-word repetition task followed a similar pattern of procedure to that of the word repetition task. The child was introduced to a new puppet that had a moveable mouth and was told that this puppet comes from a different Country and that he speaks a ‘funny language”. The child was told that the puppet would like to know if the child could also say his “funny words”.

Figure 2-2 Puppet used in the non-word repetition task

[redacted]

“trangkeroo”

As is common for repetition research involving children of this age (e.g. Roy and Chiat, 2004; Chiat and Roy, 2007), stimuli were presented live. While this had the disadvantage of children hearing slightly different acoustic versions of the stimuli and with unequal intervals between stimuli, it had the distinct benefit of maintaining a level of ‘naturalness’ and therefore increasing motivation and compliance by the young children.

The task was video-recorded to enable later broad phonetic transcription and scoring. As for the naming task, the researcher checked for consent to video both from the parent (written consent) and the child (verbal consent). Where a parent had not consented to video-recording (one participant), audio-recording was used instead.

Reliability of scoring for the repetition tasks

Children's responses on the naming and repetition tasks were video-recorded and later transcribed phonemically by the researcher. Reliability of the researcher's transcriptions was calculated using a sample of 7 of the videos (13% of the sample) that were randomly selected. These were additionally transcribed by four speech and language therapy 3rd year undergraduate students, who were trained and proficient in phonetic transcription. The two naming tasks, and the word and non-word repetition tasks from the T1 data collection, were transcribed during this process, yielding 28 separate task scores. Overall the correlation between scores calculated by the researcher and those computed by the second scorers was $r(27)=0.957$, $p<0.001$. Individual task correlations are given in table 2-13.

Table 2-13 Inter-rater reliability correlation coefficients

	Correlation
Naming task 1	$r(6)=0.97$, $p<0.001$
Naming task 2	$r(6)=0.98$, $p<0.001$
Word repetition	$r(6)=0.98$, $p<0.001$
Non-word repetition	$r(6)=0.92$, $p<0.005$

2.4.5.3. Span task

The span task aimed to assess children's ability to hold unrelated words in PSTM by asking them to repeat two-, three-, four- and five- (T2 only) word sequences drawn from a closed set bank of ten one-syllable words.

Following the word and non-word repetition tasks, children were shown the original puppet again. They were told that the puppet wondered whether they could remember two (or more) words. They were asked to listen to the puppet's words and then to repeat both the words that they had heard after the puppet.

Similar to the word and non-word repetition tasks, the span task was presented live. This procedure had the same advantages and disadvantage as the word and non-word repetition tasks. Participants completed five trials at each span length (two-, three-, four- word sequences, and additionally five-word sequences at T2).

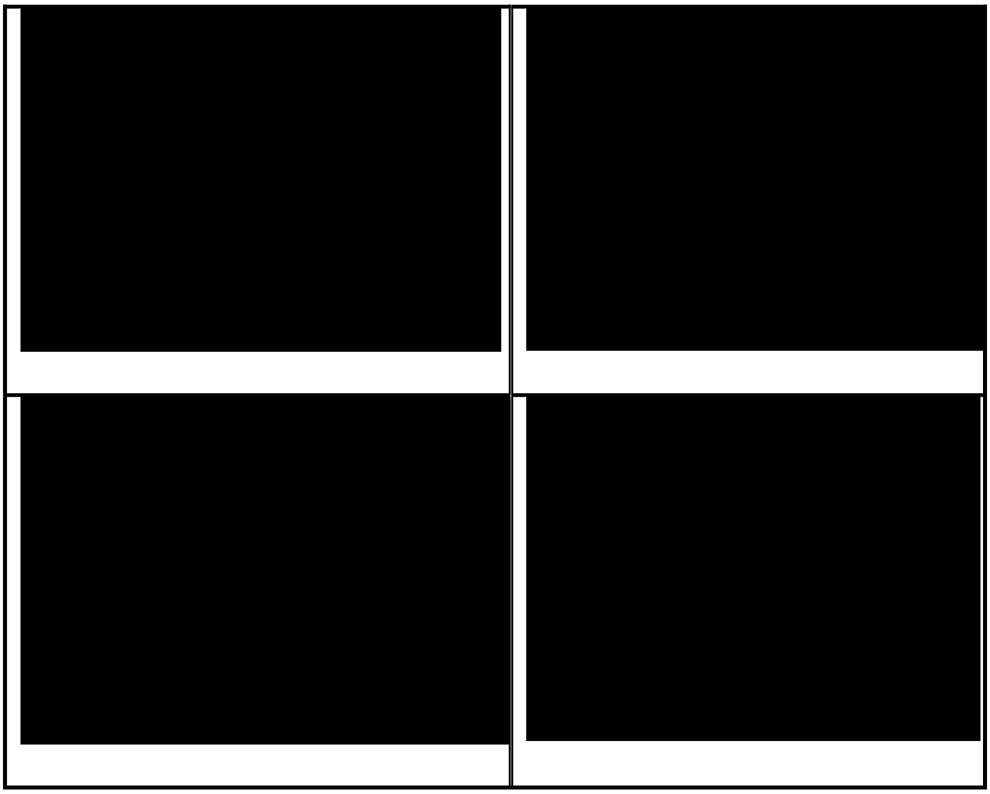
The words used in the span task were drawn from the bank of single syllable words used in the naming, recognition and repetition tasks. This ensured that data was available about whether the child knew the words and about how he/she articulated the words when his/her PSTM system was considered to be minimally stressed, i.e. when only repeating one word at a time. This allowed responses during the span task to be compared with those during naming and repetition. The same sequences of words were presented to all the children (see 2.2.5).

2.4.5.4. Recognition task (T1 only)

Target stimuli were presented together with three distractor pictures. The participants were told that they would play another computer game. In this game four pictures would pop up on the computer screen and they needed to try to find the picture that the researcher named.

As previously discussed, children who had been presented with picture set A for the naming task were presented with picture set B for the comprehension task. These picture sets (set A or set B) were combined with the same distractor pictures, i.e. all children received the same distractors, in the same positions on the screen regardless of whether they were presented with picture set A or B. Stimuli were presented one at a time with the three distractor pictures in a 2x2 grid (see figure 2-3). The position of the target item had been ‘randomly’ assigned to one of the possible positions. The task was presented to the child on a laptop computer (screen size 22.5cm x 13cm) using PowerPoint. The children were asked to point to the target picture with either “where’s the [e.g. car]” or “show me the [e.g. car]”.

Figure 2-3 Example of stimuli presented for the recognition task



2.4.5.5. Standardised assessment tasks

<u>Pre-School Language Scales</u>

<u>i) Auditory</u>

The children were told that they were going to look at some pictures, and complete a ‘listening and pointing game’. The procedure for administration then followed that described in the PLS4 manual (Zimmerman et al. 2002) for the auditory part of the test.

ii) Expressive

The children were told that they were going to answer some questions and look at some pictures. They were told that this picture game would need them to do some talking. The procedure for administration then followed that described in the PLS4 manual (Zimmerman et al. 2002) for the expressive part of the test.

British Ability Scales

i) Picture Similarities Test

The procedure for this test adhered to that described in the BAS II manual (Elliott et al., 1996). Children were presented with four pictures on an A4 page from the picture book. They were required to match a line drawing picture, presented on a small card, to the corresponding picture in the manual.

ii) Block-Building Test

The procedure for this task followed that set out in the British-Ability Scales II manual (Elliott et al., 1996). The children were asked to recreate the wooden block designs that were first completed by the researcher. The children were able to see the target pattern at all times while recreating their matching block design.

Clinical Evaluation of Language Fundamentals –Pre-School (T2 only)

Word Structure Subtest

The procedure for this task followed that described in the CELF-PS2 manual (Wiig et al. 2004). The children were asked to look at the pictures and to complete the sentence started by the researcher.

Early Repetition Battery (T2 only)

Sentence Imitation Task

The children were told that they would participate in “another puppet game”, but that this time the puppet would say some sentences and the children needed to repeat these. The procedure then adhered to that set out in the instruction manual (Seeff-Gabriel et al., 2007).

2.4.6. End of visit

At the end of each visit, all children were thanked for helping the researcher with her work and were given specific praise for an aspect of their behaviour. They were also offered a sticker. At the end of the first visit at each time point, the researcher asked the children if it would be ok for the researcher to come back to do some more ‘games’ with the child. All the children agreed to this, and the researcher arranged the follow-up visit with the parent.

Chapter 3

Participants and overview of assessment scores

3.1. Chapter overview

This chapter presents information about data relating to the population sample included in the study. It includes information on the methods of recruitment; information about the participants' age and gender and discusses any effects of these variables on the children's performance. It then presents scores on the standardised measures of language and non-verbal ability that were not the main focus of the study. It also provides an overview of the distribution of scores on the tasks that were the focus of the study.

3.2. Participants

The study aimed to recruit between sixty and eighty participants aged 3 years, 0 months to 3 years 6 months. Approximately half of the sample would be known to speech and language therapy services, while the other half would not have any known speech or language difficulties. The intention was to establish a population sample that was rich in children with a range of speech and language abilities, so that findings could be generalised to the full range of children that might be encountered in Speech and Language Therapy (SLT) clinics. A secondary aim was to compare the two groups of children: those who have speech and/or language difficulties and those who do not. The purpose of this was to investigate whether there were qualitative and quantitative differences between the groups.

NHS ethics approval was sought and granted in December 2007 to recruit children through the NHS. Following this, City University Senate Ethics approval was sought and granted in January 2008. Extensions to both of these approvals were granted (December 2011) following suspension of the study. Recruitment began following approval of extension of the study in January 2012.

Children were recruited who spoke English only, or, where children were bi-/multilingual, where English was their strongest language. Children were excluded from the study if they had a known diagnosis of Autistic Spectrum Disorder or a known global developmental delay. Following these criteria, fifty-nine children were recruited to take part in the study. Full data were collected for fifty-four of these children at T1 of the data collection.

3.2.1. Recruitment of participants through NHS SLT Clinics

For recruitment of the clinical group, the researcher approached all speech and language therapists working in the Community Early Years (CEYs) clinics within the NHS organisation where she works as a speech and language therapist. The SLT CEYs team is based in central London and currently has seven Early Years clinics. The therapists working in the clinics were informed about the study and asked to identify any children from the respective clinic caseloads who matched the inclusion criteria. Therapists were asked to inform the parents about the study when meeting with the children's parents for assessment or therapy, and to enquire whether they would like their children to take part. If so, the parents' contact details were given to the researcher, who contacted the parents by telephone to discuss the study.

Thirty-one children were recruited through the NHS. All the participants recruited through the NHS were identified from three of the seven clinics. Most of these children (27 participants) were recruited from the clinic in which the researcher worked at the time of data collection. The clinic is situated within a built-up council estate and it serves a socially diverse population in central London. A further four participants were recruited from two of the other SLT clinics within the NHS organisation. These clinics also serve populations that are socially diverse.

All children recruited through the NHS were identified as having speech and/or language difficulties. These children were at different stages of their SLT care. Some had received an assessment and were on the waiting list for therapy, while others were receiving therapy or had recently received therapy. Some children were discharged from SLT over the course of the project either for non-attendance in therapy, or because they had made progress in their communication skills and their speech and/or language difficulties were considered to have resolved to a level appropriate for their age. These children were however still included at the second time point of the study, where it was possible to make contact with them. Where the groups were analysed separately, these children were still considered to form part of the clinical group.

No parent was coerced into agreeing for their child to take part in the study and parents were made aware that participation or non-participation in the study would not affect their child's SLT care. All parents were given detailed information about the tasks involved in the study and they all gave their informed written consent.

The tasks were also explained to the children, using simple language and hand-drawn pictures. All children participating in the study gave their verbal consent to take part.

Parents also completed a questionnaire with the researcher, which included questions regarding their own profession, level of education, languages spoken to and by their children (see appendix F). The researcher also asked whether the child currently had or had a history of glue ear and this was noted on each questionnaire.

3.2.2. Recruitment of the non-clinical group

Children who were not known to SLT were recruited through nurseries local to the NHS organisation. The researcher contacted nurseries to enquire whether they would be willing to approach parents of appropriately-aged children in their nurseries about the study. Parents were given brief written information about the study and a consent-to-contact form on which they filled in their contact details. The nurseries returned these forms to the researcher who then contacted the parents to discuss the study. Children recruited in this way were socio-economically diverse.

Fourteen of the children were recruited through friends, colleagues and relatives. All of these children lived within the Greater London area. These children were not from diverse socio-economic groups: all of the children recruited in this way had one or both parents with a University education. A total of twenty-seven children were recruited through nurseries, friends, family and colleagues.

All of the parents in the non-clinical group gave their informed written consent and they also completed the questionnaire described in 3.2.1 (appendix F).

Two of the children recruited as part of the non-clinical group were referred to SLT during their involvement in the study (one for a speech disorder and one for language difficulties). Parents of a further two children were advised that a referral to SLT would be appropriate (one for language difficulties and one for a stammer), but they opted not to make a referral.

3.2.3. Withdrawals from the study at T1

Five children were withdrawn from the study at T1. Two children were unwell on their allocated days of testing, and were too old for the study by the time they had recovered from their respective illnesses. One child was heard to be stammering on the first visit and the researcher felt that extensive formal language testing would not have been beneficial and might even have been detrimental for the child. One child refused to participate in several of the tests, so data could not be reliably used. One child spoke in a whisper for all of the language tests and could not be encouraged to use a louder voice.

Therefore full data were collected from fifty-four participants, twenty-eight of whom were known to SLT and twenty-six of whom were not.

In subsequent sections the population is described in more detail with respect to characteristics that might conceivably affect task performance, e.g. age, gender, exposure to other languages, SES.

3.2.4. Age

Attempts were made for all children involved in the study to be aged between 3 years, 0 months and 3 years, 6 months at the first visit in which data was collected. However circumstances beyond the researcher’s control (e.g. incorrect information about the children’s age prior to testing) meant that two of the children fell into ages outside this range. One child was a month younger than the targeted age and one was a month older. Table 3-1 shows the distribution of ages in years and months within the sample at T1.

Table 3-1 Age range of the children at T1

Age (yrs, mths)	2,11	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,7	Total
Child recruited through SLT clinic	0	4	5	4	3	5	4	2	1	**28**
Child not recruited through SLT clinic	1	2	4	6	3	1	7	2	0	**26**
Total	**1**	**6**	**9**	**10**	**6**	**6**	**11**	**4**	**1**	**54**

3.2.5. Gender

A Chi-Square test revealed a significant effect of group by gender ($\chi 2$=5.56, $p<0.05$) i.e. more boys were represented in the clinical sample than girls. Gender distribution figures are given in table 3-2.

Table 3-2 Distribution of female and male participants in the clinical and non-clinical recruitment groups

	Male	Female	Total
Child recruited through SLT clinic	20	8	28
Child not recruited through SLT clinic	12	14	26
Total	32	22	54

The significant difference in gender distribution between the groups is unsurprising, since it is known that developmental speech and language difficulties are more prevalent among boys than girls. For example, Robinson (1991) reports a male:female ratio of 3.8:1 (although this sample included some speech and language difficulties associated with other conditions, e.g. cleft palate). Tomblin et al. (1997) report a smaller ratio of 1.5:1 between females and males (9% of boys and 6% of girls) in their large sample of pre-school children in the USA.

Previous studies in similar aged children have shown that NWR tests are not sensitive to gender (Chiat and Roy, 2007), so the difference in the clinical group compared with the non-clinical group was not considered to be problematic.

3.2.6. Language experience

The children taking part in the study were required to have English as their only or main language. For children who were bi- or multi-lingual, 'main' language was defined as being the language that the child was considered by his/her parent(s) to be the strongest speaking. Additionally, in order to meet the inclusion criteria, at least one of the parents needed to be speaking English only with the child. This was because it is known that linguistic factors that are distinct in different languages affect children's repetition (see chapter 1, section 5), although less so than for vocabulary tests (Thordardottir and Brandeker, 2013). It was however considered appropriate to include children with additional languages, as multilingualism is common in London (a DfES report from January 2005 states that 42.9% of primary age children have English as an additional language in the London borough in which the data was primarily collected). Therefore, inclusion of pupils speaking additional languages to English arguably increases the generalisability of the findings from the study to the population. Table 3-3 shows the language experience of the children who took part in the study.

Table 3-3 Participants' language experience

Language	**Group**	English only	English & French	English & Greek	English & Arabic	English & Tigrigna	English & Polish	English & Czech	English, Arabic & French	**Total**
Number of children	**Clinical**	24	3	0	0	1	0	0	0	**28**
	Non-clinical	20	0	1	1	0	2	1	1	**26**
	Total	44	3	1	1	1	2	1	1	**54**

Many children this age spend a significant amount of their waking time in nursery, which would also influence their language exposure. Of the 54 children for whom full data was collected at T1 of the study, three were not attending nursery. All other participants were attending English-speaking nurseries either part-time or full-time. The three children who were not attending nurseries at T1 were all exposed only to English at home.

3.2.7. Participation by twins in the study

There was an unusually high number of twins who took part in the study. Of the 54 children for whom full data was collected, twelve children came from a set of twins. There were two sets of identical (monozygotic) twins and four sets of fraternal (di-zygotic) twins. Four of the children (from three sets of twins) were known to speech and language therapy, the other eight children did not have identified speech and/or language needs, though two were siblings of children with identified speech and/or language needs. This high prevalence of twins makes the sample unusual. According to the Multiple Births Foundation, the incidence of twins is fifteen in every thousand births. It is known that twins are at a greater risk of language delay than singletons. For example, Rutter, Thorpe, Greenwood, Northstone and Golding (2003) and Thorpe, Rutter and Greenwood (2003) report that the large proportion of twins in their study were on average 3.1 months behind their singleton peers at 3 years old. Thorpe et al. (2003) report differences in the parent-child interaction behaviours of mothers towards twins compared to singletons, e.g. encouraging the children to speak, pointing out features of interest, elaborating upon the children's comments. They report a relationship between these interaction styles and language skills of the twins. However most of the twins in the present study's sample did not have any identified language difficulties.

The fact that twins were over-represented in the sample might present a limitation in terms of the generalisability of the study's findings.

3.2.8. Hearing status

None of the children who took part in the study had a neurological hearing impairment. Three of the children were known to have intermittent glue ear. It is known that glue ear is common in pre-school children, affecting between 10 and 30% of children aged between 1 and 3 years old (Lous, Burton, Felding, Oveson, Rovers and Williamson, 2005). It is also known that children who have glue ear are more at risk of language difficulties (e.g. Maw, Wilks, Harvey, Peters and Harvey, 1999). Following the same inclusion criteria as that used by Chiat and Roy (2007) and because the children's glue ear was known to be intermittent, it was decided that these children would not be excluded from the study.

3.2.9. Social and educational background

In order to ensure that the findings from the study were representative of the population, an attempt was made to recruit children from diverse socio-economic groups. Obtaining an accurate measurement of socio-economic status (SES) is difficult. Roy and Chiat (2013) describe that 'low SES' is a relative term that has different meanings depending on the information gathered. Typically information is sought on occupation of the main or both care-givers, levels of education and income. However these measures are significantly correlated (Roy and Chiat, 2013) and parents have been found to be more willing to disclose information about their education level and their occupation than their income (Noble, McCandliss, and Farah, 2007). If SES is defined in these ways, it is also not necessarily a static term, for example levels of education, occupation and income can change throughout an individual's life, facilitating social mobility.

In the present study, information was gathered about the main carer's highest level of educational qualification and also about their occupation. However, only information about education is reported, for three reasons. First, these measures are known to be highly correlated (Hart and Risley, 1995), second Tomblin et al. (1997) found that language ability correlated with parental education, and third it was simpler to categorise educational level than it was to categorise occupation. The researcher did not think herself sufficiently informed about the nature of the different professions/job titles and the skills and experiences required for each role to make a judgement about how to categorise these. This rendered occupation uninformative. Table 3-4 shows a summary of data obtained about educational qualifications.

Table 3-4 Main carer's highest educational qualification and comparison with 2011 census

	No formal qualifications	GCSEs or equiva-lent	A-levels or equiva-lent	Degree level or above	Other qualification (please specify)	No response
Number (SLT group)	3	3	3	12	2	5
Non-SLT group	0	1	1	20	1	3
Overall group percentage	5%	7.4%	7.4%	59.3%	5%	14.8%
Specific London borough census 2011	17.9%	21%	12%	45.1%	4.1%	-
England and Wales census 2011	22.7%	28.6%	12.3%	27.2%	9.3%	-

As evident in table 3-4, parents with no formal qualifications and with GCSEs or equivalent as their highest level of education are unfortunately under-represented in the study's sampling. It is known that membership of a low SES group is associated with poor vocabulary. For example Hart and Risley (1995) found that found that toddlers in low SES groups heard about a quarter of the number of words compared to their peers from professional families. However, the measures investigated in the present study (word repetition, non-word repetition and sentence repetition) are thought to be relatively unrelated to measures of SES, e.g. Burt, Holm and Dodd (1999). This is because they are arguably thought to be measures of processing, rather than measures that tap cultural or existing language knowledge (Roy and Chiat, 2013). Therefore the under-representation of the groups with lower educational qualifications was not considered to be problematic.

3.2.10. Time 2 recruitment

All parents were contacted approximately 11 months after their original participation in the study. Children who had been known to SLT were contacted regardless of whether they were still known to the SLT service. Parents were asked whether they would still like for their child to take part in the study by participating in a follow-up visit. Following verbal consent, children were visited twice, either in their home or, if agreed, at nursery or school.

A total of 52 children were visited for follow up assessment at T2. Two children were lost from the sample. This was due to one child and his family moving overseas and to the other child's parents being non-contactable. Both of these children were male (one monolingual, one bi-lingual in Tigrigna and English) and had been recruited through SLT clinic. Table 3-5 shows the proportion of children (divided into boys and girls) recruited through the NHS and those recruited through other methods (e.g. through local nurseries) at T2.

Table 3-5 Gender distribution across the groups at time 2

Recruitment group	Male	Female	Total
NHS SLT clinic	18	8	**26**
No identified speech/language needs	12	14	**26**
Total	**30**	**22**	**52**

A Chi-Square test revealed a significant effect of group by gender ($\chi 2=4.00$, $p<0.05$). As for T1, more boys than girls were represented in the clinical sample.

The study aimed to visit the children at T2 exactly a year after their first T1 visit. It was not possible to arrange follow-up visits exactly 12 months after the first visit for all the children (due for example to holidays, illness, availability). Table 3-6 shows the number of months between visits for the participants.

Table 3-6 Number of months between Time 1 and 2

No. of full months between phase one and two	11	12	13	14	15	Total
Number of participants	10	29	6	5	2	**52**

Table 3-7 shows the ages of the children seen at T2.

Table 3-7 Age of participants at Time 2

Age (yrs, mths)	4, 0	4, 1	4, 2	4, 3	4, 4	4, 5	4, 6	4, 7	Total
Child recruited through SLT clinic	4	4	1	3	7	3	2	2	**26**
Child not recruited through SLT clinic	3	7	1	2	1	7	2	3	**26**
Total	**7**	**11**	**2**	**5**	**8**	**10**	**4**	**5**	**52**

3.3. Distribution of scores

The following section presents score distributions for the tasks. Distributions are presented in box plots for the clinical (blue) and non-clinical (orange/yellow) groups separately and these are also combined (green) to give overall distributions. T1 data has dark colours and T2 data has light colours. Descriptive statistics for each of the tests are also provided in appendix H, and information regarding normality of the distribution of the scores is presented in appendix I. The data is presented both as separate groups and as a combined sample to reflect the way in which the data was analysed. Chapter 5 separated the sample into clinical and non-clinical groups. Data analysis in chapters 4, 6 and 7 combined the two groups to give a language-impaired enriched sample.

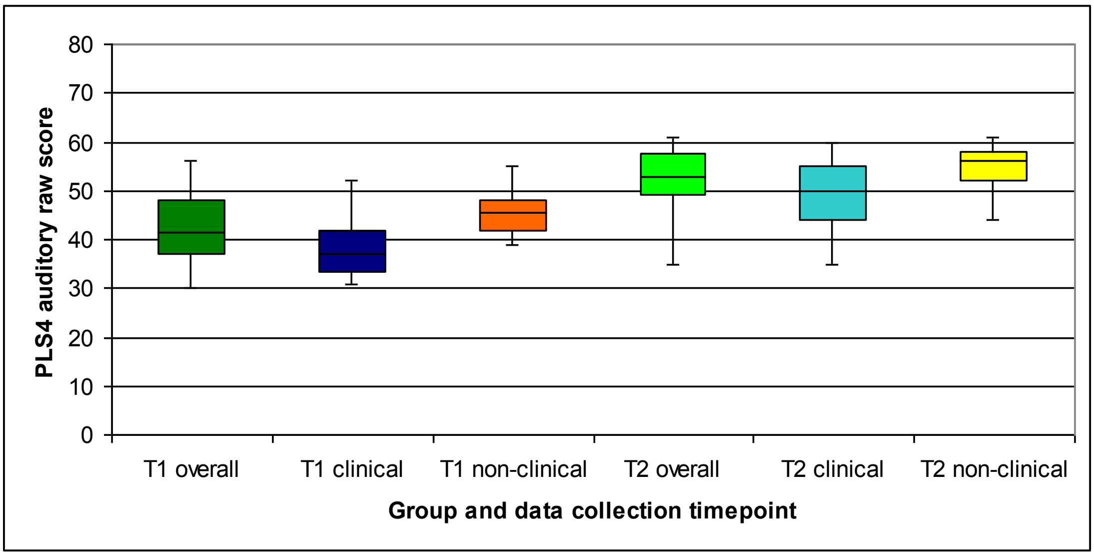
The tests of normality provided in appendix I indicate that most of the test scores were not normally distributed. Data transformations were attempted but did not correct the normality of the data for most of the tests. Non-normality of the data violates one of the assumptions of parametric statistical tests. However, the t-test is known to be robust and therefore “relatively insensitive to violations of its underlying mathematical assumptions” (Pagano, 2010, p363). Pagano (2010) explains that this is particularly true for sample sizes greater than 30. In this study, given the size of the combined sample size of 54 at T1 and 52 at T2, parametric tests were used to analyse the data, however non-parametric test results are also given in appendix J or in the main text where the non-parametric tests altered the results. Where the data was split into clinical and non-clinical groups (chapter 5), sample sizes approached 30 for the groups (28 in the clinical group and 26 in the non-clinical group at T1; 26 in both groups at T2).

3.3.1. PLS4 scores

Figures 3-1 and 3-2 present the distribution of scores on the PLS4 auditory and expressive tests.

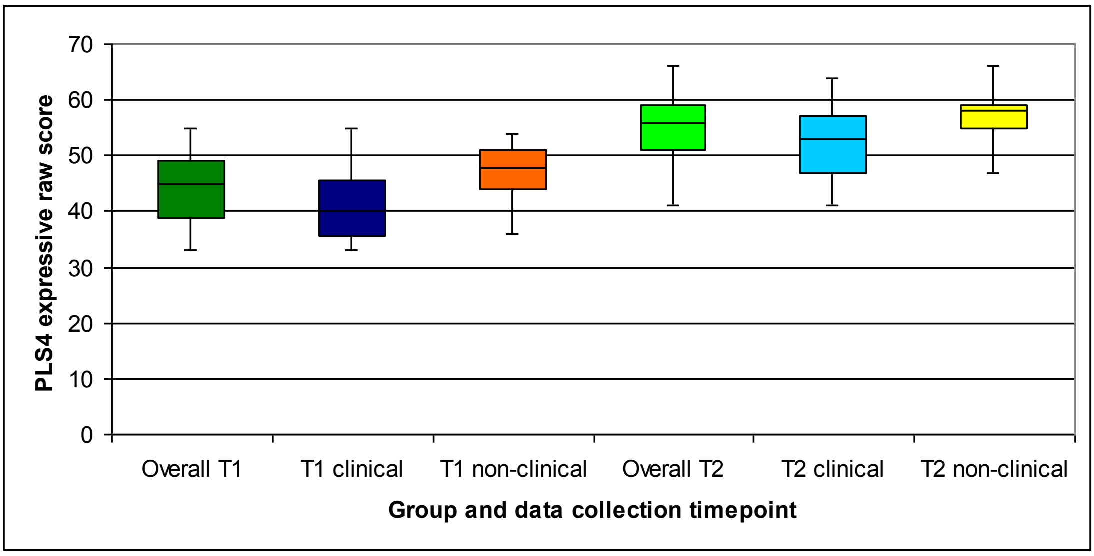
Auditory

Figure 3-1 Distribution of scores on the PLS4 Auditory Test



Expressive

Figure 3-2 Distribution of scores on the PLS4 Expressive Test



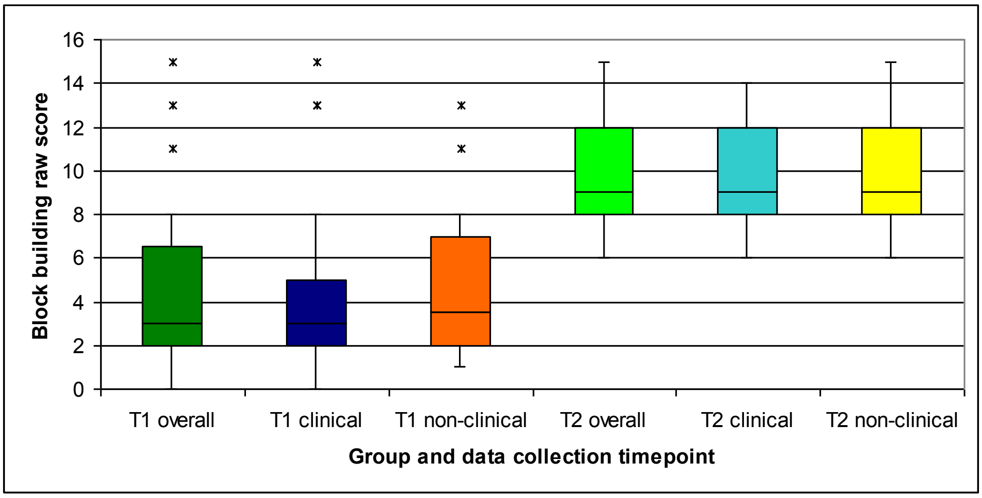
Evident from the graphs above (figures 3-1 and 3-2), there was some overlap between the clinical groups in the scores on the PLS-4 tests. This is due to some of the children in the clinical group having identified speech but no language difficulties. It is also partly due to two children in the non-clinical group scoring lower than expected on the assessment.

3.3.2. BAS scores

Figures 3-3 and 3-4 present the distribution of scores on the BAS sub-tests.

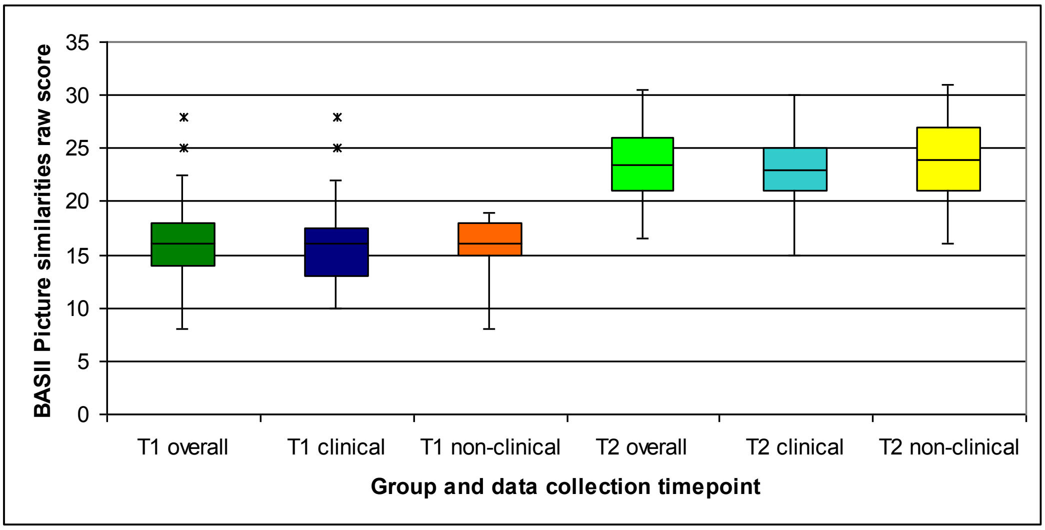
Blocks

Figure 3-3 Distribution of scores on the BAS II Block Building test



Picture similarities

Figure 3-4 Distribution of scores on the BAS Picture Similarities test



Evident from the graphs above (figures 3-3 and 3-4), there was considerable overlap between the group scores on the BAS II subtests at both time-points. T-tests (section 3.4.1) confirm that while the clinical group had language difficulties (apparent in their PLS4 scores), there was no difference between the groups in non-verbal skills.

3.3.3. Naming

Figure 3-5 Distribution of scores on the Naming task (scored for number of phonemes correct)

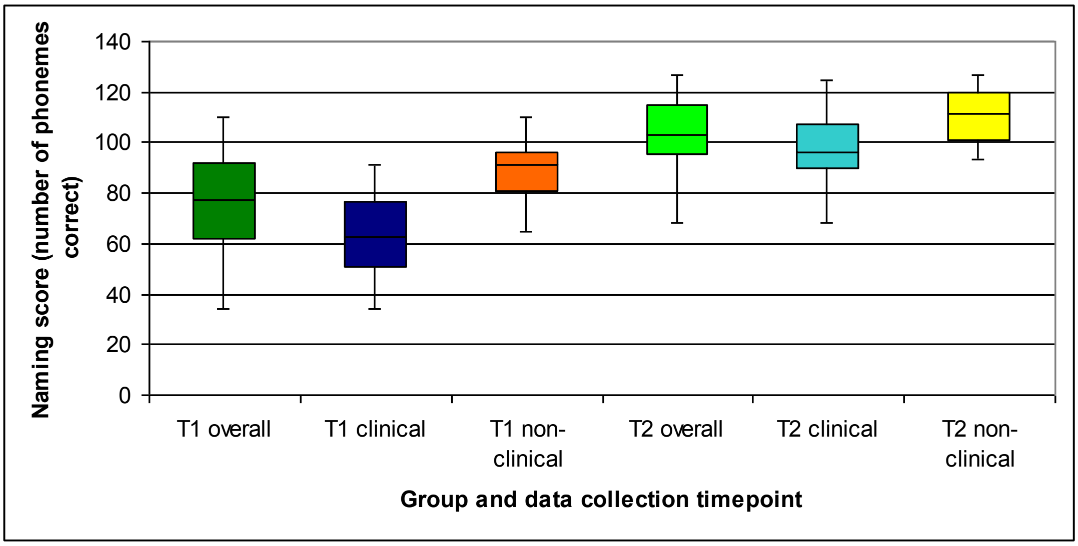
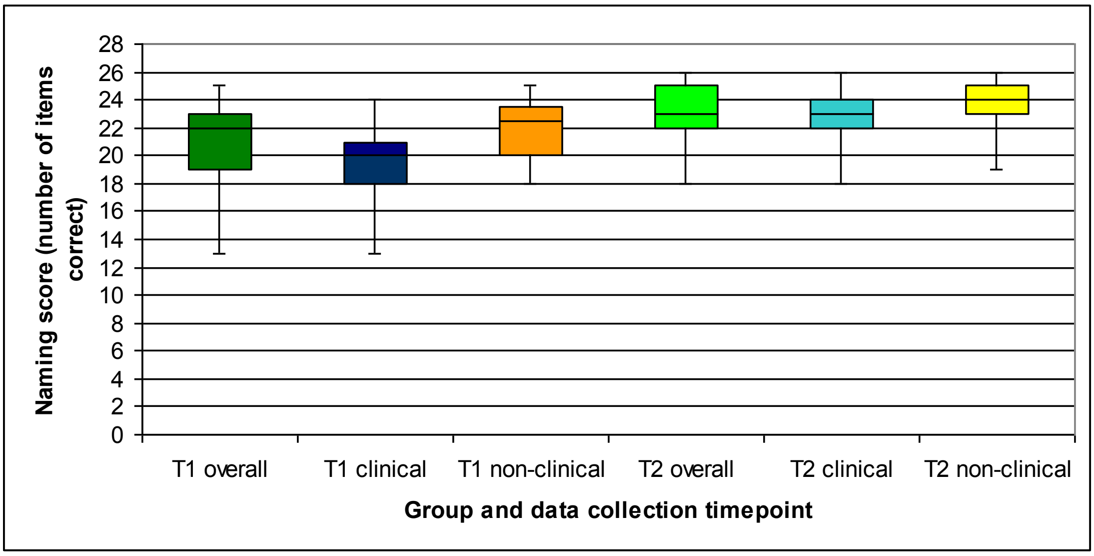


Figure 3-6 Distribution of scores on the naming task (scored for number of items correct)



As indicated in the graphs above (figures 3-5 and 3-6), there was overlap in the scores by the clinical and non-clinical groups. This was particularly evident at T2 and most apparent when the test was scored as items correct, rather than phonemes correct. Furthermore, at time 2 there was a small range of scores.

3.3.4. Word repetition

Figure 3-7 Distribution of scores on the word repetition task (scored for number of phonemes correct)

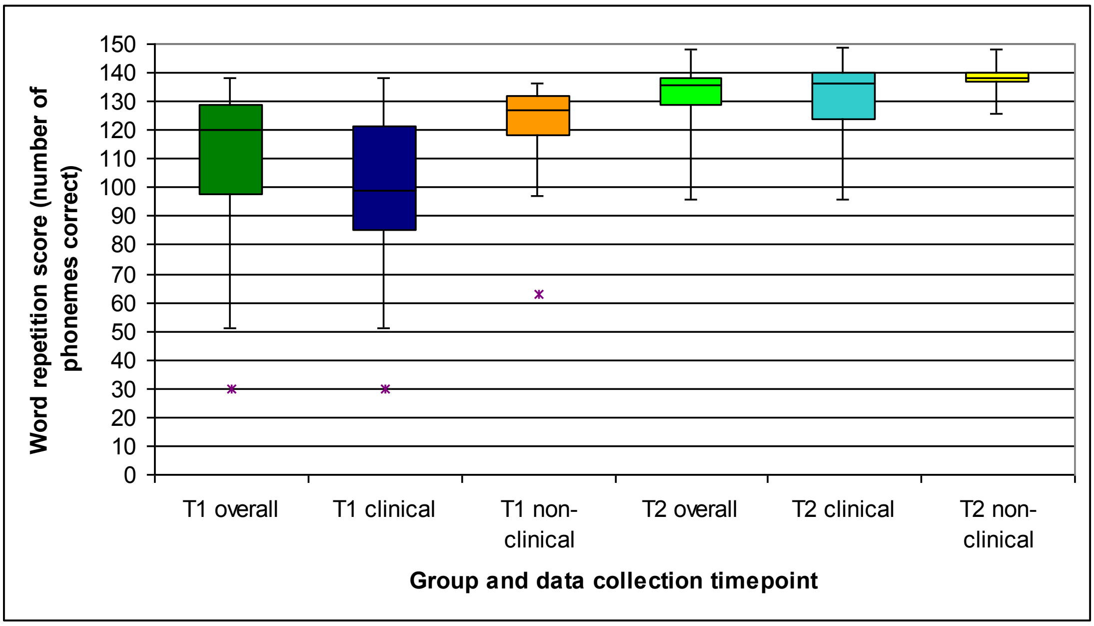
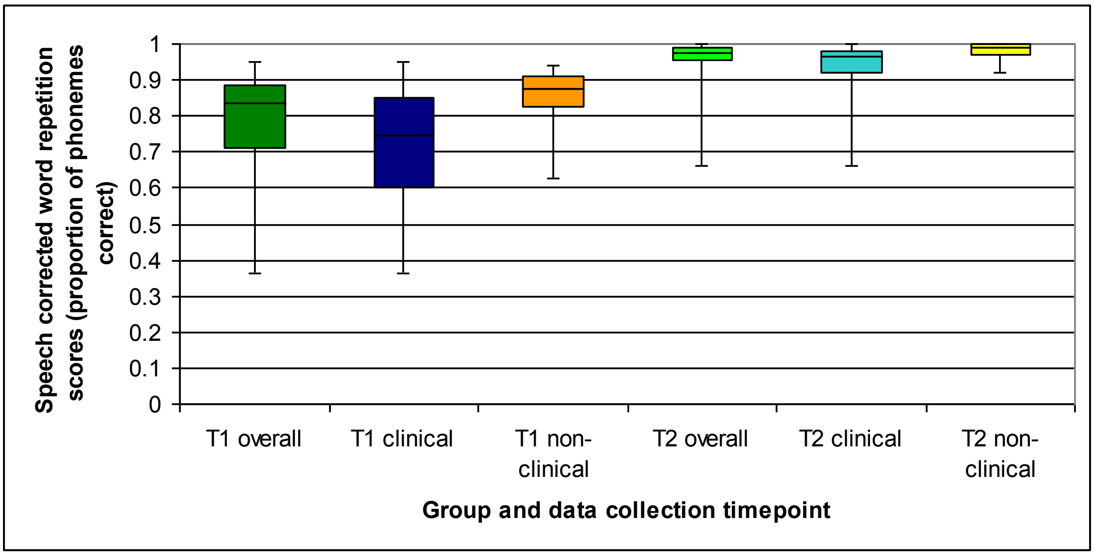


Figure 3-8 Distribution of scores on the speech corrected word repetition task (scored for known words only – proportion of phonemes correct)



As shown in the graphs above (figures 3-7 and 3-8), ceiling effects were evident at T2 for both the clinical and non-clinical groups in word repetition scores. This was particularly the case when scores were corrected for speech errors (figure 3-8).

As discussed previously (section 2.3.2), the word stimuli were split into words that the children knew and those that they did not know. Word knowledge was assessed by whether the children showed an ability to name the item. Table 3-9 shows the mean number of items that were known and the mean number of items that were unknown at T1; table 3-10 shows the mean number of items known and unknown at T2.

Table 3-9 Mean known and unknown words at T1

	1 syllable (max=10)		2 syllable (max=6)		3 syllable (max=6)		4 syllable (max=6)		Total (max=28)	
	known	unknown	known	unknown	known	Unknown	known	unknown	known	unknown
clinical group (mean and range)	8.96 (4-10)	1.04 (0-6)	4.43 (2-6)	1.57 (0-4)	4.29 (3-6)	1.71 (0-3)	1.25 (0-2)	4.75 (4-6)	18.9 (11-24)	9.07 (4-17)
Non-clinical group (mean and range)	9.31 (8-10)	0.69 (0-2)	4.77 (3-6)	1.23 (0-3)	5.15 (4-6)	0.85 (0-2)	2.23 (0-4)	3.77 (2-6)	21.5 (18-25)	6.54 (3-10)

Table 3-10 Mean known and unknown words at T2

	1 syllable (max=10)		2 syllable (max=6)		3 syllable (max=6)		4 syllable (max=6)		Total (max=28)	
	known	unknown	known	unknown	known	Unknown	known	unknown	known	unknown
clinical group (mean and range)	9.54 (9-10)	0.46 (0-1)	5.27 (4-6)	0.73 (0-2)	5.62 (4-6)	0.38 (0-2)	2.35 (0-4)	3.65 (2-6)	22.77 (18-26)	5.33 (2-10)
Non-clinical group (mean and range)	9.69 (7-10)	0.21 (0-3)	5.58 (4-6)	0.42 (0-2)	5.92 (5-6)	0.08 (0-1)	2.81 (2-4)	3.19 (2-4)	24.00 (19-26)	4.00 (2-9)

3.3.5. Non-word repetition

Figures 3-9 and 3-10 present the distribution of scores on the non-word repetition tasks.

Figure 3-9 Distribution of scores on the non-word repetition task (number of phonemes correct)

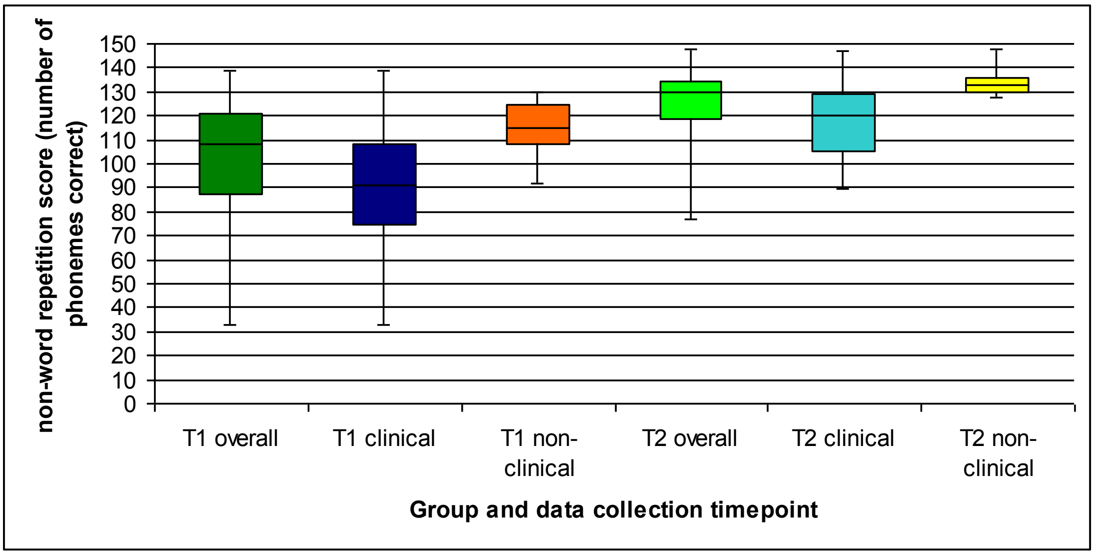
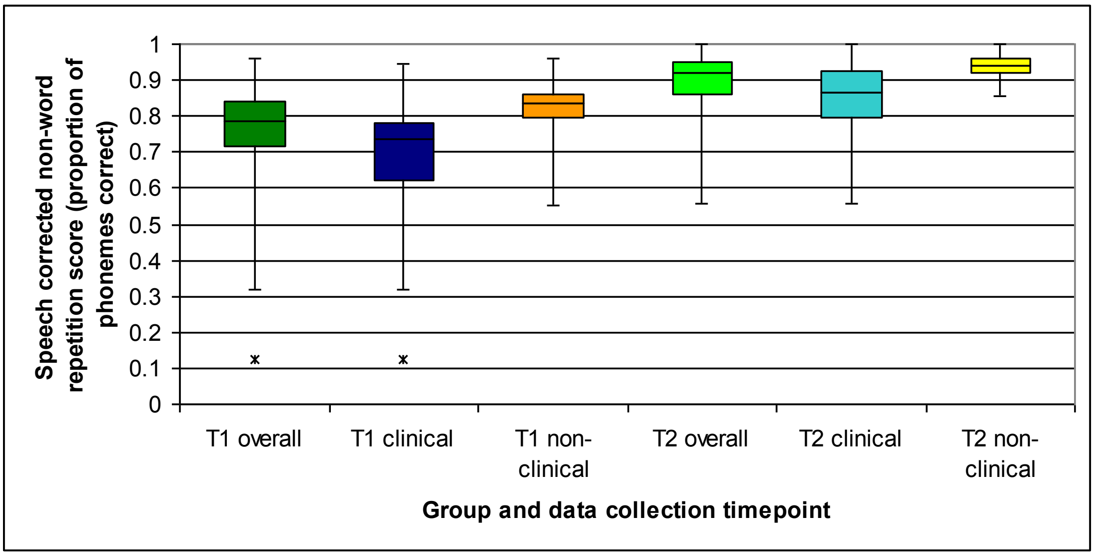


Figure 3-10 Distribution of scores on the non-word repetition task following speech error correction (proportion of phonemes correct)



The above graphs (figures 3-9 and 3-10) indicate that the range of scores on the non-word repetition task was wider than for the word repetition task. Following speech errors correction the non-clinical sample however tended towards ceiling.

3.3.6. Span task

Figure 3-11 Distribution of scores on the word span task

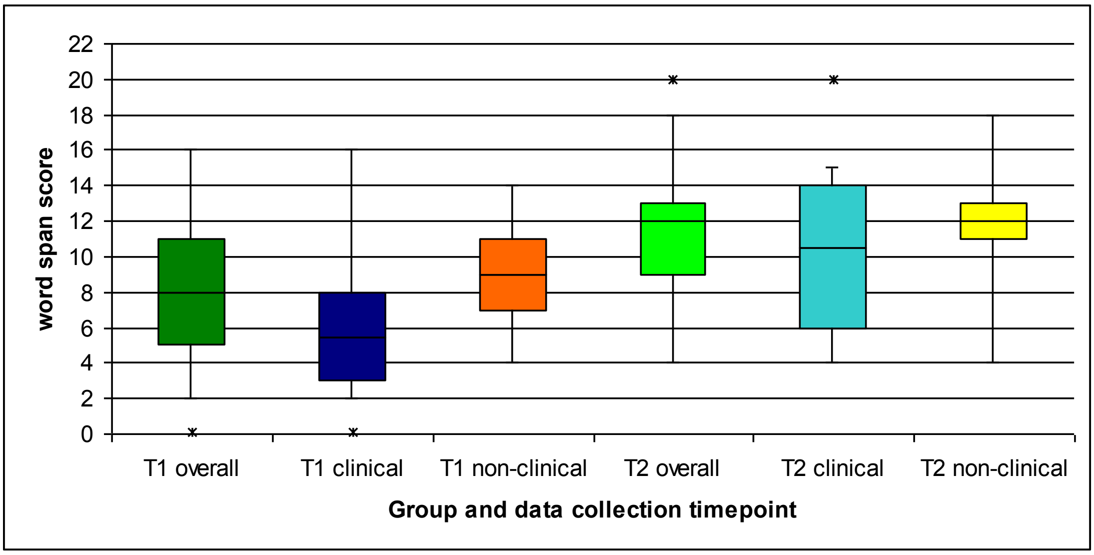


Figure 3-11 shows the distribution of scores on the span task. One child in the clinical sample scored at floor level at time 1 on this task and none of the children scored at ceiling.

3.3.7. Word Structure Test

Figure 3-12 Distribution of scores on the Word Structure Test (assessed at time 2 only)

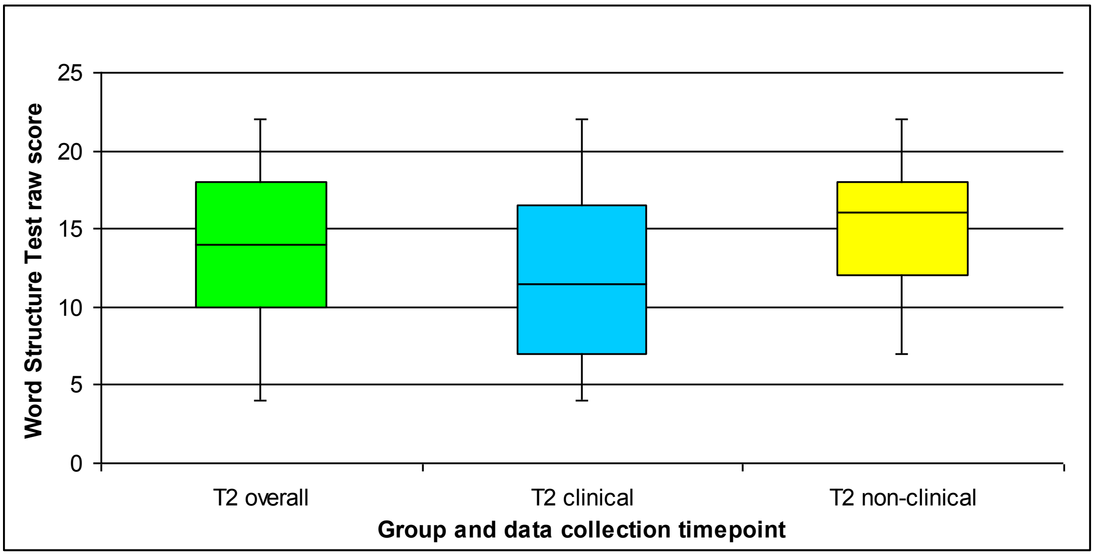


Figure 3-12 shows the distribution of scores on the word structure task (test of grammatical morpheme use). Evident in the graph, this test yielded considerable overlap of scores between the clinical and non-clinical groups.

3.3.8. Sentence Imitation Test

Figure 3-13 Distribution of scores on the Sentence Imitation task (assessed at time 2 only)

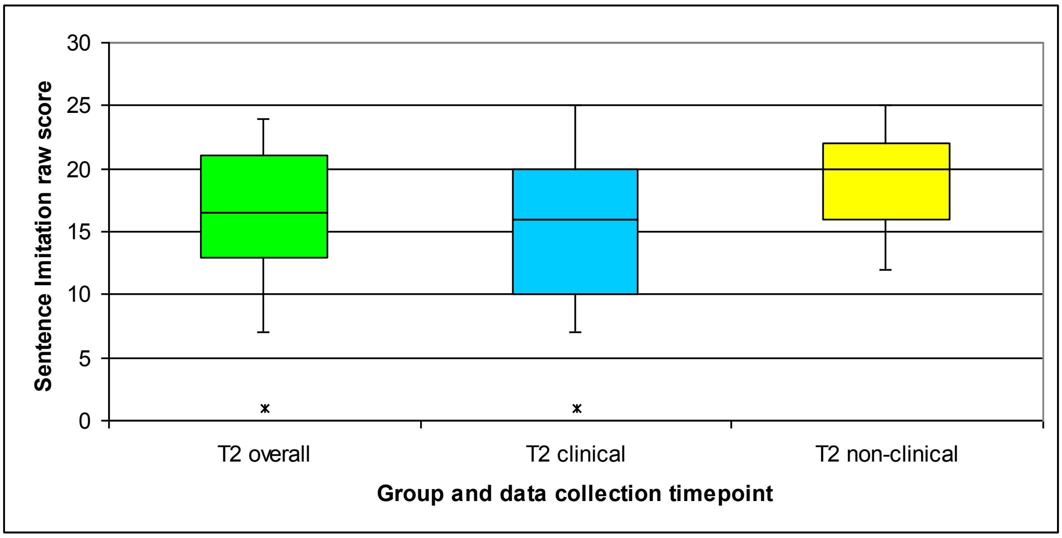


Figure 3-13 shows distribution of test scores on the sentence imitation task. Similar to the results of the word structure task, this test produced overlap in the scores by the clinical and non-clinical groups.

3.4. Effects of group, gender and age on task scores

3.4.1. Time 1 data

A series of independent samples t-tests were performed to assess effects of group (recruited through the NHS speech therapy caseload or not) and gender on the language and non-language based tasks, prior to carrying out the more detailed analyses of the results in chapters 4, 5, 6 and 7. The degrees of freedom reported reflect the number of participants included in the analyses (N-2). Therefore where degrees of freedom are less than 52, this indicates that data were missing for some of the participants due to non-compliance. The independent t-tests revealed no significant difference between the groups on the following task scores: BAS blocks $t(52)=1.063$, $p=0.292$; BAS picture similarities $t(52)= -0.179$, $p=0.859$. There were significant differences between the groups at $p=0.001$ or below for all the language-based tasks. The clinical group performed more poorly on all of these measures: PLS-auditory $t(52)=3.783$, $p<0.001$; PLS-expressive $t(50)=3.907$, $p<0.001$; Word Naming Task $t(52)=5.891$, $p<0.001$; Word Repetition Task $t(52)=4.175$, $p<0.001$; Known word repetition $t(52)=4.052$, $p<0.001$; Non-word Repetition Task $t(52)=3.503$, $p=0.001$; speech-corrected known word repetition $t(52)=3.967$, $p<0.001$; speech-corrected non-word repetition $t(52)=3.886$, $p<0.001$; Span $t(46)=3.795$, $p<0.001$. The results were consistent when t was adjusted due to unequal variances between the groups. The results were also consistent when non-parametric tests were used for those test scores that were not normally distributed (see appendix I).

There were no significant effects of gender on any of the tasks at $p<0.05$ (using both parametric and non-parametric tests), so it was not considered problematic that more boys were recruited to the study than girls.

Effects of age were also explored (although effects were not expected, given the narrow age range of the participants). Significant correlations between age and the tasks were found only for the BAS picture similarities task ($r(53)=0.32$, $p<0.05$). Although significant, this was not considered to be important for two reasons: first it does not account for much variance in the data (10.24%) and second it is not one of the tasks that were under investigation. There was no significant relationship with age for any of the other tasks; therefore, age was not included in any of the further analyses.

3.4.2. Time 2 data

Fifty-two of the original sample of 54 children participated in the follow-up phase of the project. Three further children from the sample did not complete all the tasks at T2. One child refused to attempt some of the tasks and it was possible for the researcher to visit the other two children only once at time two. Therefore, full data for time two of the study was collected from 49 of the children (30 boys, 19 girls). Partial data was collected for a further three children (1 boy, 2 girls).

As was the case for the time one data, a series independent samples t-tests were performed to assess effects of group (recruited through the NHS speech therapy caseload or not) and gender on language and non-language based tasks. Again, the degrees of freedom reflect the number of participants in the group (N-2) and where the value is less than 50 this is due to missing data for some of the children.

The independent samples t-tests revealed no significant differences between the groups on the following task scores: BAS blocks $t(49)= - 0.656$, $p=0.515$; BAS picture similarities $t(50)=1.392$, $p=0.170$); span $t(50)=1.764$, $p=0.084$.

There were significant differences between the groups (at $p<0.05$ or below) for all the language-based tasks: PLS-auditory $t(49)=2.848$, $p=0.006$; PLS-expressive $t(49)=2.515$, $p=0.015$; Sentence Imitation Task $t(49)=3.129$, $p=0.003$; CELF Word Structure subtest $t(48)=2.730$, $p=0.009$; Word Naming Task $t=4.266$, $p<0.00$; Word Repetition Task $t(50)=3.379$, $p=0.001$; Non-word Repetition Task $t(50)=4.460$, $p<0.001$ ($F(1,50)=19.27$); known word repetition $t(50)=2.053$, $p=0.045$; speech-corrected known repetition $t(50)=2.882$, $p=0.006$; speech-corrected non-word repetition $t(50)=4.107$, $p<0.001$. Results were consistent using non-parametric tests.

There were no significant effects of gender at time 2 (at $p<0.05$) for any of the tests using the independent samples t-test, or using non-parametric tests, where scores were not normally distributed (see appendix I).

A significant correlation was found between age and the PLS-expressive task ($r(49)=0.28$, $p<0.05$). As this explained very little of the variance in the data (7.84%), and because no other significant correlations were found for age, age was therefore not considered to be an important factor for these analyses and was not included in the further analyses.

3.5. Summary

This chapter has provided an overview of the participants and their scores on the tasks used in the study. As indicated in the distribution graphs, and also apparent in the data summary tables in appendices H and I, most of the data were not normally distributed. Distributions could not be normalised using data transformations, so results in the next four chapters are given using non-parametric as well as parametric tests. The next four chapters (chapters 4, 5, 6 and 7) turn attention to the research questions and the rationale behind these questions. They further report results from the specific tasks that answer the research questions; and offer interpretation of the results.

Chapter 4

Part 1: Investigating the influence of speech sound development, word knowledge and phonological short term memory on word and non-word repetition

4.1. Chapter Overview

This chapter focuses on the children's repetition of words and non-words at two time points: age 3 years and 4 years. The chapter begins with an overview of studies that have investigated NWR and WR in young children, recapping some of the background information presented in chapter 1. It identifies gaps in the existing research and it presents this study's attempts to address these.

4.2. Introduction

The ability to repeat non-words has been widely investigated in the research literature and has been proposed as a clinical marker of language disorders (e.g. Conti-Ramsden et al. 2001). NWR is a useful assessment because it is quick and simple to administer, and there is some evidence to suggest that early NWR skills predict later language skills (e.g. Chiat and Roy, 2008; Gathercole et al., 1992; although see also Chiat and Roy, 2013; Melby-Lervåg, Lervåg, Lyster, Klem, Hagtvet and Hulme, 2012). Some studies have also explored real WR in young children and have investigated the relationship between this and wider language skills (e.g. Casalini et al., 2007; Chiat and Roy, 2007; 2008; Dispaldro et al., 2009; Gathercole and Adams, 1993; Roy and Chiat, 2004). The following account discusses in turn the tasks of NWR and WR, with a particular focus on studies that have recruited young (pre-school) children without identified language-learning difficulties.

4.2.1. Non-word Repetition task

The NWR task requires a participant to repeat a nonsense ‘word’ immediately after hearing this. The ‘words’ might comprise one or several syllables and responses are scored.

4.2.1.1. NWR and phonological short-term memory

In its early use, the NWR task was put forward as a ‘pure’ measure of the phonological short term memory (PSTM) system (Gathercole and Baddeley, 1989, 1993) and a test, the Children’s Test of Non-word Repetition (CNREP), was published by Gathercole and Baddeley (1996) for this purpose. The task requires the participant to repeat made-up words (non-words) immediately after the assessor articulates these. The non-words might differ in their similarity to real words in terms of their phonological structure (e.g. uncommon combinations of phonemes, unusual stress patterns). NWR was proposed to be a ‘purer’ measure of PSTM than the span task (previously described, see 1.4.2.5) as it does not depend on long-term word knowledge to the same extent as the word span task (see Gathercole and Adams, 1993). NWR consistently shows item-length effects: longer non-words are more difficult to repeat than shorter non-words (e.g. Archibald and Gathercole, 2007; Gathercole and Baddeley, 1990; Gathercole et al., 2004). This is thought to reflect the limited capacity of the PSTM store, although alternative explanations have also been proposed and are explored in section 4.2.1.2. However, NWR is a different task to the span tasks, for example not including any verbal rehearsal element that is usually considered to aid performance in span tasks, at least in children over 5 years old (Henry, 1991) (see also chapter 1, section 1.4.2.5).

One study that directly investigated the equivalence of the span task and NWR task in measuring PSTM was that by Archibald and Gathercole (2007). Using the same syllables in their NWR task as single syllable non-word items in a span task, they compared children's performance on the two tests. Children performed better on the NWR test than they did on the span task, indicating that NWR was not an exact match to the span task. This could have been due to non-word segments being facilitated through 'prosodic chunking' (see chapter 1, section 1.4.2.6), although in a second paper (Archibald and Gathercole, 2007b) the authors controlled for prosody by keeping an even stress on all syllables and the advantage for non-words over single syllables remained. Alternatively the difference might be explained by the fact that individual syllables in a non-word repetition task are presented over a shorter period of time than they are in a span task. According to Baddeley's model of Working Memory, the phonological store component of the phonological loop is a time-limited store, where phonological material is vulnerable to decay unless they are rehearsed (see 1.4.2.5). During the span task, items are presented at intervals of one second, while in the NWR task the syllables the items are given in given in quick succession. This means that the length of time before recall is reduced in the NWR task.

4.2.1.2. NWR and language knowledge

Several researchers refuted the claim that long-term language knowledge is not accessed when repeating non-words (e.g. Hulme et al., 1991; Snowling et al., 1991) and this is now widely agreed to be the case. The position is supported by the findings set out in chapter 1 that non-words that are more 'word-like' are repeated more accurately (e.g. Dollaghan, Biber and Campbell, 1995; Gathercole et al. 2001). For example, that non-words containing familiar combinations of phonemes (e.g. Edwards, Beckman and Munson, 2004), familiar grammatical morphemes (Casalini et al. 2007) or familiar stress patterns (Dollaghan et al., 1995) increase accuracy of repetition. In acknowledging these linguistic factors, Gathercole (2006) proposed that the accuracy with which children repeat non-words depends not only on the quantity of phonological material, but also on its quality (i.e. the type of information stored). Chiat (2006) builds on this argument by highlighting that the extent to which each child will be able to draw on existing language knowledge, including phonological representations, will depend on their previous exposure to language material and consequential phonological sensitivity, e.g. sensitivity to prosodic structure.

As discussed in the introduction (section 1.4.1.2.) and illustrated in figure 1-2, additional factors are likely involved in the repetition of words and non-words. Indeed, Bowey (2006) states that NWR is: "a complex task involving several components, most involving phonological processing. These include speech perception, the construction, and encoding of a phonological representation in the phonological store, maintenance of this representation, retrieval of the representation from the phonological store, assembly of articulatory instructions, and articulation itself" (p548). Therefore, while the present study has PSTM and existing phonological representations as its focus, it cannot be ignored that difficulties at the level of speech perception and/or the programming of the movements needed for articulating the stimuli might also impede performance. These latter factors could particularly play a role where non-word stimuli are several syllables in length.

Further support for a link between NWR and language knowledge comes from studies by Gathercole and Adams (1993; 1994). They showed that children's NWR accuracy is correlated with their receptive vocabulary at age 3, 4 and 5 years, while digit span, which is the task more commonly used to measure PSTM is not related to vocabulary at these ages. They further looked at correlations between word span and vocabulary at 5 years. This is helpful because it might be that young children perform more poorly on tasks involving digits due to unfamiliarity with numbers. The authors found no correlation between word span and vocabulary either (Gathercole and Adams, 1994). They did find positive correlations between NWR and the two span tasks, suggesting that there is a role for PSTM, but their results from their study are consistent with the view that NWR taps language in addition to PSTM. A further study by the same authors used both quantitative and qualitative measures to explore the relationship between NWR with language (Adams and Gathercole, 1995). The children in their study, aged 3 ½ years, were split into two groups: one with good PSTM (measured using a NWR and digit span task) and the other with poor PSTM skills. A combination of language measures as well as recordings of the children's language during structured and free play tasks informed the study findings. The authors reported that poor PSTM skills were associated with language that was less grammatically sophisticated, showed a more restricted range of vocabulary, and consisted of shorter phrases.

4.2.1.3. NWR and its relationship with language development

Gathercole and Baddeley (1990b) proposed a relationship between NWR and word learning. They categorised 5 - 6 year old children as having high or low PSTM, based crucially on their scores on a NWR task and they assessed their ability to learn names for toys. They found that those children with good NWR accuracy learned unfamiliar names more quickly than children with poor NWR scores. The groups were however equivalent in their speed of associating familiar names with the toys. The accurate NWR group also showed better retention of the toys' names (unfamiliar and familiar) the day after testing. A further study by the authors and additional colleagues (Gathercole et al., 1992) proposed a causal relationship in younger children between NWR ability and vocabulary development (aged 4-5 years). Their study used cross-lagged correlations and found that NWR predicted later vocabulary growth (for details see chapter 7: part 4 of the study).

4.2.1.4. NWR studies with very young children

Some studies have focussed their attention on very young children, for example, Gathercole and Adams (1993), Roy and Chiat (2004) and Chiat and Roy (2007). The latter study, which included children with identified language disorders, reported that even at age 2 ½ - 4 years, children were showing lower accuracy on NWR tasks compared to their age-matched peers (see chapter 5). A further study (Chiat and Roy, 2008) reported that the children's performance on the NWR task and, interestingly, a WR task at this young age was a good predictor of their later ability to repeat sentences (see chapter 7). WR and its relationship to language development will be discussed in the following account.

4.2.2. Word repetition task

The WR task requires a participant to repeat real single words of various lengths (number of syllables) after the assessor. This task has been used in several studies involving young children, mainly to contrast with NWR performance (e.g. Casalini, et al., 2007; Chiat and Roy, 2007, 2008; Dispaldro et al., 2009; Dispaldro, Deevy, Altoé, Benelli and Leonard, 2011; Dispaldro et al., 2013a; Dispaldro, Leonard and Deevy, 2013b; Gathercole and Adams, 1993; Roy and Chiat, 2004; Vance, Stackhouse and Wells, 2005). The studies consistently find that words are repeated more accurately than non-words, suggesting that existing lexical knowledge is important.

One study that interpreted the advantage for words over non-words in relation to a theoretical model was that by Vance et al. (2005). In their study, they compared repetition of words and phonologically-matched non-words by children aged 3 – 7 years. The children performed significantly better on the WR compared to the NWR task, except at age 3 years, where accuracy was statistically equivalent between the tasks. The authors interpreted this finding in relation to the psycholinguistic model proposed by Stackhouse and Wells (1997). The older children appeared to use existing lexical representations to support recall of the familiar speech material. At 3 years old, by contrast, the children seemed to rely on bottom-up processes to support repetition. An additional interpretation of their findings might be that the children did not know the words that they were repeating. If so, this would mean that they were unable to benefit from stored-lexical support. This suggestion is partially upheld by their findings that the same children were able to name just fewer than 60% of the words used in the repetition task.

4.2.2.1. WR and its relationship with language development

Some studies have looked at word repetition in relation to vocabulary in very young children (Gathercole and Adams, 1993; Roy and Chiat, 2004). These studies have shown that WR, like NWR is correlated with receptive vocabulary in the pre-school years. Other studies have investigated WR in relation to other aspects of language development. For example, Dispaldro et al. (2009; 2011; 2013b) found a significant correlation between both WR and NWR and grammatical skills in Italian children aged 3 and 4 years. They assessed grammatical skills according to the children's use of verb morphology and their ability to complete sentences using a pronoun in the correct form. They found that WR may be a better predictor of grammatical skills than NWR in children with typical language. They interpreted their findings as word repetition reflecting lexical abilities, which subsequently affect grammatical skill development (see Dispaldro et al. 2013b). However it is difficult to draw conclusions from the results: although the 2011 study used words that were expected to be present in the children's vocabularies at 3 and 4 years, it did not directly assess whether the children knew the words in the WR task. Furthermore, they did not state whether the greater correlation between WR (compared to NWR) and performance on the grammatical task was significant statistically. Interestingly, the authors did not find the same relationship between WR and grammar in a sample of Italian children with language difficulties (Dispaldro et al 2013b) and they also could not replicate the finding in English (Dispaldro et al. 2011).

4.2.3. Confounding variables in young children's WR and NWR

So far, a recurrent problem that has been raised from the existing research is that most studies reporting results on young children's WR have not assessed whether the words were known by the children. If children did not know the words, it would mean that they had less support from existing lexical knowledge during repetition. This would confound findings about word-length as well as conclusions that can be drawn in relation to what the task measures. A further problem associated with assessing the repetition skills in general of very young children is that these children often have difficulty using certain speech sounds. As part of normal speech development, some phonemes emerge later in children's repertoires than others. This might result in pronunciation errors being incorrectly scored as repetition errors. Some studies have controlled for this, by basing the non-words around early developing phonemes (e.g. Dollaghan and Campbell, 1998), or by correcting for typical speech patterns in the scoring (e.g. Roy and Chiat, 2004, Chiat and Roy, 2007). However, the problem remains where the test is used clinically, for children whose speech sound system does not follow the typical pattern. Instead they might make unusual and/or inconsistent speech errors. These children are commonly found in clinical samples. It might be that their results are influencing the data. Alternatively, it might be exactly because the children have delayed or unusual speech sound development that NWR is a useful tool in the identification of speech and language difficulties (e.g. see Dispaldro et al., 2013a). The present study sought to address these problems.

4.3. Purpose of the study and hypotheses

The wealth of existing literature has provided us with clear evidence in support of the usefulness of NWR tasks in tapping children's language skills. There is evidence for the contribution of both PSTM and existing language knowledge in children's repetition of these stimuli. The findings are murkier in the case of words. Results are confounded through not verifying whether the children know the words. If, as some studies suggest (e.g. Dispaldro et al. 2009, 2011, 2013a, b; Chiat and Roy, 2007, 2008), WR can be used to reveal language competence through tapping underlying lexical templates then does it provide any additional information than a naming task would? If, however, young children do not draw upon their lexical templates during WR, instead favouring bottom-up processes, does the WR task provide any additional information that cannot be gathered through a NWR task?

The purpose of the study was therefore to explore the NWR and WR accuracy of children aged 3 years and again at 4 years. The study also included a naming task. This had the dual purpose of assessing the children's knowledge of the words they were repeating and also of assessing the children's phonological representations for these words. The task was completed twice to gain a measure of consistency of speech sound errors. Having both theoretical and clinical motivations, it attempted to identify the underlying mechanisms influencing NWR and WR accuracy. Specifically this part of the study sought to explore the contribution of speech skills, word knowledge and PSTM on pre-school children's WR and NWR.

This part of the study combined findings from the clinical and non-clinical groups (see chapter 3, part 2 for rationale). In the next part of the study (chapter 5), the data were split into the two groups to explore qualitative and quantitative differences between the two groups' performance on the tasks.

The following predictions were made:

- Word knowledge will influence word repetition accuracy; i.e. where available, children will use existing word knowledge in their repetition.
- Speech sound skills will affect performance on both WR and NWR.
- PSTM will affect non-words but not known words, due to access to stored lexical representations for the latter.

This study proposed the following set of hypotheses:

- If word knowledge affects performance, there will be a sliding scale of performance: known words will be repeated more accurately than unknown words and unknown words will be repeated more accurately than non-words. The difference between unknown words and non-words was predicted because children were likely still to be familiar with the unknown words, even if this familiarity was not sufficient to lead to accurate naming (see table 1-1, chapter 1).
- If speech sound skills affect performance there will be a significant difference in scores on WR and NWR following speech error correction (scores that are corrected for speech errors will be higher than those that are not corrected).
- If NWR (but not WR) taps PSTM then there will be clear word length effects for non-words but not for words (due to phonological information fading from the PSTM store).

- If NWR taps PSTM then there will be a significant correlation between word span and NWR.
- If WR does not tap PSTM then there will not be a significant correlation between word span and WR (for known words).

4.4. Methods

4.4.1. Participants

The study recruited 59 children to the study at T1 of the study, when the children were aged 3 - 3 ½ years (see general methods chapter 2, part 1 for more information regarding demographic information, methods of recruitment and information about the children's development). Data for this part of the study were collected from 54 of these children at T1 and 52 of those 54 at T2, a year later, when the children were 4 – 4 ½ years old. However, as is the case for other studies reporting data from similar aged children (e.g. Chiat and Roy, 2007; Gathercole and Adams, 1993) there was a degree of non-compliance at both time points. In the task descriptions below, the number of children completing each task at each time point is given in brackets.

4.4.2. Tasks

Relevant to this part of the study, the children were assessed using the following tasks:

- Word repetition (54 children at T1, 52 children at T2)
- Non-word repetition (54 children at T1, 52 children at T2)
- Picture naming (same word stimuli as used in the word repetition task) (54 children at T1, 52 children at T2)
- Word span (49 children at T1, 52 children at T2)

The children's non-verbal and general language skills were also assessed at both time points. Their non-verbal skills were assessed using the BAS II (Elliott et al., 1996) and their language skills were assessed with the Preschool Language Scales 4 (Zimmermann et al., 2009).

For further information about the specific tasks completed by the children and the assessment environment, see general methods chapter 2, section 4).

4.4.3. Methods of scoring

4.4.3.1. Word knowledge

Word 'knowledge' was established using responses from the first administration of the naming task. Those items that the child was able to name were considered 'known' while unnamed or incorrectly named items were 'unknown'. Where responses were ambiguous, due to limited intelligibility the responses from the second naming task were also examined. Where vowel sounds were correct and there was consistency in the child's naming across the two occasions, the word was scored as 'known'. The participants' word repetition performance was rescored according to whether the item was 'known' or 'unknown'. An item was considered 'known' if the child named the item in the first naming task (see general methods, section 2.3). The proportion of phonemes correct was then calculated for each of the children for each of the stimulus categories (known words, unknown words, non-words)

4.4.3.2. Speech errors

In order to investigate the effect of children's speech errors on their repetition accuracy, speech errors during the two naming tasks were first examined. Where children consistently made a speech error (developmental or otherwise) on the items that they named, the error was allowed for in their repetition i.e. points were awarded as though the repetition had been accurate. Where they had not attempted to name a given item, error patterns from the other named items were scrutinised. Where speech error patterns were evident, additional points were awarded for repetition attempts that matched these error patterns. This method of scoring is described in greater detail in chapter 2 (part 2.3).

4.5. Results

4.5.1. Does children's knowledge of words affect repetition accuracy?

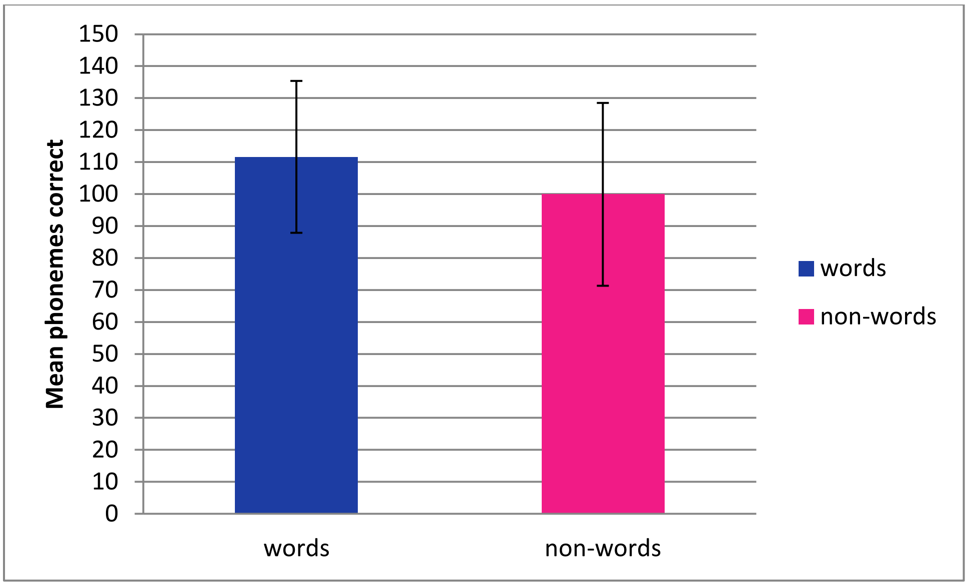
Recapping the hypotheses set out in the introduction, if children's knowledge about words aids their accurate repetition, then the following findings would be expected:

1) Greater repetition accuracy for words compared to non-words.
2) Greater repetition accuracy for words that the children were able to name ('known' words) compared to the words that they were unable to name ('unknown' words). In turn, better repetition of the real words that the children were unable to name ('unknown words') compared to non-words.

4.5.1.1. Findings at T1

The bar chart below (figure 4-1) shows the mean correct number of phonemes for all 54 children at T1 on the word and non-word repetition tasks. On this and subsequent graphs, error bars represent 1 SD from the mean.

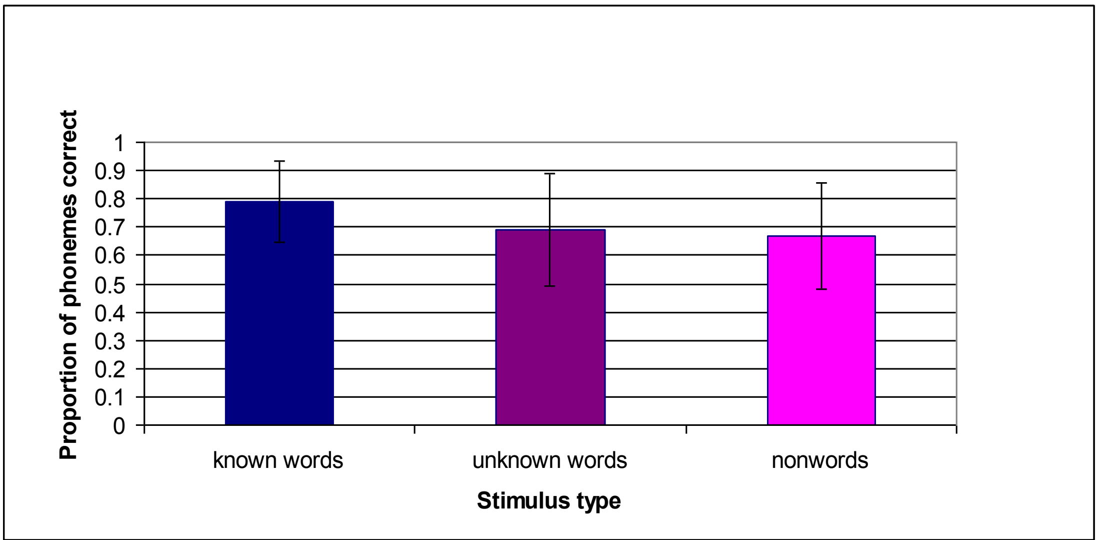
Figure 4-1 Mean number of phonemes correct for words and non-words at T1



Repetition accuracy for the 28 words and 28 non-words was compared. The number of correct phonemes (including vowel sounds) was counted for each of the participants. The maximum score was 149 phonemes correct. Means were compared using a paired sample t-test. This revealed that children repeated words ($M=111.61$, $SD=23.75$) significantly more accurately than non-words ($M=99.89$, $SD=28.59$); $t(53)=5.41$, $p<0.001$. As the data were not normally distributed (see appendix J, non parametric tests were also employed. The results were consistent with the results from the analyses using parametric tests (appendix J).

The mean number of 'unknown' words at T1 was 7.72 (range=3-15 words). Figure 4-2 shows the mean proportion of phonemes correct for words that are known, words that were ‘unknown’ and non-words.

Figure 4-2 Mean proportion phonemes correct for known words, unknown words and non-words at T1



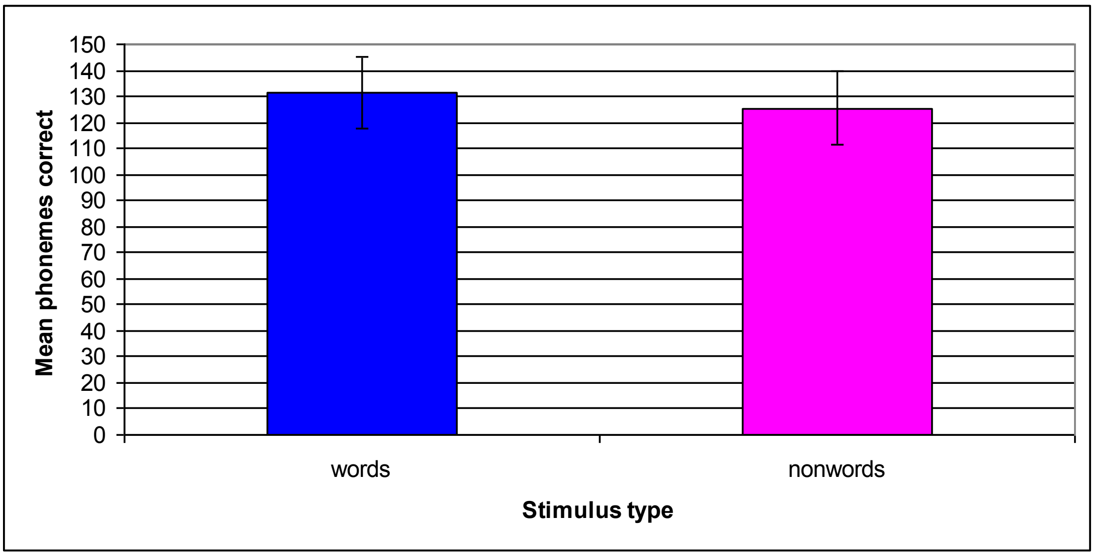
A repeated measures ANOVA was performed to explore differences in repetition accuracy across the different word types. Partial eta squared values are given as the effect size estimates. According to Cohen (1988), a value of 0.0099 represents a small effect size, 0.0588 represents a medium effect size and 0.1379 signifies a large effect size. The ANOVA revealed that repetition was affected by the stimulus type: $F(2, 104)=22.58$, $p<0.001$, $\eta_p^2 = 0.303$. Planned comparisons, applying a Bonferroni correction, revealed that scores on known words (M=0.79, SD=0.14) were repeated more accurately than non-words (M=0.67, SD=0.19) ($p<0.001$) and more accurately than unknown words (M=0.69, SD=0.20) ($p<0.001$). There was no difference between scores for unknown words and non-words ($p=0.79$).

As the data were not normally distributed (appendix I) non-parametric tests were also used. These results were consistent with the results from the analyses using parametric tests (appendix J).

4.5.1.2. Findings at T2

The bar chart below (figure 4-3) shows the mean correct number of phonemes for all 52 children at T2 on the word and non-word repetition tasks.

Figure 4-3 Mean number of phonemes correct for words and non-words at T2

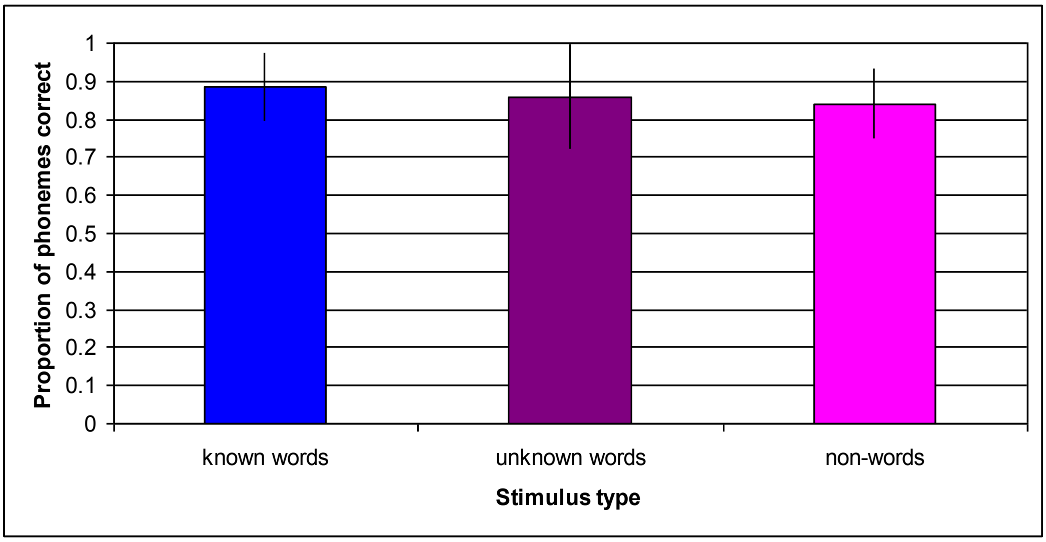


Mean scores on these tasks were compared using a paired sample t-test. This revealed that children repeat words (M=131.37, SD=14.01) more accurately than non-words (M=125.40, SD=14.17); $t(51)=5.22$, $p<0.001$.

As the data were not normally distributed (appendix I) non-parametric tests were also used. Results were consistent with the parametric tests (appendix J).

Using the same procedure as T1, proportion scores were derived for each child for *known* words, *unknown* words and non-words. At time 2 the mean number of unknown words was 4.62 (range=2-10 words). Figure 4-4 shows the mean proportion of phonemes correct for each of these stimulus types. Analysis of the data revealed a different finding from those at T1: no significant advantage for known words over unknown words was found.

Figure 4-4 Mean proportion of phonemes correct for known words, unknown words and non-words



As the data were not normally distributed (known words D(52)=0.253, p<0.001, skew= -2.78, kurtosis=9.49; unknown words D(52)=0.212, p<0.001, skew= -2.38, kurtosis=6.62; nonwords D(52)=0.177, p<0,001, skew= -1.18, kurtosis=1.36), non-parametric testing was conducted. These are reported in favour of the parametric test results, as they yielded slightly different results. A Friedman's ANOVA revealed a chi square value of 23.23, p<0.001. Post hoc testing using Wilcoxon tests (applying a Bonferroni correction so that results were considered significant if p<0.0167[1]) revealed that children repeated words (Mdn= 0.91) more accurately than non-words (Mdn=0.88), T=4.23, p<0.001). They repeated unknown words (Mdn=0.89) more accurately than non-words (Mdn=0.88), T=2.64, p=0.008. There was no significant difference in their repetition of known words (Mdn=0.91) compared to unknown words (Mdn=0.89), T=1.56, p=0.12, NS.

4.5.2. Do children's speech sound skills affect repetition accuracy?

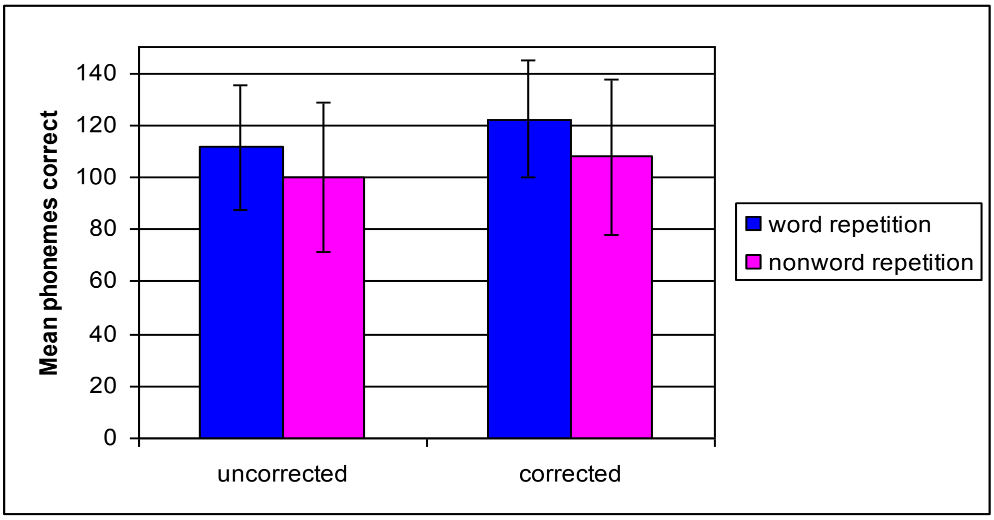
To recap the hypothesis specified in the introduction, if speech sound skills affect performance there will be a significant increase in scores on WR and NWR following speech error correction

[1] p<0.05 divided by the number of tests, i.e. 3

4.5.2.1. Findings from T1

Figure 4-5 shows the uncorrected repetition scores from T1 and the repetition scores that have corrected for speech errors.

Figure 4-5 T1 repetition scores before and after speech error correction



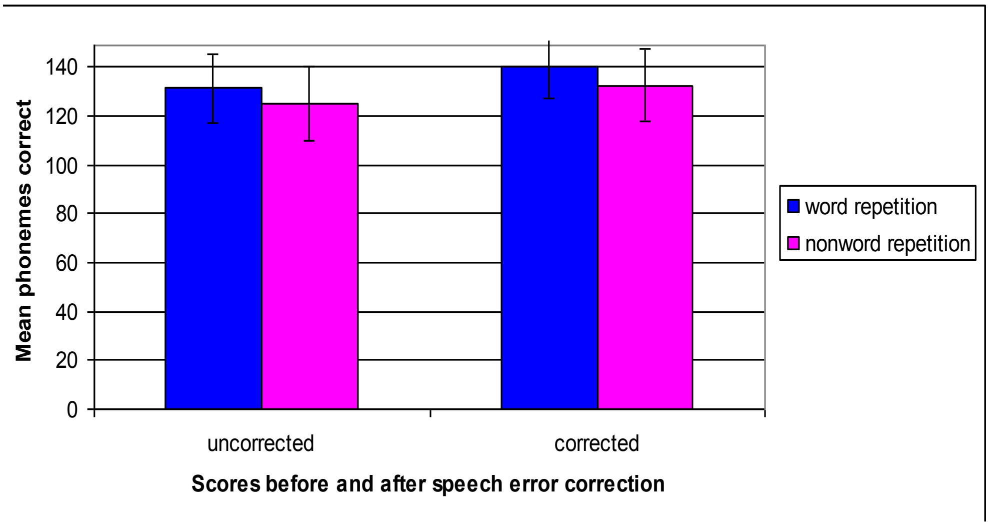
Mean scores were compared before and after speech error correction, using paired-sample t-tests. A significant difference was found in children's word repetition scores when speech errors were corrected compared with when they were not allowed for (t(53)=6.18, p<0.001). This was also the case for non-word repetition (t(53)=3.58, p<0.005) representing more accurate performance when speech errors were allowed.

Following speech error correction there remained a significant difference between scores on the word repetition test (M=122.17, SD=22.50) and the non-word repetition test (M=107.87, SD=29.88): t(53)=5.80, p<0.001. Results were consistent using non-parametric tests (appendix J), which were conducted due to non-normally distributed scores.

4.5.2.2. Findings from T2

Figure 4-6 shows the uncorrected and corrected repetition scores from T2Figure 4-6 shows the original repetition scores from T2.

Figure 4-6 T2 repetition scores before and after speech error correction



Performance on the word and non-word repetition tasks was compared before and after allowing for speech errors, using paired-sample t-tests. A significant difference was found in children's word repetition scores when speech errors were corrected compared with when they were not corrected: $t(51)=9.97$, $p<0.001$. This was also the case for non-word repetition: $t(51)=9.78$, $p<0.001$.

After speech error correction there remained a significant difference between scores on the word repetition task ($M=140.06$, $SD=13.09$) compared to the non-word repetition task ($M=132.69$, $SD=14.76$): $t(51)=5.02$, $p<0.001$.

As the data were not normally distributed and therefore violated the assumptions made by parametric tests, non-parametric tests were also conducted and were consistent with the results reported above.

4.5.3. Does children's phonological short-term memory influence repetition accuracy for words and non-words?

The hypotheses stated that if the repetition tasks tap PSTM then item length effects would be evident. The present study predicted this to be the case for non-words but not for known words. For the known words, the PSTM store is not taxed as children should be able to use their stored lexical representations when repeating these items. For non-words however, longer items would tax the PSTM more, due to phonological information fading from the PSTM store. This would lead to more errors for these longer non-words. In addition, if the repetition tasks tap PSTM, then there will be a significant correlation between word span and non-word repetition scores, but not word repetition scores.

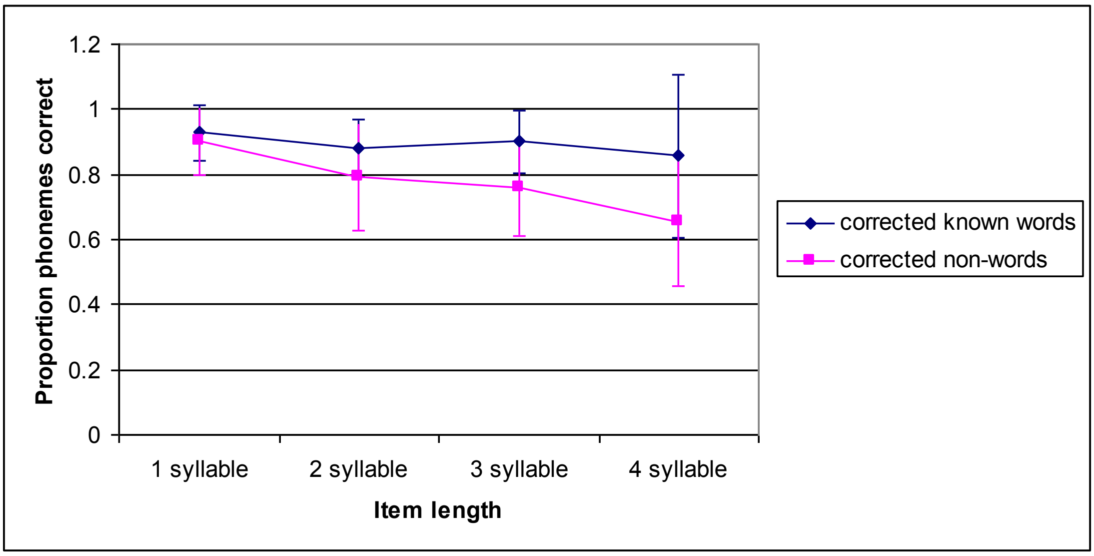
Due to the nature of the study, there were several different ways to analyse the data. For example stimuli could be coded simply as words and non-words. However, unknown words were found to be treated as non-words at T1 and there was some evidence that they were treated as real words at T2. This may have been due to the way in which word knowledge was established in the present study and the possibility that judgements about knowledge were based on individual pictures. It might be that a child would have known a given word in a different context. In order to avoid any resulting confounds, data are presented for known words and non-words only. Additionally, it was decided to present data only for speech error-corrected data. However, complete data (where speech errors have not been corrected and where unknown words are included in the analysis) can be found in appendix K. This shows a similar pattern of performance in the case of non-words, but a different pattern for known words.

As described in chapter 2 (section 4: methods for scoring the data), scores for both repetition tasks were divided into separate word length scores (1, 2, 3 and 4 syllables). Each of the word repetition scores was divided into known words and non-words. Proportion scores were calculated based on the total possible score for the known words at each syllable length. Five of the children had not demonstrated knowledge of any words at 4 syllables and one of these children had not demonstrated knowledge of any words at 3 syllables. These children's data were therefore removed completely from the analyses, so that data at each item length was compared to equally sized data samples at the other word lengths.

4.5.3.1. Findings from Time 1

Figure 4-7 presents the data for speech error-corrected known words and non-words across the different word lengths at T1.

Figure 4-7 Mean proportion phonemes correct for known words and non-words at T1 (it should be noted that proportion scores are reported, therefore scores cannot exceed 1)



To explore differences in repetition accuracy by the group on words and non-words of different lengths, a repeated measures ANOVA was conducted. Mauchly's test showed that the assumption of sphericity was violated, so degrees of freedom were corrected using Greenhouse Geisser estimates. Partial eta squared values are given as the effect size estimates. The ANOVA revealed a main effect of 'word' type: $F(1, 47)=58.07$, $p<0.001$, $\eta_p^2=0.553$. It also revealed a main effect of item length: $F(2.25,105.9)=23.48$, $p<0.001$, $\eta_p^2=0.584$. Pairwise comparisons revealed that 1 syllable items were repeated more accurately than all other items at $p<0.001$, 2 syllable items were repeated more accurately than 4 syllable items ($p<0.05$), there was no significant different between 2 and 3 syllable items ($p=1$), there was no significant difference in repetition of 3 and 4 syllable items ($p=0.053$).

There was a significant interaction between 'word' type and item length: $F(2.06, 96.9)=12.59$, $p<0.001$, $\eta_p^2=0.475$. Simple effects analysis revealed that there were significant effects of item length for both words (and non-words.
To further explore the interaction between word type and syllables, separate repeated ANOVAs were calculated for words and non-words. As reported in the simple contrasts above, there was a significant effect of item length for words: $F(3,141)=2.98$, $p<0.05$ $\eta_p^2=0.06$. Planned comparisons, using repeated contrasts revealed that this significant effect was driven entirely by 1 syllable words being repeated more accurately than the longer words ($p<0.05$). All the other comparisons were not significant (i.e. 1syll>2syll=3syll=4syll).

There was also a significant effect of item length for non-words (for non-words, Mauchly's test for normality was significant, so Greenhouse Geisser values are given): $F(2.05, 109)=30.07$, $p<0.001$, $\eta_p^2=0.362$. Planned comparisons, using repeated contrasts revealed a significant difference between all comparisons at ($p<0.004$) (i.e. 1syll>2syll>3syll>4syll).

As the data was not distributed normally, analysis was repeated using non-parametric tests. The results were consistent with the parametric tests and are reported in appendix J.

The finding that word length effects are largely not present for known words but they are for non-words is consistent with the hypothesis that children draw upon PSTM for non-words but not for words.

Correlations were explored between scores for the words that were known and the span task, and for the non-words and the span task. A significant correlation was found between the non-words and the span task: $r(46)=0.48$, $p=0.001$. A significant correlation was also found between known word repetition scores and the span task: $r(46)=0.33$, $p=0.025$.

As repetition tasks (non-word, word, span) are similar tasks, they likely involve shared skills in addition to those under investigation. In order to assess the variance shared by the individual repetition tasks and the span task (but not shared by repetition tasks more generally), the variance shared between the two repetition tasks was partialled out from the correlations. A significant correlation remained for non-word repetition and span ($r(43)=0.38$, $p=0.01$) but not for known word repetition and span ($r(44)= -0.001$, $p=0.99$).

To further confirm the above findings, multiple linear regression analysis was performed. NWR was inserted as the dependent variable and word WR and span were inserted as the predictors. A forced entry method was used. Overall the model was significant ($F(2,44)=25.57$, $p<0.001$), explaining 51.6% of the variance (adjusted R-squared). Both WR and span were significant predictors of NWR ($t=5.37$, $p<0.001$ and $t=2.70$, $p=0.01$, respectively).

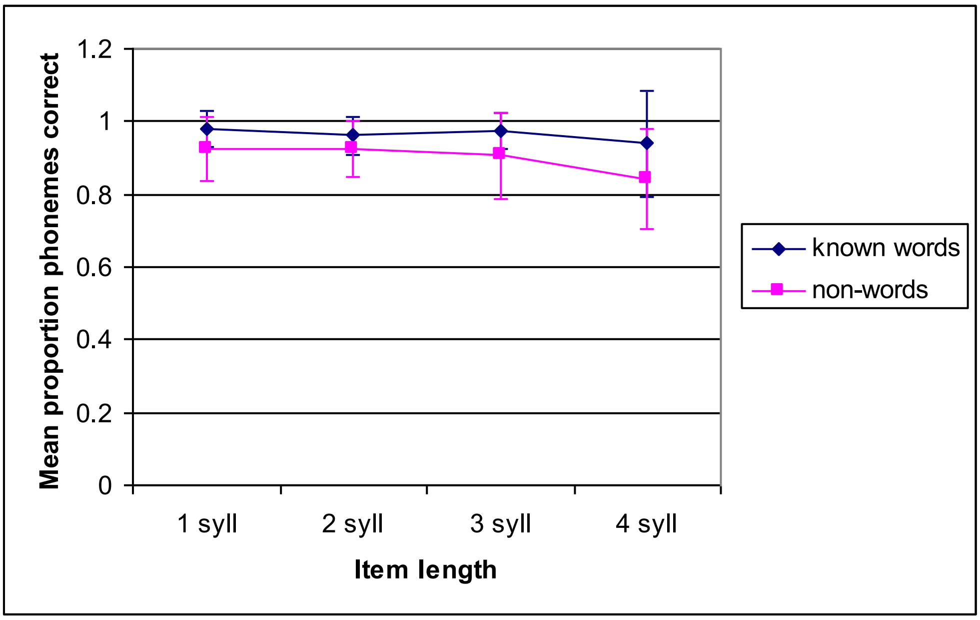
A multiple linear regression analysis, inserting WR as the dependent variable and NWR and span as the predictors was also performed. Overall the model was significant ($F(2,44)=18.81$, $p<0.001$), explaining 43.6% of the variance (adjusted R-squared). Consistent with the partial correlation, NWR was a significant predictor ($t=5.37$, $t<0.001$), but span was not ($t=-0.01$, $p=0.99$).

The results of the regression analyses are consistent with the hypothesis that PSTM is involved in the repetition of non-words but not words.

4.5.3.2. Findings from Time 2

Figure 4-8 presents the data for corrected known words and non-words across the different word lengths for T2.

Figure 4-8 Mean proportion phonemes correct for known words and non-words at T2 (it should be noted that proportion scores are reported, therefore scores cannot exceed 1)



A repeated measure ANOVA was conducted to explore differences in repetition accuracy across different length 'items' and across word and non-words. The

ANOVA showed a significant main effect of 'word' type (words and nonwords): $F(1,49)=68.6$, $p<0.001$, $\eta_p^2 = 0.583$. There was also a significant main effect of item length: $F(3, 147)=13.43$, $p<0.001$, $\eta_p^2 = 0.215$. There was a significant interaction between the 'word' type and item length: $F(3, 147)=5.85$, $p=0.001$, $\eta_p^2 = 0.107$.

Again, this was explored using separate ANOVAs. There was a significant effect of length for word repetition: $F(3,147)=3.44$, $p=0.019$, $\eta_p^2 = 0.066$. Planned comparisons, using repeated contrasts however revealed that none of these comparisons reached significance at $p<0.05$.

There was a significant effect of length for non-word repetition (Greenhouse Geisser values are given, as Mauchly's test for sphericity was significant: $F(2.45, 125)=11.02$, $p<0.001$, $\eta_p^2 = 0.18$. Planned comparisons using repeated contrasts revealed that the significant effect was entirely driven by 4 syllable non-words and all other item lengths (at $p<0.005$). No other contrasts were significant at $p<0.05$ (i.e. 1syll=2syll=3syll>4syll).

Non-parametric tests were also conducted to explore the results above (due to non-normally distributed data). The results of the tests were consistent with the parametric tests.

As for T1, length effects were more evident for non-words than words at T2. There were no word length effects for known words. There was some evidence of length effects for non-words: 4 syllable non-words were repeated less accurately than all other non-word lengths. This partially supports the hypothesis that PSTM is implicated in the repetition of non-words but not words. However, all the results at T2 were confounded by ceiling effects.

As was the case for the T1 data, correlations were explored between scores for the words that were known and the span task, and for the non-words and the span task. Results should be interpreted with caution, as ceiling effects for the repetition tasks were evident. A significant correlation was found between the non-words and the span task: $r(51)=0.462$, $p<0.001$. A significant correlation was also found between known word repetition scores and the span task: $r(51)=0.296$, $p=0.033$.

As for T1, the variance shared between the two repetition tasks was partialled out from the correlations. A significant correlation remained for non-word repetition and span ($R(49)=0.382$, $p=0.006$) but not for known word repetition and span ($R(49)= -0.092$, $p=0.523$, NS).

Multiple linear regression analysis with the NWR score inserted as the dependent variable and WR and span inserted as the predictors revealed a significant model ($F(2,49)=35.02$, $p<0.001$), explaining 58.8.6% of the variance. WR was a significant predictor of NWR ($t=6.50$, $p<0.001$) and span was also a significant predictor ($t=3.123$, $p=0.003$).

Multiple linear regression analysis with WR inserted as the dependent variable and NWR and span inserted as the predictors revealed a significant model ($F(2,49)=32.84$, $p<0.001$, explaining 55.5% of the variance. NWR was a significant independent predictor ($t=7.46$, $p<0.001$), but span was not ($t=-0.64$, $p=0.52$).

Again, at T2 the results of the regression analyses were consistent with the hypothesis that PSTM is involved in NWR but not WR.

4.6. Discussion

4.6.1. Does children's knowledge of words affect repetition accuracy?

At 3 years and 4 years of age, children repeated real words more accurately than non-words. These results support previous studies (e.g. Casalini et al., 2007; Chiat and Roy, 2007; Dispaldro et al., 2009; Gathercole and Adams, 1993; Sundström, Samuelsson and Lyxell, 2014; Vance et al., 2005) and likely reflect the influence of long-term lexical knowledge on their repetition. In the current study, this finding remained when speech errors were corrected and this is discussed further in the next section (section 4.6.2).

A novel comparison made by the present study was between words that were known to the children (as assessed by their ability to name these during the picture-naming test) and words that they had not be able to name and were therefore considered *unknown*. At T1, known (named) words were repeated more accurately than the words the children did not know. The finding indicates that children are processing words at a 'deeper' level of processing than they do for either non-words or for unfamiliar real words. It is reasonable to assume that they are using their existing knowledge of the words to support their recall (see 1.4.1.2). Division of the words into known words and unknown words resulted in the same pattern of results at T2 in terms of numerical score (known real > unknown real > non-words). However, at T2 accuracy on all items was greater. Ceiling effects are also evident for the repetition of the known words so that the gap between them and the other two word types closed somewhat, and resulted in the difference between known and unknown real words no longer being significant.

Chiat and Roy (2007) also indicated ceiling effects in the scores of their sample of 315 children (overall for the combined word and non-word repetition score). They did not however discuss whether ceiling effects were more apparent for word repetition scores compared to non-word repetition and whether this was problematic for the analyses.

In the present study, the *unknown* words were repeated with the same statistical accuracy as non-words at both time points. Although this aspect of the investigation was novel in relation to other studies involving single item repetition, the finding is consistent with a study by Boyle and Gerken (1997) that compared 2 year old children's repetition of familiar-, unfamiliar- and non-words within sentences. The authors also found that children repeated known words (nouns and verbs) more accurately than both unfamiliar words and non-words, which were repeated with equivalent accuracy.

The finding should also be considered alongside the study by Dispaldro et al. (2009) in which repetition of early-acquired and later-acquired words is investigated. Although this is not the same comparison as made in the present study, it is possible that early-acquired words were processed like known words but that the later-acquired ones were processed by the pre-school (Italian) children, like unknown ones. However, in contrast to the findings presented here, Dispaldro and colleagues found no difference in the accuracy with which their participants repeated the two sets of words.

Two possible explanations for the different findings between the studies can be considered. First, it might be that the participants in the study by Dispaldro et al. (2009) had similar knowledge for the later-acquired as they did for the early-acquired words used in the study. Although the children were younger than the age that the ‘late acquired’ words would typically have been learnt, it might be that the children did in fact know these words. Indeed, no assessment was made of the children’s familiarity with either set of words. Knowledge of the late acquired words would have led the children to treat these similarly to the early acquired words during repetition.

A second possible explanation for the difference in findings between the present study and that of Dispaldro et al. (2009) is that the ‘unnamed’ words in the present study were perhaps phonologically more complex than the words that the children did name. This could lead the children to repeat these less accurately than the familiar, phonologically less complex words. In order to examine this possibility, it is first necessary to consider what is meant by phonological complexity. Chapter 1, section 1.5.2. described phonological complexity in relation to atypical stress patterns and presence or absence of consonant clusters. Longer words (in terms of syllable number) may also be more complex to articulate, having fewer templates to draw upon. A post hoc examination of the items that the children were least likely to name (where 10 or more children did not correctly label the item at T1) was carried out using the following features as risk factors: number of syllables; typicality of stress; and presence of consonant clusters. The results are reported in detail in appendix L and summarised here. Words that might be considered to be maximally difficult, because they had three of these risk factors i.e. *trampoline* and *binoculars*, were not found to be the most difficult items for the children to name. The words *macaroni, harmonica* and *avocado*, which had only two risk factors each, were named by fewer children. Furthermore, the word *umbrella,* which has three of the risk factors, was named correctly by all but 6 of the children at T1 and all but one child at T2. The ‘unnamed’ words in the present study were therefore not phonologically more complex than the words that the children did name. The key difference was therefore more likely to be that they were unknown.

The finding that children did not show enhanced repetition of unknown (i.e. unnamed) words compared to non-words is somewhat surprising: even if children did not have sufficient knowledge of the words to enable them to label a picture of the object, it seems likely that they would already have encountered these words at some point in their lives. Indeed a possible limitation of the study is that the naming task, designed to measure the children's word knowledge, might not have accurately measured this. If the children had heard the words before, it might be assumed that they would have formed a phonological representation for these words, even if overall the lexical representation were less well specified. This rationale led to the hypothesis that these words would be repeated more accurately than the non-words, but the hypothesis was not confirmed by the findings. Two possible interpretations for the surprising findings arise and are discussed as follows.

The first possibility is that the children had in fact never encountered the words that they were unable to name. This seems plausible for some of the words, e.g. *macaroni, avocado*, *harmonica* and *binoculars*. If the children had never heard the words before, this would lead them to treat these as non-words. It seems unlikely though for some of the other words, e.g. *toe*, where it is more likely that the high rate of difficulty naming this item was due more to the unsatisfactory elicitation of the target by the picture. Many of the children persisted in calling this picture a foot or feet, despite the researcher's attempts to focus their attention more narrowly on the *toe*.

The second possibility is that even if the children had previously encountered the words, without sufficient visual and/or semantic information it was not possible to retrieve the phonological representations for the words. The unnamed words would therefore be treated as non-words, benefitting only from sub-lexical knowledge (the phonological templates in the model, see chapter 1, figure 1-5). Linked to this possibility is the consideration about how the non-words for the present study were formed. These, being created from the real word syllables, were phonotactically similar to the real words. Therefore they could presumably enjoy more support from existing sub-lexical templates than other less word-like stimuli might do. There may have been a more ‘stepped’ finding (i.e. known words > unknown words > non-words, as had been expected) had the non-words been created to be less word-like.

To summarise, the consistent finding that words are repeated more accurately than non-words is in line with previous studies (e.g. Chiat and Roy, 2007; Dispaldro et al., 2009; 2011; 2013a,b; Roy and Chiat, 2004; Vance et al., 2005). It likely reflects activation of the child’s long-term lexical knowledge during word repetition. Because of this activation, the PSTM is unlikely to be over-taxed: the phonemic and sub-phonemic detail do not need to be stored as these are already assembled in the lexicon. Returning to the model of word and non-word repetition set out in the introduction (chapter 1, figure 1-5), the present study supports the view that known words are processed at a deeper level of processing than non-words. At T1 there is evidence that when these words are not known, they do not benefit from the same depth of processing. This latter finding had not previously been investigated in the research literature and could have important implications for the use of WR tasks clinically.

4.6.2. Do children’s speech sound skills affect repetition accuracy?

The second question investigated whether speech sound errors affect performance on the repetition tasks. The results showed that there was a significant difference in children's repetition scores before and after speech error correction. This was the case for words and non-words, i.e. speech errors were affecting both words and non-words similarly. These findings were consistent at phase one and two. As already discussed above, following the speech error correction, the 'word knowledge' effect remained after speech error correction.

As discussed in this chapter's introduction, previous studies that have focussed on NWR in young children have used different methods to reduce effects of young children's typical speech errors. Studies have, for example, based the non-words exclusively around early-developing phonemes (e.g. Dollaghan and Campbell, 1998), or they have reported correcting for errors that are known to occur in typical development (e.g. Chiat and Roy, 2007; 2008), or they have looked for evidence of the same speech errors elsewhere in the repetition task and corrected for these where found (Ellis Weismer, Tomblin, Zhang, Buckwalter, Chynoweth, and Jones, 2000). These approaches likely work effectively for children whose speech and language is developing typically. However, many speech errors made by young children with speech and language disorders do not follow a typical developmental pattern. As discussed in this chapter's introduction, it might be that errors produced by these children during repetition are not corrected in these tests, as there would be no means of distinguishing these from repetition errors. It might be, therefore, that these speech errors are reducing the children's potential repetition scores. Conversely, other repetition tests might assume developmental speech errors that are actually errors made due to inaccurate repetition.

This study overcame some of the speech-related problems by exploring both children's naming and repetition of the same stimuli. This was a novel contribution by the present study. The participants had the opportunity to name the items twice and this served as a measure of consistency. This was important since almost 10%

of children with speech disorders are known to have inconsistent speech, whereby they pronounce the same word differently each time they say it (Broomfield and Dodd, 2004) All (and only) repetition errors that were also present in the children's naming were corrected. As the same syllables were used in the non-words as the words, the information from the naming could also be used to correct errors in non-words. One limitation of this approach was for children who could not name some of the items. For these items, the available naming data was used to determine whether any phoneme error in the repetition task could be explained by a speech delay or disorder.

Most other studies that correct for speech errors have not presented their results before and after the correction. One exception to this is the study by Dispaldro et al. (2013a). In their study, they reported results using two methods of scoring. One method calculated percentage phonemes correct and they corrected for typical speech errors. The other method computed percentage whole words correct and did not allow for speech errors. The results were similar, but not identical for the methods adopted. The more time-consuming method of correcting for speech errors and scoring for percentage phonemes correct identified that Italian children with SLI performed differently to their age-matched peers on their repetition of words compared with non-words (there was a statistical interaction between these variables). The difference between accuracy by children with SLI compared to performance by their peers was greater for non-words than for words. In contrast, when speech errors were not corrected and items were scored for whole words, no interaction between these variables was found. This is an important finding, as it indicates that WR tasks might identify children with language difficulties partly due to speech errors.

The results from the present study indicated that speech errors were affecting words and non-words in a similar way, when the participants are considered as a single group (i.e. not distinguishing typically developing children from those identified as having a language impairment). The present study did find other evidence of speech errors masking the results though. There was evidence that speech errors affected different length known word items differentially (see appendix K for uncorrected speech word-length effects).

4.6.3. Does children's phonological short-term memory affect repetition accuracy of words and non-words?

The results suggest that phonological short-term memory is important for repetition of non-words but not words. There were two main findings in support of this: presence and absence of 'word' length effects and results of the correlation and regression analyses. Given the earlier findings that unknown words were treated similarly to non-words, the data that were of most interest to compare were the *known* words and non-words.

When the participants were aged 3 - 3 ½ years there were word length effects for non-words but not known words. They made more errors on non-words as the item length increased. Word-length effects have also been identified by previous studies, where participants of various ages repeated non-words (e.g. Chiat and Roy, 2007; Dollaghan and Campbell, 1998; Ellis Weismer et al., 2000; Gathercole and Adams; Gathercole et al., 1992; 1993; Gray, 2003).

Removal of the unknown items resulted in no evidence of word length effects for the known words, except in the case of one syllable words, which were repeated more accurately than the other words. This finding strongly supports the position that children draw upon their long-term representations when they repeat known words and that repetition performance is therefore not constrained by the limited capacity PSTM system. This effect was more evident when speech errors had been corrected (see appendix K for comparison).

At T2, comparison of the items (grouped by lexical status) across different lengths also revealed that words showed no evidence of word length effects, consistent with the T1 findings. There was some limited evidence of length effects for non-words: 4 syllable non-words were repeated significantly less accurately than all the other items and significantly better than the real words of this length. It is however difficult to interpret the findings from T2 as ceiling effects were apparent for both lexical groups.

Results from the correlations and regressions were consistent with the T1 findings. A significant correlation was found between NWR and span that was maintained when the variance shared with word repetition was partialled out. This was not the case for the WR task. The findings were also supported by the regression modelling. As before, this provides further support for the view that PSTM is involved in the repetition of non-words but not known words at 4 years old.

The finding that non-words but not known words are vulnerable to length effects is consistent with the view that PSTM is involved in the repetition of non-words but not words. The findings are consistent with the studies by Dispaldro and colleagues (2009, 2011, 2013a,b), who found that the repetition of non-words by Italian children was sensitive to word-length effects, indicating dependence on PSTM, but that the repetition of words was not. The results are also similar to those reported in a study by Gathercole and Adams (1993). The 3 year old children in their study showed a tendency to repeat longer non-words less accurately, but they did not show the same tendency for real words. However, the authors of this study did not report whether the difference in repetition accuracy was significant. They also included only 1-3 syllable words and, like the studies by Dispaldro and colleagues, they assessed the children's general vocabulary, rather than specifically whether they knew the words that they were repeating.

While the favoured explanation for the presence of a 'word' length effect in the repetition of non-words but not words is the PSTM account, an alternative explanation can be considered. The repetition of unfamiliar sequences of phonemes presumably requires the construction of a new motor plan prior to articulation. Longer non-word items would require a greater level of motor programming, which increases the opportunities for errors to be made and might lead therefore to errorful articulation. The scoring method adopted by the present study (i.e. scoring percentage phonemes correct) reduces the impact of this possibility. However the design does not unfortunately allow for the ability to separate out errors arising from inadequate storage from those arising from motor programming.

The findings in the present study and those in the studies of Italian children by Dispaldro et al (2009, 2011, 2013a,b) differ from those of Chiat and Roy (2007, 2008), who found word-length effects for both words and non-words by their clinical group: shorter items were repeated more accurately than longer items regardless of lexical status. The findings also differ from the findings from the sample of English children in Dispaldro and colleagues' study (Dispaldro et al., 2011). Interestingly their study used the same stimuli as those used in Chiat and Roy's studies (Chiat and Roy, 2007, 2008). A likely reason is that the present study removed from the analyses word items that were not named by the child and therefore were considered to be unknown. Chiat and Roy (2007, 2008) and

Dispaldro et al. (2011) did not make this distinction in their analyses. Another possible reason is that the present study included four-syllable items. Indeed it was the four syllable items that showed the greatest distinction between the mean scores for the (known) words and non-words (see figure 4-7). The findings by Chiat and Roy (2007; 2008) are discussed further in the following chapter (Chapter 5).

The second set of findings in support of the role of PSTM in non-word repetition but not word repetition came from the results of the correlation and regression analyses. Both known word and non-word repetition accuracy were significantly correlated with span. This result is consistent with a study by Gathercole and Adams (1993) in their assessment of similarly aged children. That NWR and WR both correlate with the span task might be because all three tasks require similar skills in addition to those under investigation, such as ability and motivation to focus attention on a verbal task, motivation to attempt repetition, and the ability to articulate sounds in sequence. Of interest was whether PSTM influenced NWR and WR after these task similarities had been accounted for. Therefore, variance shared between NWR and WR was partialled out and residual correlations were explored between the span task and NWR and WR. As predicted, this resulted in a significant correlation being maintained between NWR and span, but not between WR and span. This was the case whether the WR and NWR tasks were corrected for speech errors or not (see appendix K), which reinforced the robustness of the findings. The findings were additionally supported by the regression modelling, providing strong support for the view that PSTM is involved in the repetition of non-words but not known words. This finding is especially interesting given that the non-words used in the present study were formed from re-ordering the syllables of the real words thereby creating non-words which were similar to the real words in phonotactic properties. Despite this, the strong relationship between non-words and span remained.

The presence of ceiling effects at age 4 years is surprising, as NWR has been found to be sensitive as an indicator of language difficulties for older children and adults (e.g. Archibald and Gathercole, 2006; Gallon, Harris and van der Lely, 2007; Marshall and van der Lely, 2009). There are four possible explanations given here for the ceiling effects found in the present study. First, non-words were based very closely on the real words. No independent measure of wordlikeness was obtained, but there was no significant difference between the phonotactic frequency scores for the non-words compared to the words (see chapter 2.2). Furthermore, the non-words were recreated purely by reorganising the syllables from real words, necessarily making them wordlike. It might be that this wordlikeness supported the children's repetition of these; sub-lexical phonological templates were more available to them meaning that there was less reliance on PSTM. Clinically, the finding has implications for the sensitivity of tests of NWR where the non-words are so closely matched to real words in identifying difficulties in older children.

A second explanation for the presence of ceiling effects relates to the careful design of the present study to eliminate speech errors from the scoring. It might be that other studies using less rigorous methods to avoid speech errors in their scoring, count errors of articulation and phonology as repetition errors in children's non-word repetition. Lexical items might be more prone to these speech errors as the words become longer.

A third explanation is in relation to the methods of scoring used in this study. Many tests of NWR use an all-or-nothing method to score the children's responses. This means that a single phoneme error leads the whole item to be scored as incorrect. Therefore each single phoneme error would have more profound effects on the overall score than in the methods used here. In contrast, the percentage phoneme correct method is more likely to lead to scores clustering around the top.

A final explanation for the findings might link to the methods of analysis used in the present study. As was discussed in the methods, it was decided to combine the sample of children recruited through SLT clinic and those who had no identified speech or language needs. It might be that splitting the groups would reveal an item-length effect for the clinical group's repetition of the non-words and this provides the rationale for splitting the groups in the next chapter (Chapter 5).

To summarise, results from the correlation and regression analyses at both phases implicate the role of PSTM in NWR but not in WR. The robustness of this finding is confirmed by the finding of item-length effects for non-words in the data at T1.

4.6.4. Summary of Part 1: Role of word knowledge, phonological short-term memory and speech sounds in NWR and WR

The results of this study show that, in children aged 3 and 4 years, both the ability to hold phonology in STM and the ability to temporarily activate lexical and sub-lexical representations in long term memory affects NWR. In the case of words that are known, children draw upon their stored lexical representation of the items. In the case of non-words, they likely draw upon sub-lexical knowledge, such as phonotactic probabilities and phonological templates. However at 3 year olds, this sub-lexical information is not sufficient to eliminate item-length effects, which are due to the overburdened PSTM system for the unfamiliar phonological information. The study found an effect of speech errors: children's repetition scores on both WR and NWR tasks were improved following correction of these. As noted in section

4.6.2., there was some evidence that these speech errors affected different length items differentially (see appendix K for uncorrected speech word-length effects).

4.6.5 Theoretical and clinical implications

These findings emphasise the need for studies involving NWR to make concurrent speech assessments and to consider the results of these in their scoring. The findings also caution against using WR as a tool to identify children with language difficulties, as it is apparent that this task measures a combination of 1) the children's knowledge of the words to be repeated and 2) their ability to repeat unfamiliar phonological sequences. Where the former skill is measured, the results are subject to ceiling effects. Instead of using the WR task, word knowledge might better be assessed with a test of vocabulary. Where the latter skill is measured, the task is actually an unknown word repetition task and so the ability to repeat phonological sequences might better be tapped using a NWR task.

The main clinical implication of the results relates to assessment. If NWR tasks are used clinically, then the present study would indicate that they need to be used in conjunction with an assessment of phonology and articulation. This would ensure that any errors on the task are due to inaccurate repetition, rather than delayed or disordered speech development. A second assessment-related implication is in the use of word repetition tasks to tap children's language skills. Findings from this study suggest that where words are known, these are repeated accurately and where they are not known the repetition test may be serving more as a measure of vocabulary knowledge/NWR.

Chapter 5

Part 2: Qualitative and quantitative differences in word- and non-word repetition accuracy by children with and without identified speech and language difficulties

5.1. Introduction

5.1.1. Non-word repetition: performance by children with language disorders

Several researchers have examined the NWR skills of children with spoken language disorders, using a range of different NWR tests in English (e.g. Archibald and Gathercole, 2006; Bishop et al., 1996; Chiat and Roy, 2007, 2008; Conti-Ramsden, 2003; Conti-Ramsden et al. 2001; Dollaghan and Campbell, 1998; Gathercole and Baddeley, 1990a; Marshall, Harris and van der Lely, 2003; Marshall and van der Lely, 2009; Marton and Schwartz, 2003). The studies have consistently shown that these children's NWR abilities are inferior to other children the same age and to children matched for language ability. Similar results have been found in languages other than English, for example, Dutch (de Bree, Rispens and Gerrits, 2007), French (Thordardottir, Kehayia, Mazer, Lessard, Majnemer, Sutton, Trudeau, Chilingaryan, 2011) Italian (Dispaldro et al. 2013a), Slovak (Kapalková, Polišenská and Vicenová, 2013), Spanish (Girbau and Sckwartz, 2007) and Swedish (Kalnak, Peyrard-Janvid, Forssberg and Sahlén, 2014).

Although it is clear that cross-linguistically children with language difficulties perform more poorly on tests of NWR compared to their peers (although, see Stokes, Wong, Fletcher and Leonard, 2006), it is not clear why this is. Three hypotheses are discussed:

i) NWR tests tap PSTM which is impaired in children with language difficulties.

ii) NWR tests tap existing language knowledge, which is impaired in children with language difficulties

iii) NWR tests tap a combination of PSTM and existing language knowledge, both of which are impaired in children with language difficulties

In support of the first hypothesis, an early study by Gathercole and Baddeley (1990a) compared a small sample of 7-8 year old children who had identified language disorders with a sample of age-matched and language-matched control participants. They found that the clinical group showed impaired performance on both a NWR and a span task compared to both of the control groups. In the case of non-words, the reported effect was explained entirely by performance on the 3- and 4- syllables non-words. This was seen as evidence in support of impaired PSTM in children with language difficulties, the longer non-words requiring more resources from the PSTM system.

Impaired PSTM cannot however explain entirely the findings of a later study by Archibald and Gathercole (2007). They compared children (aged 7-13 years) with and without SLI on their ability to repeat non-words and equivalent strings of single syllables. The authors demonstrated that the children with SLI were disproportionately impaired on the NWR task compared to the span task. This is an important finding to the present study, as it emphasises linguistic or para-linguistic influences on performance by children with language difficulties. However, the finding was not replicated in a sample of French children with SLI (Leclercq, Maillart and Majerus, 2013). Furthermore, it is difficult to draw out the implications of the finding by Archibald and Gathercole (2007): which difficulty (if either) is causing the other? Does NWR reflect an aspect of language processing that is deficient in children with language disorders and this affects their language acquisition (e.g. Gathercole and Baddeley, 1990a; Gathercole et al., 1992)? Or does NWR tap existing language knowledge and children with language disorders have less of these resources to draw upon (e.g. Graf Estes et al., 2007)?

5.1.2. Word repetition: performance by children with language disorders

So far evidence has been presented in support of PSTM and lexical influences that potentially cause children with language difficulties to repeat non-words less accurately than their peers. Further support for the lexical hypothesis comes from studies examining word repetition as well as non-word repetition accuracy. A few studies have shown that young children with identified speech and language difficulties also repeat these real words less accurately than their language-unimpaired peers in English (e.g. Chiat and Roy, 2007) and in Italian (e.g. Casalini et al., 2007; Dispaldro et al., 2013a, b). Findings from each of these studies will be described in the following account.

Casalini et al. (2007) compared Italian children aged 5-8 years old (divided into two age brackets: “pre-school” and “first grade”), with and without a language disorder, on their repetition of non-words and words. The non-words were split into two separate groups: those that contained Italian grammatical morphemes, and those that did not contain grammatical morphemes. They found that all of the children repeated words more accurately than non-words with familiar morphemes, and that both types of stimuli were repeated more accurately than the non-words without familiar morphemes. The children in the group with SLI showed impaired performance across all measures and there was no significant interaction between group and type of stimulus. Additionally they compared performance by children with different sub-types of SLI. When analysed separately, there was no significant difference in performance by the different clinical groups on their repetition of the stimuli in either the younger or the older sample. These findings are important for the present study as they provide rationale for including a heterogeneous sample of children from SLT clinic.

In another study published in 2007, Chiat and Roy compared young children's (aged 2 ½ -4 years) performance on a word repetition test and phonologically matched non-word repetition test. Their very large sample of 483 participants included 168 children who were receiving speech and language therapy. Their findings were comparable to those of Casalini et al, (2007): the clinical sample showed impaired performance across both lexical categories. It could therefore be that the children's impaired language system was influencing their performance, rather than a difficulty with PSTM. An interesting further finding of Chiat and Roy's study was that the clinical sample (but not the control group) showed similar item-length effects in their repetition of words as they did in their repetition of non-words (i.e. shorter words were repeated more accurately than longer words). Two possible interpretations for this finding are:

> 1) Longer words are more difficult to repeat for children with language difficulties due to phonological properties of the words, making these more difficult to articulate.
>
> 2) As the 'word length effect' is usually associated with limitations in PSTM capacity the clinical group were using their PSTM for word repetition as well as non-word repetition.

The second interpretation (above) is consistent with the explanation given by Vance et al. (2005) for performance by the 3 year olds in their study. However, an alternative explanation might be that the children in Chiat and Roy's study did not know the words that they were repeating and therefore these were treated the same as the non-words. Dispaldro et al. (2013a) report similar performance by older Italian children and this study will be discussed in more detail below.

Dispaldro and colleagues (2013a,b) compared repetition performance by 4-6 year old children who had a diagnosis of SLI with younger, language unimpaired peers (Dispaldro et al., 2013b) and with age-matched peers (Dispaldro et al., 2013a). They used both real word and non-word stimuli of 1-4 syllables, assuming the children's knowledge of the real words based on pre-determined age-of-acquisition norms. Their results showed generally poorer performance by the clinical group on both WR and NWR, with some evidence of a greater impairment for NWR (depending on scoring method adopted). Compared to the age-matched peers (but not the language-matched peers), the clinical group additionally showed greater difficulty as the lexical items (words or non-words) increased in length.

Frustratingly, Dispadro and colleagues did not report whether a 3-way interaction was present. However, interestingly the authors report a correlation between the NWR and WR for the clinical group but not for the control group (Dispaldro et al. 2013a). Taken together, the effects relating to lexical length and the significant correlation found between the tasks for only the clinical group raises the question again about the group's knowledge of the vocabulary. The children in the clinical group are known to have difficulties with language-learning, so it could be argued that a lack of familiarity with the real word vocabulary might have led them to treat the real words in the same way as non-words.

In summary, studies have shown that young children with language difficulties repeat real words less accurately than their peers. Therefore, these studies have suggested that WR as well as NWR could serve as a marker of language impairment. Two possible interpretations are provided for the findings of the existing studies:

i) Children with language difficulties repeat real words less accurately than their peers because word repetition taps existing vocabulary and children with language difficulties do not have the assessed-words in their vocabularies.

ii) Children with language difficulties repeat real words less accurately than their peers because they cannot access the stored representations in their lexicons and therefore must rely on their PSTM.

5.1.3. Aims of the study

This part of the study aimed to investigate whether a sample of pre-school children (aged 3 and 4 years) who had identified speech and language difficulties at the first data-collection point would show the same pattern of effects in their repetition of known words and non-words as their language-unimpaired peers do. The study differs from previous studies that have investigated word repetition by preschool children in that it included only real words that the children definitely knew. This was assessed by their ability to correctly name a picture representing the stimulus. The present study was therefore interested to discover whether children with identified speech and language difficulties draw upon their stored lexical knowledge during repetition of known words.

5.1.4. Hypotheses

If children with language difficulties draw upon lexical knowledge during repetition of known words, the following would be expected:

- The clinical sample would repeat known words with equal accuracy to the non-clinical sample, as they are using their existing knowledge of these words.
- Item length effects would be present for both groups for non-words (i.e. shorter non-words would be repeated more accurately than longer non-words for both the groups), as children draw upon PSTM during the repetition of these items.
- Item length effects would not be present for known words for either group, as children draw upon their existing word knowledge during repetition of these items.

5.2. Methods

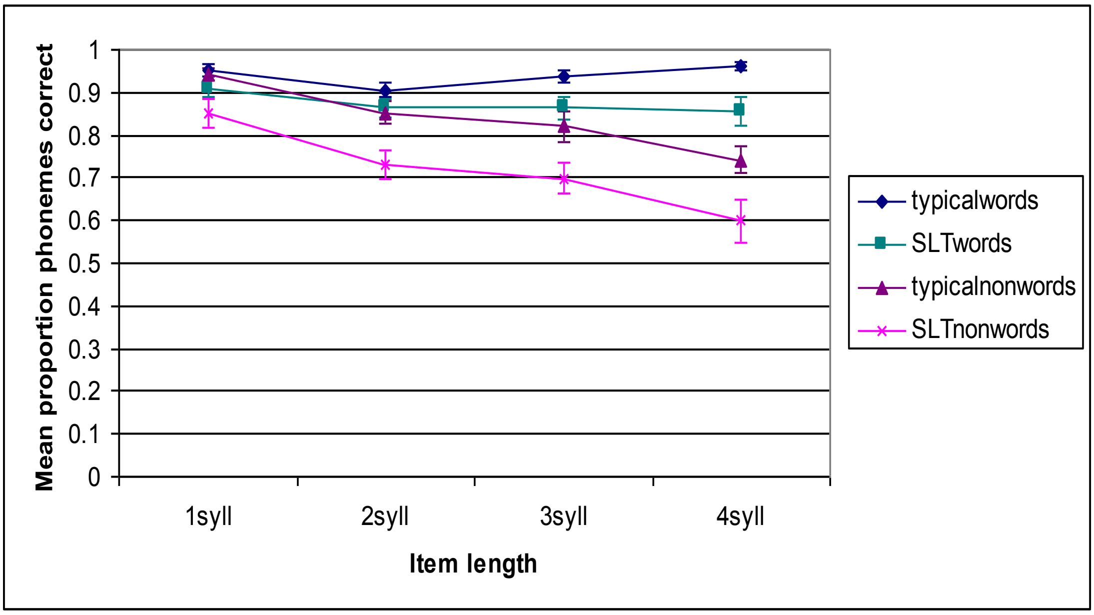
This part of the study used identical methods to part 1 (see chapter 3). However for this part of the study, the participants were grouped according to whether they were known to speech and language therapy (SLT) for speech and/or language needs at the T1. Those who were known to SLT represented the clinical group and those who were not known to SLT were categorised as the non-clinical group. This method of categorisation did not take into account scores on the standardised language measures. Indeed, there was some overlap in these scores (reported in chapter 2).

This part of the study used the corrected speech scores (see chapter 2 and 3) in its analyses to explore the specific hypotheses presented above.

5.3. Results

5.3.1. Results at T1

Figure 5-1 Mean scores on the WR and NWR tasks for each item length (for ease of reading, error bars represent standard error, rather than standard deviation from the mean)



(SLT=clinical group, typical=non-clinical group)

To explore differences in repetition accuracy by the groups on known words and non-words of different lengths, a mixed design ANOVA was performed. Group (clinical or non-clinical) was the between subject variable and lexical type (known word or non-word) and length (1, 2, 3 or 4 syllables) were the between-subjects variables. Some of the children had not been able to name any words at some of the word-lengths, meaning that they did not have the opportunity to obtain a repetition score for some of the word lengths. These children's data were not included in the analysis. This meant that the clinical group had a sample size of 22 participants and the non-clinical group had a sample size of 23 participants. Mauchly's test of sphericity was significant, indicating that the variances of the differences between levels are unequal. As the assumption of sphericity was violated, Greenhouse-Geisser estimates are reported. Partial eta squared values are given as the effect size estimates.

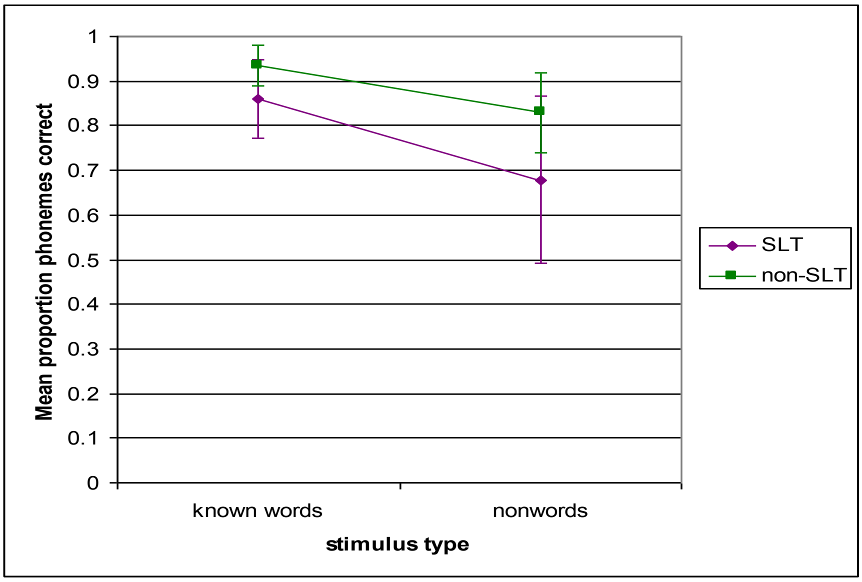
<u>Main Effects</u>

The analysis showed a significant main effect of group, $F(1, 46) = 17.12$, $p < 0.001$, $\eta_p^2 = 0.271$. The non-clinical children (M=0.88, SD=0.13) repeated the items more accurately than the clinical children (M=0.77, SD=0.13). There was a significant main effect of lexical type, $F(1, 46) = 63.47$, $p<0.001$, $\eta_p^2=0.580$. Known words (M=0.90, SD=0.07) were repeated more accurately than non-words (M=0.75, SD=0.15). A significant main effect of 'word' length was also found, $F(3, 46)=24.34$, $p<0.001$, $\eta_p^2 = 0.346$ and the planned comparisons associated with this were reported in chapter 4 (section 4.5.3.1).

<u>Interactions</u>

There was a significant interaction between group and word type: $F(1, 46)=4.34$, $p=0.043$, $\eta_p^2 = 0.086$, indicating that the groups performed differently on the different types of stimulus. The mean group values are plotted on the graph in figure 5-2.

Figure 5-2 Group mean proportion scores for known words and non-words



Post-hoc paired sample t-tests (applying a Bonferroni correction such that the test is considered significant if p<0.0125[2]) revealed that both groups were more accurate repeating known words compared to non-words (clinical group: words>non-words, t(22)=5.76, p<0.001; non-clinical group: words>non-words, t(24)=5.51, p<0.001).

A post-hoc independent samples t-test revealed that the difference in scores on the known word repetition task and the non-repetition task were both significant (known words: t(46)=14.56, p<0.001; non-words: t(46)=7.39, p<0.001). However the graph (figure 5-2) indicates that the difference between the groups is greater for non-words than for known words (though may be confounded by large standard deviations in the case of non-words).

There was no significant interaction between group and item length (F(3,141)=1.75, p=0.16) and no 3-way interaction (F(3,141)=0.75, p=0.48).

[2] p<0.05 divided by the number of t-tests, i.e. 4

A significant interaction was found between item type and item length ($F(3,48)=12.32$, $p<0.001$, $\eta_p^2=0.211$: non-words are affected by length and known words are not. This has been discussed in Part 1 (see chapter 4).

As the data were not distributed normally (see appendix I), the two groups were also compared on their WR and NWR scores using non-parametric tests (mean proportion phonemes correct). The results are reported in appendix J and were consistent with the parametric tests.

5.3.2. Results at T2

As for the data at T1, a mixed design ANOVA was carried out, using the same between group and within group variables. As for T1, data were excluded where children had not been able to name the word stimuli at any length. At T2, the clinical group had a resulting sample size of 24 participants and the non-clinical group had a resulting sample size of 26 participants.

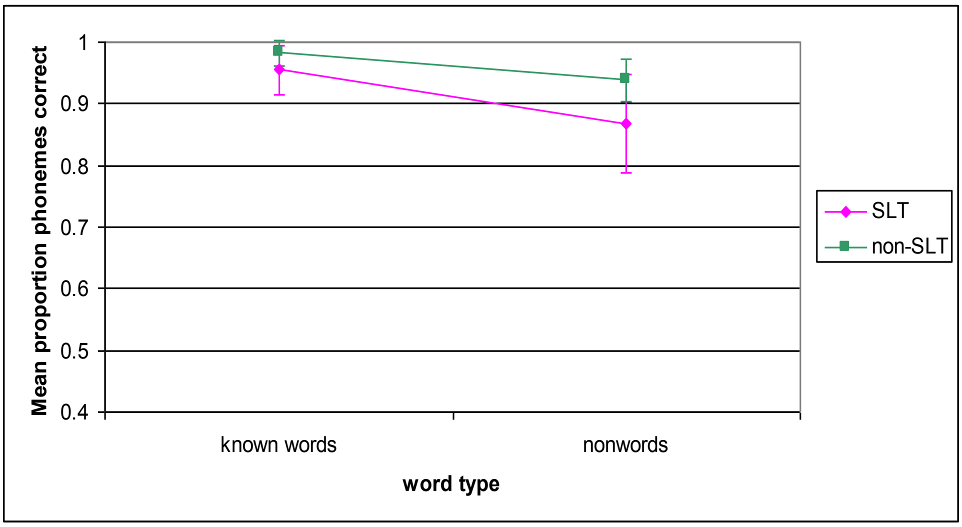
Main Effects

The analysis showed a significant main effect of group, $F(1, 48)=17.56$, $p<0.001$, $\eta_p^2=0.268$. The non-clinical children (M=0.96, SD=0.06) repeated the items more accurately than the clinical children (M=0.91, SD=0.06). There was a significant main effect of lexical type, $F(1, 48)=81.40$, $p<0.001$, $\eta_p^2=0.629$, known words (M=0.97, SD=0.03) were repeated more accurately than non-words (M=0.90, SD=0.06). A significant main effect of 'word' length was also found, F(3, 48)=13.465, p<0.001, η_p^2 =0.219 and the planned comparisons associated with this were reported in chapter 4 (section 4.5.3.2).

<u>Interactions</u>

As for T1, a significant interaction was found between group and word type, F(1, 43)=8.71, p=0.005, η_p^2 =0.154, indicating the groups performed differently on the different stimuli. The mean group values are plotted on the graph in figure 5-3.

Figure 5-3 Mean proportion scores for known words and non-words (note that the y-axis has been manipulated to emphasise the group x 'word' type interaction)



Post-hoc paired sample t-tests (applying a Bonferroni correction so that results were considered significant if p<0.0125) revealed that both groups were more accurate repeating known words compared to non-words (clinical group: words>non-words, t(23)=6.55, p<0.001; non-clinical group: words>non-words, t(25)=6.60, p<0.001).

A post-hoc independent samples t-test revealed that the difference in scores on the known word repetition task and the non-repetition task were both significant (known words: t(48)=3.14, p=0.003; non-words: t(48)=4.05, p<0.001). As for T1, however the graph suggests that the interaction is explained by a greater effect for non-words than known words.

There was no significant interaction between group and item length ($F(3, 46)=1.80$, $p=0.150$) and there was no 3-way interaction ($F(3, 46)=0.73$, $p=0.534$).

There was an interaction between item type and item length ($F(3, 46)=5.86$, $p=0.001$, $\eta_p^2=0.109$) and this has been discussed in Part 1 (chapter 4).

As the data were not distributed normally, analysis was also made using non-parametric tests. The results of these were consistent with the parametric tests and are reported in appendix J.

5.4. Discussion

5.4.1. Recap of the study's aims

This part of the study aimed to investigate whether children with identified speech and language difficulties show the same pattern of effects during repetition of known words and non-words as do their language-unimpaired peers. It aimed to tease out whether the reported impairment in word repetition by language-impaired children (e.g. Casalini et al., 2007; Chiat and Roy, 2007; Dispaldro et al. 2013a, b) could be due to the children not knowing the words to be repeated. The present study differed from the other studies in its inclusion only of real words that were definitely known by the participants. The predictions were that the clinical group would perform equally as well as the non-clinical group on their repetition of words that they knew (following speech error correction), but that they would perform more poorly than their age-matched peers when repeating non-words.

5.4.2. Results summary and interpretation

The results indicate that the participants who are known to speech and language therapy perform less well overall than the non-clinical group at both time-points. An interaction was found between group and word type, which showed a trend towards the clinical group performing disproportionately less well on non-words compared to known words. This would be consistent with the study's hypothesis; that children with language difficulties do not have a difficulty repeating familiar lexical items, but that they have a limited store of these. This limited store results in them having less well defined sub-lexical templates to support their storage of unfamiliar lexical items during non-word repetition (see also Chiat 2006 for a similar explanation).

The interaction between 'word' type and length was explored in chapter 4 (Part 1). The present study found no interaction between group and item length and no 3-way interaction (group, item type and item length). This is consistent with the predictions: children with language difficulties draw upon lexical knowledge in a similar way to their peers when repeating known words of different lengths.

5.4.3. Results in relation to other studies

The findings should be considered in the light of those of Casalini et al. (2007), Chiat and Roy (2007) and Dispaldro et al. (2013a,b). All of these studies recruited young participants with identified speech and language difficulties and they compared repetition performance (words and non-words) against a control group. Similarities and differences in the results of the studies will be discussed in turn.

Consistent with the present study, Casalini et al. (2007) found that pre-school children with language difficulties repeat words and non-words less accurately than their peers. Unlike the present study, theirs did not find a significant interaction between the group and the type of stimulus. Instead, the effect was present regardless of whether the children were repeating real words, non-words that used familiar morphology or entirely unfamiliar non-words. Therefore the results from the study by Casalini and colleagues suggested that all of the repetition tasks could distinguish clinical from non-clinical groups and that presumably all the tasks were drawing on similar skills. Casalini and colleagues did not investigate item length effects.

Chiat and Roy (2007) also indicated repetition difficulties for both words and non-words by their clinical sample compared to the control participants. However, they analysed the results of the two groups separately so direct comparisons between the groups cannot be made. In their separate analyses, Chiat and Roy (2007) reported that the control sample showed item length effects (i.e. longer items were repeated less accurately than shorter items) for non-words but not words. The clinical sample, by contrast, showed item length effects for both words and non-words. Their study indicated therefore that the clinical group were using similar skills in their repetition of real words and non-words, whereas the non-clinical group were presumably using their lexical knowledge in their repetition of words, but not for non-words. The present study did not find this 3 way interaction[3]. Instead it showed that the clinical group showed equivalent patterns of performance to the non-clinical group for real words of different lengths and non-words of different lengths. While both groups showed word length effects, indicating the role of PSTM, the clinical group showed lower scores across all syllable lengths to the non-clinical group. This again indicates that the clinical group were not able to benefit from the same sub-lexical processing that the non-clinical group do.

[3] Although the present study did not find a 3-way interaction as suggested by the findings in Chiat and Roy (2007), there was a tendency in the same direction (see figure 5-1).

The study by Dispaldro et al. (2013a) used two different methods of scoring. Of most interest to the present study are their results when they scored proportion phonemes correct, as these are directly comparable to the present study. Their study revealed some similar findings to the present one: children with language difficulties repeated both words and non-words less accurately than the language-unimpaired peers, but a significant group by ‘word’ type interaction showed that the clinical sample performed disproportionately more poorly when repeating non-words. This was also the tendency in the present study, but it was not confirmed by post-hoc analysis. Their study however also yielded some different results from the present study. The first difference related to the interaction between word type and word length found in the present study. Dispaldro and colleagues found no such interaction. Instead both shorter words and non-words were repeated more accurately than longer words and non-words. This finding is inconsistent with the hypothesis that children are using their stored lexical representations for the words. If this is the case, no word length effect should be present for real words. A further difference in their findings was a group by word length interaction: children with language difficulties were affected more by the length of the items than the language-unimpaired group. No such interaction was revealed in the present study. As discussed previously, the authors did not report whether their results revealed a 3-way interaction, which is a disappointing, as this might have proven very informative.

The various differing findings by Casalini et al. (2007), Chiat and Roy (2007) and Dispaldro et al. (2013a) compared to the present study might all partially be explained by the exclusion of ‘unknown’ words from the word repetition task in the present study. If the children in the other studies did not know some of the words used in the repetition task, this would mean that they would have no choice but to treat these as non-words. This might explain the word length effects (for real words) reported in the latter two studies as well as the group by word-length interaction and the non-interaction between word type and word length both reported in Dispaldro and colleagues' study. Children in the clinical groups are less likely to have known the real words than the children in the non-clinical groups. Therefore the explanation would be upheld if the studies had shown 3-way interactions, whereby clinical groups were showing similar word length effects for words and non-words but the non-clinical groups showed these only for non-words.

Counter to the above argument, the study by Dispaldro et al. (2013a) did attempt to ensure the children's knowledge of the words used in their study. They selected words that "were assumed to be to be known by preschool children, based on norms reported in Barca, Burani, and Arduino (2002)" (p328). However they did not explicitly test the children's word knowledge and so it might be that the clinical group, in particular, were unfamiliar with the words.

Another possible reason for the discrepancy in findings among the studies is the use of different NWR and WR tests. It is apparent that different tests draw on different skills depending on the composition of the non-words used (Gathercole, 1995; and see Graf Estes et al., 2007 for a review). An example of how this can affect children's performance is illustrated in a small study by Archibald and Gathercole (2006). They compared children's performance on two different NWR tests: the Children's test of Nonword Repetition (CNREP, Gathercole et al., 1994) and the Nonword Repetition Test (NRT, Dollaghan and Campbell, 1998). They assessed children with SLI (aged 7-11 years), a group of age-matched control participants and a group of younger children who were matched for language skills. The study found that children with SLI performed more poorly than both of the other groups on the CNREP task (when non-verbal skills were controlled).

However, the children performed more poorly than the age-matched group only (not the language-matched group) on the NRT.

In their meta-analysis, Graf Estes et al. (2007) confirmed the findings reported by Archibald and Gathercole (2006): the difference between scores by children with SLI compared with language unimpaired peers is greater for the CNREP compared to the NRT. They summarised four possible variables that cause the difference in performance: i) the CNREP contains several long non-words (up to five syllables; ii) the CNREP includes non-words which contain later developing phonemes and clusters of phonemes, therefore making this more phonologically complex; iii) the CNREP adopts a whole-item scoring method, meaning that a single phoneme error leads to the whole item being scored as incorrect; iv) the CNREP contains non-words which are judged to be more word-like than those used in the NRT. It is possible that any of these same variables might have caused the disparity between the findings of the present study and those of the other studies.

In summary, the balance of the evidence points to the differential performance between the clinical and non-clinical groups being more marked in the case of NWR than WR. This was a tendency found in the present study in which, unlike others, known words were distinguished form unknown words. Furthermore, the interaction between word-type and word-length found here for both groups, suggests that children with language difficulties draw upon lexical knowledge in a similar way to their peers when repeating known words of different lengths. Evidence therefore suggests that NWR tasks are more effective than WR for distinguishing clinical from non-clinical groups.

5.4.4. Possible confounding variables

The results of the present study should be considered with caution. First, ceiling effects were present in the data. Ceiling effects are unfortunately impossible to prevent in the case of repetition of known words. This is because the present study took a binary approach to assessing children's word knowledge: either there was evidence that a child knew the word (they could label a picture representation of it), or there was not. The method of scoring adopted, whereby phonological errors were corrected further meant that any speech errors could not reduce the ceiling effects. Ceiling effects might result in the illusion of an interaction between variables, where none is present.

A second possible confounding variable is the way in which the groups were allocated. The groups were split only according to whether the children were known to speech and language therapy or not. As presented in the general results chapter (chapter 3), there was overlap between the groups in the children's scores on the PLS-4 assessment. This was due in part to the inclusion into the study's clinical group of children who had phonological difficulties only. This also led to a highly heterogeneous clinical group. The advantage of this is that the sample is more generalisable to other clinical populations. Indeed, it is striking that many of the results were consistent with some previous studies, given the methods of grouping the participants.

5.4.5. Theoretical Implications

From a theoretical perspective, the results of the present study support the proposed model of repetition (chapter 1) to some extent. Children with speech and language difficulties draw upon long-term lexical storage during repetition of familiar words in a similar way to their language unimpaired peers. This is evidenced by the item length effects found for non-words but not known words for both groups. It is possible to conclude that during repetition of known words, children do not draw upon their PSTM, whereas for non-words they do.

However, if children with speech and language difficulties were able to draw upon their lexical knowledge during repetition in exactly the same way as their peers, a group difference in known word repetition would not be expected. This was however found in the present study. The finding cannot be explained by any consistent phonological difficulties present for the children, because these were corrected (based on the children's naming). Two other explanations are therefore proposed. The first explanation is that some of the clinical sample may have had *fuzzy* or ill-defined phonological representations for some of the word stimuli. This would mean that their production of these words might be inconsistent. The present study aimed to identify any children presenting this way, by requiring the participants to engage in the naming task twice. This method of assessment, being non-standardised, might not however have identified children with inconsistent phonological disorders. Indeed, while Dodd estimates that children presenting with this type of disorder represent around 9-10% of the population (Broomfield and Dodd, 2004), no children in the sample were identified to be presenting this way.

An alternative explanation for the clinical group performing less well when repeating words that they knew might be that some of the clinical sample did not recognise some of the words when they were presented out of context as part of the repetition task. It should be remembered that a repetition task necessarily requires different skills to a naming task. Children with language difficulties are classically more adept at processing information visually than verbally. It could therefore be that during a picture naming task, the presence of the picture facilitates the child's retrieval of the phonological label. It might be that the same is not true when the child hears the word (i.e. a visual or semantic representation is not retrieved). If so, this would mean that the children would process these words at a shallower level of processing, through a sub-lexical route, treating these unrecognised words as non-words during repetition. This is possible in the Stackhouse and Wells (1997) model presented in chapter 1, and is also not impossible in the novel model proposed in this study. However, according to the present study's model, if this is the case an item length effect would be present for the known words processed this way, as they would require more resources from PSTM (or would be subject to motor programming errors - see chapter 4). Figure 5-1 does indicate that there is a possible tendency in this direction, but analyses were not pursued as no 3-way interaction had been identified

5.4.6. Clinical Implications

Results from the study are consistent with the wealth of previous studies that have shown that children with language difficulties show considerable impairment on NWR tests in comparison with age-matched peers. The results of the present study suggest that non-words of 2 syllables or more are particularly useful. The results would caution against using WR tests as a tool to identify children with language difficulties. This is because the present study found that where children know the words, their repetition tends to be accurate. Where they do not know the words, they treat these as non-words and therefore the task does not differ from a NWR test. Performance on the task is therefore confounded by a given child’s familiarity with each test item.

Chapter 6

Part Three: The Influence of grammatical knowledge and phonological short-term memory on children's repetition of sentences

6.1. Chapter overview

This chapter discusses the influence of existing language knowledge and PSTM on sentence repetition (SR). It presents data from T2 in the study, when the children were 4 years old.

6.2. Introduction

It is widely documented in the research literature that children's ability to repeat sentences identifies those children who have language difficulties (e.g. Conti-Ramsden et al., 2001) and some evidence that accuracy on the task also identifies those people with a history of spoken language difficulties but whose difficulties appear to have resolved (e.g. Moll, Hulme, Nag and Snowling, 2015). Clinically, several pre-school speech and language therapy assessments incorporate different versions of the task, e.g. the *CELF P2* (Wiig et al., 2004), the *PLS -4* (expressive) (Zimmerman et al., 2002), the *ERB* (Seeff-Gabriel et al., 2007), the Grammar and Phonology Screening test (van der Lely, Gardner, Froud and McClelland, 2007) and the Test of Language Development (Newcomer and Hamill, 2008). Over the past decade a wealth of research studies has investigated which mechanisms influence children's performance on their repetition of sentences. These studies have focussed their attention on two main influences on sentence repetition, which the present study will also explore: verbal short-term memory and existing language knowledge. Each of these factors will be considered in the account that follows.

6.2.1. Involvement of verbal short-term memory in sentence repetition

Sentence repetition is an immediate recall task, so a logical leap is to assume that it involves verbal short term memory. Indeed, several studies have put forward evidence for the role of PSTM to the task (e.g. Alloway and Gathercole, 2005; Alloway, Gathercole, Willis and Adams, 2004; Baddeley, 1986; Willis and Gathercole, 2001). If sentence repetition is influenced by the PSTM system then the properties of this system should be reflected in performance on task. Some studies have investigated factors that affect sentence repetition and are also generally considered to be properties of the PSTM system.

6.2.1.1. Properties of PSTM reflected in SR

As discussed in the general introduction (figure 1-4 and section 2.2.5.), PSTM is typically measured by requiring participants to repeat lists of single syllable words. To recap, accuracy of word list recall is known to be affected by the following factors: the length of the words to be recalled, the position of the words to be recalled, whether the participant is able to rehearse the words, and the phonological similarity of the words to be repeated. These factors affecting list recall are thought to reflect properties of the PSTM system. These properties are explored in relation to sentence repetition in the following account.

The word length effect, i.e. the tendency for poorer recall of word lists when these contain longer items, is thought to reflect the limited capacity of the phonological loop. Willis and Gathercole (2001) showed that this effect was also found in sentence recall. They aimed to investigate the contribution of PSTM to children's (aged 4 - 5 years) repetition and comprehension of sentences. Recognising that the children's language systems were not yet fully developed, Willis and Gathercole (2001) hypothesised that the young children might rely more on PSTM than adults during sentence recall. They hypothesised alternatively that the children's sentence processing might be constrained by their limited, developing PSTM. They manipulated sentences by altering the number of syllables contained in the nouns. The authors also manipulated the sentences linguistically: varying these according to six different syntactic structures. Their study, drawing on data from 30 children, revealed that both sentence length and sentence type affected performance. Children showed more difficulty on longer sentences and they made more errors when the sentences contained embedded clauses and relative clauses. Therefore, while their study provides some evidence for PSTM, it also emphasises that grammatical skills are also important.

A second property of the phonological loop that has been investigated in relation to sentence repetition is the effect of word position. In a list of words to be repeated, those that occur at the start and at the end of the list are recalled more accurately. These effects are known as primacy and recency effects respectively. Alloway and Gathercole (2005) provide some limited evidence in support of primacy, but not recency effects for words during sentence repetition. They found that children aged 4 - 5 years were less likely to make errors on words that occurred at the start of sentences than in medial or final positions. Baddeley, Hitch and Allen (2009) found similarly that recall of constrained sentences by adults showed much less pronounced effects of word position than that which is characteristic of a word list task. The constrained sentences used in their study deliberately included words selected from a limited word-set, and which were presented without prosody and with restricted grammar. The purpose of this was to make the task more similar to the span task. Both studies showed some evidence of effects of word position, but much less so than for a span task.

A third property of the phonological loop is the effect of articulatory rehearsal and disruption to recall when rehearsal is prevented. This is well documented in list recall (Gathercole and Baddeley, 1995). Evidence that this also occurs in sentence recall comes from the paper described above by Baddeley and colleagues (2009). They required adults to perform a sentence recall task and while the sentence was presented, the participants were required to simultaneously articulate “1-2-3-4”. The authors found that this dual-task requirement disrupted sentence recall similarly to list recall. This was seen as evidence for the involvement of the phonological loop. However in their study, they also reported disruption to both the list recall and the sentence repetition task when participants divided their attention between these tasks and tasks thought to tap other components of the WM system (visuo-spatial sketchpad and central executive). Activation of the visuo-spatial sketchpad and central executive should not occur during simple phonological loop tasks (i.e. a list recall task). This therefore suggests either that the interference may be caused simply by dividing attention, or that these other components of the WM system are required for the tasks.

The final property of the PSTM system described above, the phonological similarity effect, is reported to impair accurate recall of sentences similarly to the way it affects recall of word lists (Baddeley, 1986). However, while those in support of the WM model view this as a property of the PSTM system, others consider the language system to be responsible for the effect. Acheson and MacDonald (2009), for example, note that speech errors made in the articulation of tongue twisters are similar to those produced during recall tasks.

6.2.1.2. Different SR accuracy by high and low PSTM groups

A different method of assessing the influence of the PSTM system on sentence recall is to compare performance on a sentence repetition task by participants with high and low span scores. One such study is that by Alloway and Gathercole (2005). They recruited children aged 4–5 years and grouped them according to high or low PSTM scores (a combined measure of span and NWR). The children were matched on non-verbal assessment, though unfortunately language skills were not assessed. Participants repeated sentences of two types: syntactically simple and syntactically complex. The authors found that children with high PSTM scores repeated both sets of sentences more accurately than the other group, supporting previous studies' findings that there is a role of PSTM in the task. They also examined the type of the errors in the sentence. They found that the children with poorer PSTM were more likely to make errors of omission, addition, errors of word order and they were more likely than the other group not to respond (Alloway and Gathercole, 2005). The authors interpreted this result as indicating that PSTM might have the role of maintaining the structure of the sentence, i.e. word order.

There are two connected criticisms with this interpretation of the study described above. First, word order is generally dictated in English by the verb and to the linked argument-structure to that verb. In simple sentences, like those used in this study, there are limited opportunities to alter the order of the sentence, while conforming to correct verb-argument structure. Second, the children in the study were not assessed on their language skills and this limitation is acknowledged by the authors. It might be that several children presenting with low PSTM in the study, were also experiencing language difficulties. Where these difficulties affected the children's grammar, this would very likely lead to omission of function words and word order errors during repetition.

6.2.1.3. Summary of the role of PSTM in SR

So far, discussion has focussed on the involvement of PSTM and to a lesser extent wider WM in the recall of sentences. While there is some evidence for the influence of PSTM on SR, this is far from clear. Studies have highlighted problems with this interpretation and have instead emphasised the contribution of language knowledge on repetition. It is evident that the span task, classically used to measure PSTM clearly differs from the recall of sentences, where individual words have semantic, syntactic and prosodic relationships with each other. Discussion now turns to the other end of the spectrum; to studies that have emphasised language skills in the repetition of sentences to the exclusion of PSTM.

6.2.2. Influence of existing language knowledge on SR

6.2.2.1 Sentence manipulation: semantics, syntax and prosody

As described in the introduction (chapter 1, section 1.6.), there are different influences of language knowledge on recall of sentences. There is evidence, for example that the amount of information recalled by children aged 4 -5 years old is influenced by the semantic relationships between the words to be recalled (Polišenská, 2011). Polišenská (2011) found that Czech and English children recalled sentences more accurately where these were semantically plausible, compared to sentences that were semantically implausible but where syntactic relationships were respected (e.g. "*I have seen an angel*" compared to "*I have read an uncle*"). Alloway and Gathercole (2005) also demonstrated that semantics influences children's SR in their analysis of 4 -5 year old children's errors. They found that these were more likely to be semantic in nature.

Grammatical influences on children's repetition of sentences are also evident. First, Devescovi and Caselli (2007) observed that children only repeat sentence structures that are found in their spontaneous speech. The finding predicts that a sentence stimulus containing morphological or syntactic structures that a child has not yet mastered would render his/her repetition errorful. Seeff-Gabriel, Chiat and

Dodd (2010) found exactly this in the case of morphology: children with SLI showed particular difficulties repeating function words and inflections. Presumably the children in the study had not yet mastered these morphemes. This finding was replicated by Riches (2012) during elicited language tasks in a study of older children with SLI. The children in this latter study showed the same morphosyntactic errors in their repetition as they did in their narrative.

The study by Polišenská (2011) described above also investigated the role of syntax in repetition of sentences. The English sentences to be recalled by the 4-5 year old children were manipulated so that the order of the words was altered (e.g. the sentence "*I have seen an angel*" was altered to "*Seen I an have angel*"). This was found to impair children's recall. The finding is confounded however in three ways. First, by manipulating the syntax of the sentence, the overall sentence meaning is also lost. Second, through manipulation of the word order but maintenance of the original prosody of the sentence, words that would normally not receive stress do so and vice versa. Finally, from a constructivist perspective, the syntactic manipulation means that words that are statistically unlikely to occur together do so in the syntactic manipulation (e.g. "an have"). An attempt was made to address one of these confounds. Prosody was removed from the presentation of both types of sentence during two of the study's experimental conditions: rather than using natural intonation, words were presented in a list format. The loss of intonation impaired children's recall, but only to a small degree. This is an interesting finding as it raises the question of whether children continued to encode these sentences using an internal prosody to aid their recall in the case of the syntactically sound sentences. The findings from this study, through its attempts to systematically separate out the influences on children's repetition, highlight the mutually dependent nature of the different multi-dimensional components of language.

6.2.2.2. Syntactic and semantic priming

Potter and Lombardi (1990, 1998) also investigated effects of syntax and semantics. They refuted claims that PSTM explains performance on sentence repetition tasks. Instead they proposed that during the task, the overall meaning of the sentence is preserved at a conceptual level and the form of the sentence is restored using the recently activated semantic and syntactic representations. They referred to this as the *regeneration hypothesis*. Their evidence came from a series of experiments that showed that sentence repetition by adults was subject to semantic (Potter and Lombardi, 1990) and syntactic priming (Potter and Lombardi, 1998). Semantic priming was evident in sentence repetition by 4 year old children. A description of this study follows.

In Potter and Lombardi's study with young children (Potter and Lombardi, 1990, experiment 7), the task involved slightly delayed recall of a sentence. Prior to recall, participants were presented with a list of four nouns, one of which was an approximate synonym to a noun in the sentence to be recalled. For example, the children in their study were presented with the sentence: "my friend got a rabbit for his birthday" and were then presented with four nouns comprising three non-synonyms and the word "bunny". Potter and Lombardi (1990) found that children were vulnerable to recalling "bunny" erroneously, but not the non-synonyms in their repetition of original sentence. In order for these lures to affect the children's repetition, the sentence must have been processed at a deep level (see figure 1-9).

The authors' later study (Potter and Lombardi, 1998) used a similar experimental design to demonstrate the presence of syntactic priming in sentence repetition. In this study they showed that language-unimpaired adults are prone mistakenly to use the structure of the clause they have most recently heard when repeating an earlier clause. For example, participants heard the two-clause sentence: “*Joe fed the baby pudding* [ditransitive] *and sold some diapers to the neighbour* [prepositional dative]”. They were found to make the first clause syntactically consistent with that which had been most recently heard (thus, “*Joe fed the pudding to the baby*”) during their repetition. While the effect is established in adults, the authors did not report on any attempt to replicate this finding with young children, but there is some evidence from subsequent studies that children’s developing language systems respond similarly to syntactic priming (Pickering and Ferreira, 2008; Riches 2012).

6.2.2.3. Summary of the influence of language knowledge on SR

In summary, evidence in support of existing language knowledge influencing sentence repetition comes from two main sources. The first source is studies that have manipulated linguistic aspects of the sentences to be recalled and observed differences in recall accuracy. The second source is studies that have disrupted the memory trace of the sentence through priming. One further piece of evidence that existing language representations are important is the observation from studies involving language unimpaired adults show that when errors occur in sentence recall, there is fidelity to the gist of the sentence (e.g. Jarvella, 1971; Saffran and Martin, 1975). However, caution should be shown in generalising adult data to children, as children’s developing language and wider cognitive systems might differ from more established adult systems.

6.2.3. Combined influence of PSTM and language knowledge on SR

While discussion so far has focussed separately on contributions of PSTM and language knowledge to sentence repetition, perhaps a more comprehensive explanation for performance on the task is that both these bottom-up and top-down systems are involved, i.e. there is a role for both PSTM and existing language. Most of the studies already reported have recognised this dual-contribution. One such study, reported in part earlier (section 6.2.1.1.), is that by Willis and Gathercole (2001), which is described next.

In the second experiment presented in their study, Willis and Gathercole (2001) used a similar design to that by Alloway and Gathercole (2005) (reported earlier). They discussed data from two groups of children; one with low and the other with high PSTM (measured using a digit span and NWR test). These children were matched however on their verbal reasoning and non-verbal skills. The authors investigated the children's repetition and also their comprehension of the sentences. The sentences varied in terms of morphosyntax; sixteen different sentence types were selected from the *Test for Reception of Grammar* (Bishop, 2003). The children were required to repeat the sentence and then to identify the picture that corresponded to the meaning of the sentence. The children scored one point respectively for each sentence correctly repeated and understood. Children with high PSTM performed better on the repetition but not comprehension of the sentences. Willis and Gathercole (2001) concluded that their data provided evidence that PSTM supports immediate recall of sentences in young children. From their analysis of the children's performance on the different types of sentences, they proposed that sentence repetition does depend on access to grammatical knowledge, but to a lesser extent than PSTM.

Further evidence for the combined influence of PSTM and language knowledge in sentence repetition comes from adult patients with neurological impairments resulting in PSTM impairments (e.g. Martin, 1993; Saffran and Martin, 1975). These patients were able to recall the gist of a sentence (due to intact language skills), but not the individual words (due to impaired PSTM). This finding highlights the dissociative contributions of PSTM and language to the task.

Riches (2012) investigated sentence repetition by children with impaired language. He compared performance by children aged 6-7 years with SLI, to age-matched and language-matched control participants. The aim of the study was to establish the contributions of PSTM, WM and syntactic representations to performance on sentence repetition tasks. The study manipulated sentences according to length and complexity. The children were assessed using a syntactic priming task, a NWR task (which was the measure of PSTM), two working memory tasks and a narrative task. The syntactic priming task was found to be the strongest predictor of performance on the sentence repetition task. However, the NWR task was also a significant predictor of sentence repetition among the children with SLI. Riches proposed this, together with a further finding that children with SLI were more affected by delaying their recall of the sentences, as evidence in support of the role of PSTM in sentence repetition. He discussed that children with SLI use a combination of PSTM, WM and long-term language knowledge during repetition of sentences.

A further study that manipulated the length and complexity of sentences to be repeated is reported by Moll et al. (2015). Their study investigated whether primarily language skills or PSTM were responsible for impaired sentence recall among school-age children with dyslexia. They found that both factors explained a significant amount of the variance in sentence repetition, but that when each skill-type (language or PSTM) was controlled for in turn (using scores from independent measures of morphology and PSTM), the children with dyslexia performed more poorly due to differences in language skill, not PSTM.

In an attempt to explain some of the advantages in recall of sentences compared to word lists, advocates of the WM model have proposed the involvement of the central executive (e.g. Jefferies, Lambon-Ralph and Baddeley, 2004), and the episodic buffer (e.g. Alloway, 2007; Alloway et al., 2004; Baddeley et al. 2009). This latter component of working memory is reported to allow chunking of the information to be recalled, achieved through the integration of information from long-term memory. However, as previously discussed (section 1.4.2.), this part of the model is not well specified or well evidenced in the research literature.

Alternatively, Alloway (2007) explains sentence repetition according to Martin's multi-systems model (e.g. Hanten and Martin, 2000; Martin, Lesch and Bartha, 1999). Martin's model incorporates separate phonological, lexical and semantic systems, each with distinct buffers. Alloway (2007) explains that the sentence is encoded semantically and then converted into a lexical code before finally being transformed into a phonological code, ready for articulation.

6.3. Purpose of the study and hypotheses

This part of the study aimed to contribute to the research investigating the influence of PSTM and language knowledge to children's repetition of sentences. More precisely it sought to evaluate the contributions of concurrent grammatical skills and PSTM when children were 4 years old. In so doing it aimed to contribute to the theoretical understanding of the processing involved in sentence recall, as well as to provide rationale for specific clinical interventions arising from the assessment in this population.

It predicted that children repeat sentences accurately to the extent that 1) the vocabulary and sentence structure in the sentence is familiar, and 2) they can use their well-developed PSTM skills to store the novel strings of connected words. It proposed the following set of hypotheses:

- If PSTM is important in sentence recall, then word span will correlate with performance on the sentence repetition task.
- If grammar is important in the recall of sentences then scores on the word structure task (from the *CELF preschool 2*) will correlate with scores on the sentence repetition task.
- If both these factors are separately important in the recall of sentences then they will emerge as independent predictors when subject to regression analysis.

6.4. Methods

52 of the original sample of 54 children were recruited for the second part of the study. The children were aged 4 - 4 ½ years. A detailed breakdown of the composition of this group can be found in chapter 3. As discussed in the general methods (chapter 2), the children were assessed on two subtests of the *BAS II* (Elliott et al., 1996) and on the *PLS 4* (Zimmerman et al., 2002).

The children also completed the experimental tasks (listed below). Some of the children were however non-compliant with the tasks and so actual figures are given in brackets below. Relevant to this part of the study, the children completed the following tasks:

- The *Word Structure sub-test* (50 children): this test is part of the *CELF pre-school -2* (Wiig et al., 2004) assessment. This sub-test is considered to tap children’s use of grammatical morphemes. The task requires the child to look at a picture or pictures and to complete a sentence that the assessor begins. For example, there is a picture of a girl waving from a window to a boy and he is waving back. The assessor begins: “She is waving at him and he is waving at…”. The child must complete the sentence using the word that is semantically consistent with the picture and morphologically consistent with the example. The test therefore depends upon morphological priming.

- A *span task* (52 children): this task had been designed for the first part of the study. Children were presented with pre-determined random sequences of a close set of single syllable words (nouns). The sequences varied from 2 - 5 items and there were five strings at each length. The children completed the strings of 2 items first, followed by 3 items and this continued to increase until the five item strings were complete. The children were required to repeat these words in sequence, and a score of 1 point was awarded for each correct string.
- The standardised *sentence imitation* task in the *ERB* (Seeff-Gabriel et al. 2007) (51 children): The task involves 27 sentences and two practice sentences that the children must repeat verbatim.

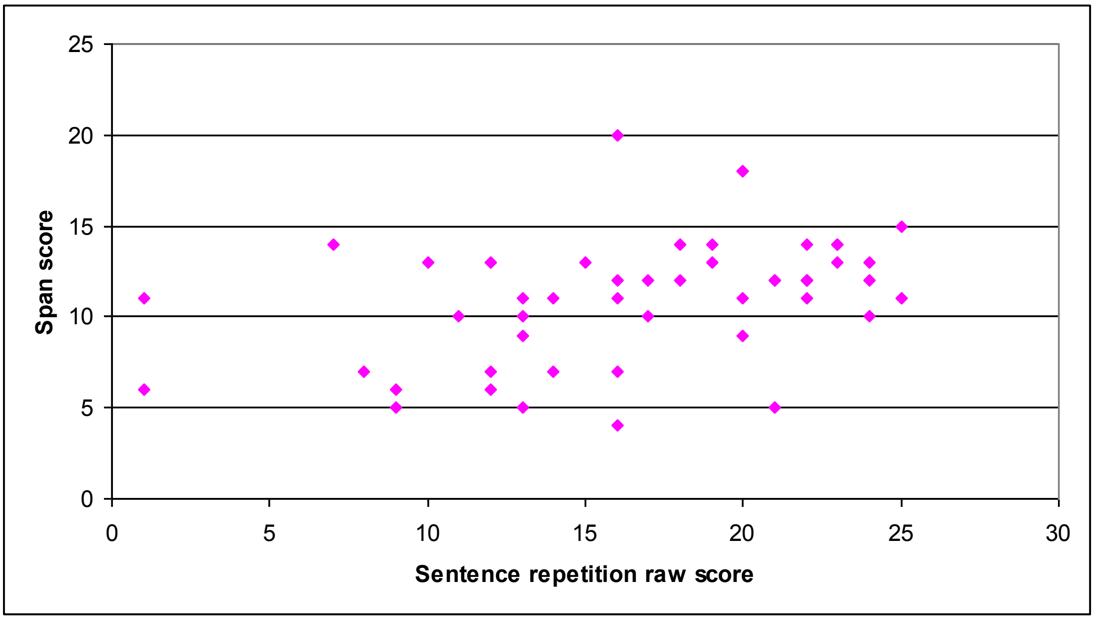
The children were seen on two occasions. During the first visit they completed the BASII subtests and the span task (and additional tasks not included in this part of the study). During the second visit they completed the three standardised language measures. The sentence repetition task and span task were both recorded using a video camera, following written consent by the parents and verbal consent by the children.

6.5. Results

Correlations were calculated for each of the variables (PSTM and grammar). There were two outliers on the sentence repetition task, both with standardised residuals less than -2. Exclusion of these cases from the data did not result in any differences in the significance of the predictors. Therefore an inclusive method of analysis was adopted: all of the children's scores were included in the analyses.

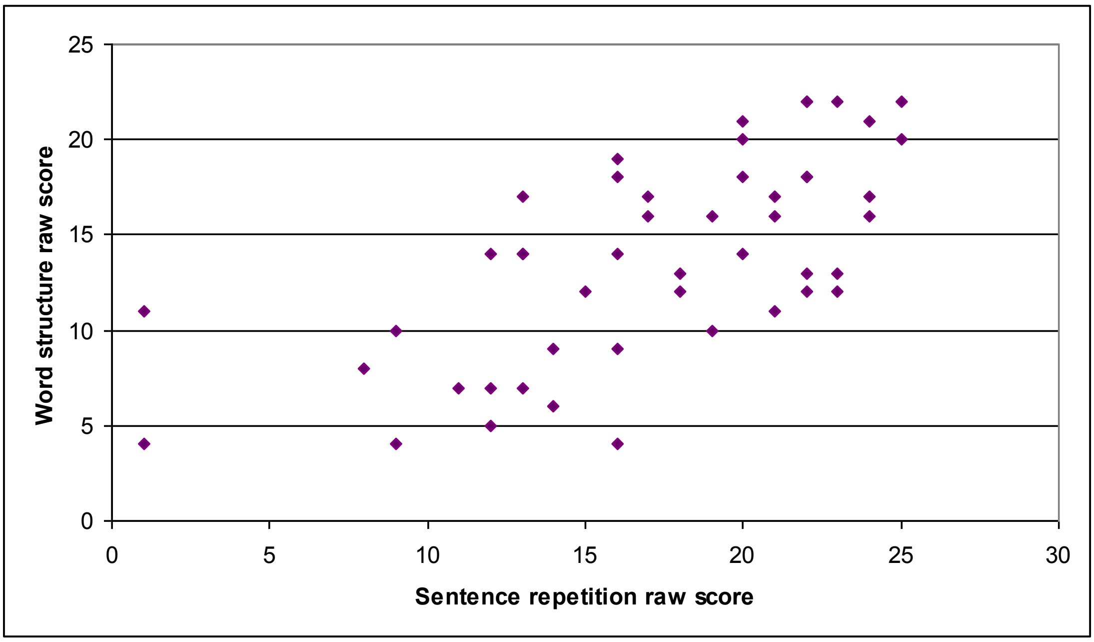
There was a significant correlation between the span task and *sentence imitation task* $r(49)=0.409$, $p=0.003$. This was also confirmed using non-parametric tests (calculated due to non-normally distributed scores, see appendix I): $r_s(48)=0.43$, $p=0.002$), indicating that children who obtained higher span scores generally also repeated sentences more accurately. The scores are plotted on the scatterplot below (Figure 6-1).

Figure 6-1 Scores on the span task and the sentence repetition task



In order to investigate the influence of grammar on sentence recall, correlations were calculated between scores on the *word structure subtest* of the *CELF pre-school 2* and the *Sentence Imitation* task. There was a significant correlation between the grammar score as measured by the *word structure* task and the *sentence imitation* task $r(48)=0.663$, $p<0.001$ (also confirmed using non-parametric tests: $r_s(48)=0.68$, $p<0.001$), indicating that children obtaining high scores on the word structure task also generally achieved higher scores on the sentence imitation task. This is illustrated in figure 6-2.

Figure 6-2 Scores on the word structure task and the sentence repetition task



The two variables: span and grammar (as measured by scores on the word structure task) were entered into a regression equation to predict scores on the sentence repetition task. There were no effects of age on any of the test variables (see chapter 3 (3.4.1)) so this factor was not considered in the equation. However, BAS scores (from both of the subtests) did correlate significantly with the two test variables, so these scores were inserted first into the equation. The model was significant, explaining 50.2% of the variance (adjusted R-squared). The BAS scores accounted for 34.2% of the variance. ($F(4)=12.57$, $p<0.001$). Using a forced entry method, grammar (scores on the word structure task) was found to be a significant independent predictor of sentence repetition Beta=0.408, $t=2.94$, p=0.005. Span (Beta=0.184, t=1.55, p=0.128) was not an independent predictor of sentence repetition.

6.6. Discussion

The following account describes the results of the study and relates these to the previous research findings. It discusses the implications of the findings and also discusses some of the limitations of the study.

6.6.1. Does PSTM influence SR?

The present study predicted that PSTM is employed during sentence repetition. This is based on the premise that a greater ability to hold unrelated items in PSTM, as is required by the span task, would lead to an increased ability to hold related words in PSTM and therefore to a longer sentence being retained. If so, there would be a significant correlation between performance on the span task and scores on the sentence repetition task. Consistent with the predictions, the results showed a significant correlation between the tasks. This therefore suggests that the span task and sentence repetition task share underlying skills or sets of skills. Assuming that word span measures PSTM, the results suggest that PSTM plays a role in sentence repetition.

This finding is consistent with findings reported by e.g. Alloway et al., 2004; Alloway and Gathercole, 200; Baddeley, 1986; Willis and Gathercole, 2001. However, as is clear from the scatterplot (figure 6-1), the relationship is imperfect; some children perform comparatively very well on the sentence repetition task, while more poorly on the span task. The reverse relationship is also true, indicating that additional factors are clearly involved.

6.6.2. Does grammar (morphology) influence SR?

Based on the findings by Potter and Lombardi (1998) that there is a syntactic influence on sentence repetition, and the report by Chiat and Roy (2008) that function word score from the sentence repetition task was the best indicator of their participants' language scores, the present study was interested in the contribution of grammar to sentence repetition performance. More precisely it investigated the role of children's morphological skills. The test used to assess this ability was the *CELF pre-school 2 Word Structure* task. There was found to be a significant correlation between the tasks. This finding is consistent, for example with those by Conti-Ramsden et al., (2001). They found similar size correlations to the one reported in the present study (.62 and .57) between the SR task and the two grammatical tasks used in their study.

6.6.3. Which is the best predictor of SR: PSTM or grammar?

When the results of the span task and the grammar test were entered into a regression model, grammar emerged as the only independent significant predictor. This finding is interesting as it appears to discount the role of PSTM in sentence repetition. This is consistent with a recent study by Klem, Melby-Lervåg, Hagtvet, Lyster, Gustaffson and Hulme (2015). They assessed 4-6 year old Norwegian children's sentence repetition performance together with their vocabulary and grammar skills at three time points spanning 2 years. They used simplex modelling to demonstrate that sentence repetition taps language ability only. Although they did not have an independent measure of PSTM, they proposed that their model of sentence repetition is explained so well by language measures that it leaves explanations from WM, whether the episodic buffer or PSTM, redundant.

That PSTM does not emerge as an independent predictor in the regression model might be explained by underlying skills shared with the grammar task. Post-hoc examination of correlations between the grammar and span tasks shows that there is a significant correlation ($r(49)=0.50$, $p<0.001$).This indicates that the tasks share underlying skills. There are different ways to interpret these inter-relationships and three hypotheses are presented.

If the tests are pure measures of the skills (span=PSTM, word structure=grammar) then a first interpretation might be that one skill causes another skill to develop (e.g. PSTM causes vocabulary development, see Gathercole et al., 1992) or that a further factor encourages development of all the skills.

It is likely however that the tasks are not pure measures of any skill. Instead several factors are implicated in the execution of these tasks. It could be, for example, that the word structure task requires activation of the PSTM system. Interpreted in this way, it could be argued that the child has to hold in PSTM the word used by the assessor in order to select a morphologically consistent word as his/her response. For the example given in section 6.4. above, the child must hold the assessor's word 'him' in PSTM while the assessor continues to speak in order to later select the correct morphologically equivalent item "her" as required in the response. While the word 'him' is held in the phonological store, the assessor's continued talking might disrupt the storage of the word, acting as irrelevant speech. A correct response might further require activation of the central executive (see chapter 1, section 1.4.2.) to inhibit response with a morphologically or semantically inconsistent alternative lexical item (e.g. "him" or "the girl"). If the word structure task involves the PSTM then this would mean that the variance accounted for by span would be included within the word structure factor when entered into the regression.

Alternatively it could be argued that both the span task and the word structure are priming measures. In the case of the span task this could be argued to be a semantic priming task to some extent; recently activated lexical items are selected during recall, while some other mechanism (possibly PSTM) maintains the correct order. The word structure task involves morphological priming: a close neighbour to the recently activated lexical item is selected in the response. An explanation in terms of priming is consistent with that of Potter and Lombardi (1990, 1998) and also with STM interpretations of Cowan (2008) and Martin and colleagues (e.g. Hanten and Martin, 2000; Martin et al., 1999).

A final explanation is that PSTM is not in fact involved in SR and that any shared variance indicated by the correlation results from common factors to all the tasks such as motivation and attention. This interpretation aligns with that of Klem et al. (2015).

The fact that the word structure task does emerge as an independent predictor of the SR task indicates that these tasks share underlying skills. In the case of comparisons with the word structure task, both tasks most obviously require the child to use the correct grammatical morphemes. This is consistent with Chiat and Roy's (2007) observation that repetition of function words was closely related to the children's overall sentence repetition score. As previously discussed, studies have shown that children only repeat grammatical structures in their repetition that they would also use spontaneously (e.g. Devescovi and Caselli, 2007). Presumably this finding can also be generalised to the word structure task, which requires the child to generate a grammatical word for which they have only heard a close syntactic neighbour.

6.6.4. Summary of the findings

The present study indicates that SR shares a stronger relationship with a grammatical task than it does with a span task. This indicates that language knowledge, rather than PSTM is tapped by the SR task, or that language knowledge is tapped to a greater extent than PSTM in the SR task.

6.6.5. Limitations

It should however be considered that different SR tasks likely measure different skills. The extent to which the findings can generalise to other SR tasks is questionable. Following early criticism that non-words, originally designed as a pure measure of PSTM, in fact reflect phonological and lexical skills, Gathercole (1995b) wrote a paper entitled “Is non-word repetition a test of phonological memory or long-term knowledge? It all depends on the non-words”. In the paper she demonstrates that less word-like non-words are more closely related to other measures of PSTM, while more word-like non-words relate more closely to other measures of language. Sentences that have flexible word order seem also to depend more on PSTM. For example, sentences containing lists of adjectives or nouns or sentences involving embedded clauses, depend more on resources from the PSTM system to maintain fidelity to the order (see Riches Loucas, Baird, Charman and Simonoff, 2010).

6.6.6. Theoretical Implications

The findings suggest that there may be a role for PSTM in repeating sentences in children aged 4 years old, but that grammatical ability is more important in this. Returning to the model of sentence repetition set out in the introduction (section 1.6., figure 1-9.), the findings suggest that children use their existing language knowledge to repeat the sentences accurately. It is possible that where existing grammatical knowledge is limited, this might lead to their greater reliance on PSTM, which is subject to capacity limitations and therefore increases the risk of repetition errors, although the findings of the present study cannot confirm whether this is the case. As Riches (2012) states, the sentence repetition task depends on several different types of representations being activated and maintenance of this information in STM. It is probably due to its complexity that it is a good indicator of language difficulties (Riches, 2012).

Chapter 7

Part Four: Exploring the predictive relationship between non-word repetition, word repetition and sentence repetition

7.1. Introduction

NWR and SR have both been identified as useful tasks that identify children with language disorders (Conti-Ramsden et al., 2001). Some studies have suggested links between skills needed for the tasks. For example, Willis and Gathercole (2001) found significant concurrent correlations between scores on these tasks in children aged 4-5 years. Furthermore Riches et al. (2010) found a strong correlation between scores in these tasks in typical and clinical groups of 14 -15 year olds. The aim of this chapter is to explore the relationship between the two tasks. Specifically it aimed to identify which underlying skills the tasks have in common that account for their correlation. Through identifying the shared underlying skills, this would illuminate potential avenues for intervention. The account below presents some of the similarities in factors affecting children's performance on NWR and SR.

7.1.1. Similarities in variables affecting repetition accuracy for non-words and sentences

Evidently there are similarities between any task that requires repetition of language-based material. First the participant needs to focus his/her attention on the task; next he/she must perceive the speech material accurately. As discussed in chapter 1 (1.4.1.3) accurate perception will depend on existing experience of language. For example, there are parallels here with studies showing adults' inability to perceive some phonemes that do not occur in their particular language (e.g. Goto, 1971; Miyawaki, Strange, Verbrugge, Liberman, Jenkins, Fujimura, 1975) and to perceive syllabic structures that are not typical in the language (e.g. Dupoux, Kakehi, Hirose, Pallier and Mehler, 1999; Peperkamp, Vendelin and Dupoux, 2010): perception is influenced by expectation. Repetition tasks also require at least some short-term storage of the verbal material followed by its accurate articulation (see figures 1-5 and 1-9).

In addition to the task similarities described above, previous studies have revealed some interesting matches between NWR and SR tasks in the difficulties encountered by children with and without language disorders as they perform these. Previous studies investigating NWR and SR separately have discovered effects of the amount of information to be recalled (stimulus length), the complexity of the information to be repeated (stimulus complexity) and effects of syllable or word stress on repetition accuracy. These factors will be explored in the following account, considering first how each factor influences repetition of non-words and then its effect on the repetition of sentences.

7.1.1.1. Length

Studies show consistently that there is an effect of length of the stimulus to be remembered on accurate repetition. In the case of non-words, those with fewer syllables are recalled more successfully by young children than those with more syllables. This has been reported by several studies including Casalini et al. (2007), Chiat and Roy (2007); Dispaldro et al. (2009); Dollaghan and Campbell (1998); Ellis-Weismer et al., (2000); Gathercole and Adams (1993); Gathercole et al. (1994); Gray (2003); Chiat and Roy (2007); Roy and Chiat (2004).

Repetition of sentences has also been shown to be influenced by the number of words contained in the sentence. As sentences increase in length they become more difficult to repeat (Willis and Gathercole, 2001; see section 6.2.3). This effect was also confirmed for a clinical sample of children by Seeff-Gabriel et al., (2010). In their study that reports children's performance on a sentence imitation test, Seeff-Gabriel et al. (2010) demonstrated that a clinical sample of children aged 4 years to 6 years performed more poorly on long sentences compared to shorter ones. The control group in this study however performed close to ceiling therefore no effect of length could be established.

7.1.1.2. Complexity

Another factor affecting children's repetition of non-words (as well as words) is the complexity of the phonological structure of the stimuli. Here, this is defined as the type and combination of phonemes selected for the stimuli. Young children typically have a small repertoire of phonemes leading them to simplify and mispronounce words that contain later developing phonemes. NWR tests vary in how they manage this typical developmental speech pattern when scoring young children's responses (see 4.2.3.). As previously discussed (section 2.2.2.1.) the presence of consonant clusters also impairs performance by clinical groups of children (SLI and/or dyslexia), particularly where these are located word medially (Marshall and van der Lely, 2009). However, while consonant clusters are late to emerge in children's speech repertoires compared to singleton consonants, the finding in this study did not appear to be caused by impaired articulation skills.

A study by Devescovi and Caselli (2007) suggested that for children aged 2 -4 years, repetition performance reflected their use of grammar in spontaneous productions. In keeping with this finding, other studies investigating sentence repetition in young children with SLI show that they have particular difficulty repeating function words and inflections (McGregor and Leonard, 1994; Seeff-Gabriel et al., 2010). Function words and inflections may be assumed to increase the complexity of a sentence for young child. Furthermore, discussed in chapter 6 (section 6.2.2) children show greater difficulty repeating sentences that are more grammatically complex, particularly where children have language disorders (e.g. Moll et al., 2015; Riches, 2012; Willis and Gathercole, 2001)

7.1.1.3. Stress

Reported in the design for the stimuli (chapter 2.2.2.1.), in learning words, English-speaking children prefer a prototypical trochaic stress pattern, where the main stress falls on the first syllable (Gerken, 1994, 1996). Words that do not conform to this template tend to be more vulnerable to consonant omission (e.g. Gerken, 1994; 1996; Carter and Gerken, 2003; McGregor and Leonard, 1994). Gerken (1994) reported on children's tendency to prefer trochaic, rather than iambic structures also during repetition of words and phrases.

McGregor and Leonard (1994) hypothesised that children with SLI might find it difficult to repeat certain grammatical morphemes depending on their position in the sentence. This might be due to difficulties perceiving or producing the morphemes in these sentence positions. They compared repetition performance by children with SLI (aged 3 years, 7 months to 5 years, 4 months) and language-matched control participants (who were younger). They found that the clinical group showed impaired performance compared to the younger children on the sentence repetition task. They also found that both groups were more likely to omit function words when these were unstressed at the start of a sentence. However, the children with SLI were not differentially more likely to demonstrate this tendency.

7.1.1.4. Summary

In summary, while it is likely that similarities in task demands (e.g. motivation, attention) explain some of the relationship between NWR and SR, effects of length, complexity and stress are also known to affect children's repetition of non-words and sentences. Effects of length seem to affect both children with and those without language difficulties, indicating that PSTM influences performance. In addition, there is some limited evidence that effects of positional-dependent syllable or word stress and phonological and morphosyntactic complexity may affect children with language disorders differently to those without language difficulties. This highlights the role of the children's language knowledge or skills on performance in the tasks. These latter findings further offer a suggestion of why repetition tasks might be useful clinically.

7.1.2. Longitudinal Studies

Some longitudinal studies have investigated NWR in relation to later language skills. As such they have considered the predictive ability of NWR to wider language skills. For example, Gathercole et al. (1992) examined its relationship with vocabulary. They assessed children at four different time points between the ages of 4 years and 8 years. Using cross-lagged correlations, they found what they interpreted to be causal relationships between children's NWR ability at 4 years of age and their receptive vocabulary a year later. They found reciprocal relationships between the two measures at the time-points thereafter. It appeared that NWR tapped a skill that underlies vocabulary learning in young children. If NWR measures the phonological loop component of Baddeley's Working Memory model (see chapter 1), then Gathercole and colleagues' finding seemed to imply that the phonological loop is key to word learning (e.g. Baddeley et al. 1998).

A very large scale study by Melby-Lervåg, Lervåg, Lyster, Klem, Hagtvet and Hulme (2012) sought to replicate the study by Gathercole et al (1992). They assessed 219 children aged 4-7 years on their NWR ability and their vocabulary and they looked at relationships between the measures at different time-points using a reportedly more robust method of analysis. Their study refuted the claims about links with language development on two counts. First, using two methods of analysis, they could not replicate the study's findings in their own large data-set. Second, they re-analysed the original study's findings using these different methods of analysis and found that the reported predictive relationship between NWR and vocabulary was unsubstantiated.

Findings from another longitudinal study (Chiat and Roy, 2008) revealed an apparent predictive relationship between non-word repetition and sentence repetition. The authors aimed to test the hypothesis that early phonological skills (sensitivity to prosodic patterns) drive later morphosyntactic skills, and therefore that early phonological processing difficulties lead to later morphosyntactic problems. In their study involving children aged 2 ½ - 4 years, they used a combined score from a non-word and real-word repetition task as their measure of phonological processing. Children's early WR and NWR performance was the strongest predictor of their morphological accuracy in a sentence repetition task (function words correctly repeated), when assessed 18 months later.

It was Chiat and Roy's study that inspired the longitudinal aspect of the present study. Their study was used as a foundation upon which to construct the hypotheses. The following account summarises the methods and findings of the study by Chiat and Roy (2008). Following this, explanation will be given as to how the present study sought to build on their findings.

Chiat and Roy (2008) recruited 209 children aged between 2 ½ and 4 years from speech and language therapy caseloads in central London (T1). They collected data 18 months later, at T2, from 187 of the children originally recruited. The children had been referred to speech and language therapy for concerns about language development (not speech); they did not have a diagnosis of autism or hearing impairment, and there were no concerns regarding their general non-verbal skills.

At T1, the children were assessed on their ability to repeat words and non-words and these tasks combined were considered by the authors to tap the children's phonological skills (combined processing and memory). They were also assessed on a range of tasks that were designed to tap the children's socio-cognitive skills (response to acted facial expressions showing different emotions, adult-initiated joint attention and symbolic understanding). At T2, the children were assessed on two tasks judged to measure their morphosyntactic skills: use of grammar in a picture question-response task; and repetition of function words in a sentence imitation task. Established and validated questionnaires were also used to investigate the children's social communication skills at the second time-point. The children's general language and non-language skills were also assessed using standardised tools at both time-points.

Consistent with their hypotheses, Chiat and Roy (2008) found that the WR and NWR task scores (their measure of phonological skills) emerged as the best predictor of the children's function word score on the sentence imitation task (their measure of morphosyntax). This relationship was found after the children's scores had been adjusted for developmental speech errors. The children's phonological skills were not correlated with their later social communication skills, but their early socio-cognitive skills were associated with these. The authors proposed that the findings support theories of skill-development specificity in neuro-developmental disorders such as SLI and autism.

7.1.3. Purpose of the study and hypotheses

While the study by Chiat and Roy (2008) looked specifically at children's repetition of function words during sentence imitation, the present study aimed to investigate relationships between WR, NWR and SR more generally. This had the clinical rationale that sentence repetition tasks are widely used in standardised tests that assess children's language skills. However, while sentence repetition has been established both in the research literature and clinically to identify language needs, it is unclear which skills it measures overall, and therefore which skills to target in therapy.

The present study sought to explore whether the reported predictive relationship between the repetition tasks is mediated by early language knowledge and/or ability to hold phonological information in memory (involvement of PSTM) in both tasks. The present study sought to answer this question by controlling for length, stress and complexity across the word and non-word stimuli.

Therefore the present study did not aim to be compared directly to that of Chiat and Roy (2008). Indeed, it differed in its methods in several ways. These include:

- Obtaining measures of the children's WR and NWR performance at both time-points to explore concurrent as well as predictive relationships.
- Separating out the relationships between i) NWR and SR and ii) WR and SR, as the present study (chapters 4 and 5) found NWR and WR to tap different skills.

- Considering the contribution of PSTM and word knowledge to performance on all the repetition tasks.

Based on the findings in chapters 4 and 5, children repeat real words accurately to the extent that they have correct stored representations for these. Also based on the findings from chapters 4 and 5, they repeat non-words accurately to the extent that they have sufficient stored vocabulary to support familiar word patterns as well as well-developed phonological short-term memory skills to store the novel strings of phonemes temporarily. Therefore the present part of the study predicted that a combination of PSTM and existing word knowledge contribute to successful SR and that this is the reason that NWR and WR predict SR.

The research questions were therefore as follows:

- Does NWR predict SR concurrently and predictively?
- Does WR predict SR concurrently and predictively?
- If NWR is a combined measure of word knowledge and PSTM, and WR is a measure of word knowledge (see chapter 4) then do these skills explain the relationship between NWR and SR?

7.2. Methods

See general methods (chapter 2, section 2.4. and chapter 3) for information about procedure and participants.

Relevant to this part of the study, the children were assessed at both time points using the following tasks:

- Word repetition

- Non-word repetition
- Picture naming (same word stimuli as used in the word repetition task)
- Word span

The WR task was scored in conjunction with the picture naming task, so that only words that the child could name (*known* words) were included in the analysis. This enabled the task to be certain to measure different skills from the NWR task. The NWR task served to provide a measure of the children's accuracy repeating novel strings of phonemes, while the WR task aimed to measure accuracy repeating familiar phonological patterns. Further details about the methods for designing the two tasks can be found in the methods for selecting the stimuli (section 2.2.). The naming task was used primarily to assess the children's knowledge of the stimuli to be repeated in the WR task, and also as an independent measure of the children's speech skills. However, for the purposes of this part of the study, it is proposed also as a proxy to the children's vocabulary knowledge. The extent to which this is a valid approximation of the children's vocabulary was not however verified, and this should be considered when interpreting the results. The word span task included single syllable strings of words that the children were required to repeat in sequence. This served as an independent measure of the children's PSTM.

The children were also assessed using other tasks that were not relevant to the specific questions of this part of the study (see general methods chapter 2).

At T2, the same assessment tasks were completed. Additionally the children were assessed using the standardised Sentence Imitation Test from the Early Repetition Battery (Seeff-Gabriel et al., 2007). Fifty-one of the children completed this task.

For further information about the assessment conditions for the children, see general methods chapter 2). Additionally a grammar task was completed at T2 (see chapter 6). This was not included as a predictor variable in this part of the study, as it did not form part of the specific question for this study (i.e. do NWR and WR predict SR). Furthermore, as the grammar task was not part of the assessment at T1, it could not be used predictively as a factor in SR.

7.3. Results

Predictive and concurrent correlations were explored. Predictive correlations are defined as those correlations that consider task scores at T1 and relate them to the sentence repetition task at T2; concurrent correlations are those reporting scores on all the tasks only at T2. It should be noted that ceiling effects were present for scores on the picture naming, WR and NWR tasks at T2 and these should be considered when interpreting the results of the concurrent correlations.

In the following results summary, initially there is a recap of the research question. Both predictive and concurrent correlations between the tasks follow.

It should be noted that scores on the WR and NWR tasks have been adjusted so that they allow for any speech error patterns (developmental or disordered) that were identified for each child. These include structural and segmental phonological errors. As previous studies have indicated that children's omission of weak syllables affects their inclusion or omission of function words (see 7.1.1.3). The WR and NWR scores were therefore rescored counting any weak-syllable deletion errors as repetition (rather than phonological) errors. At T1 this did not affect the word repetition results, but did affect the NWR results. Therefore both results will be reported for non-words at T1. At T2, the recalculation of the scores affected neither the WR scores nor the NWR scores so only speech corrected scores that include weak-syllable deletion as a developmental error are provided. Correlations involving the speech uncorrected WR and NWR scores can be found in appendix M.

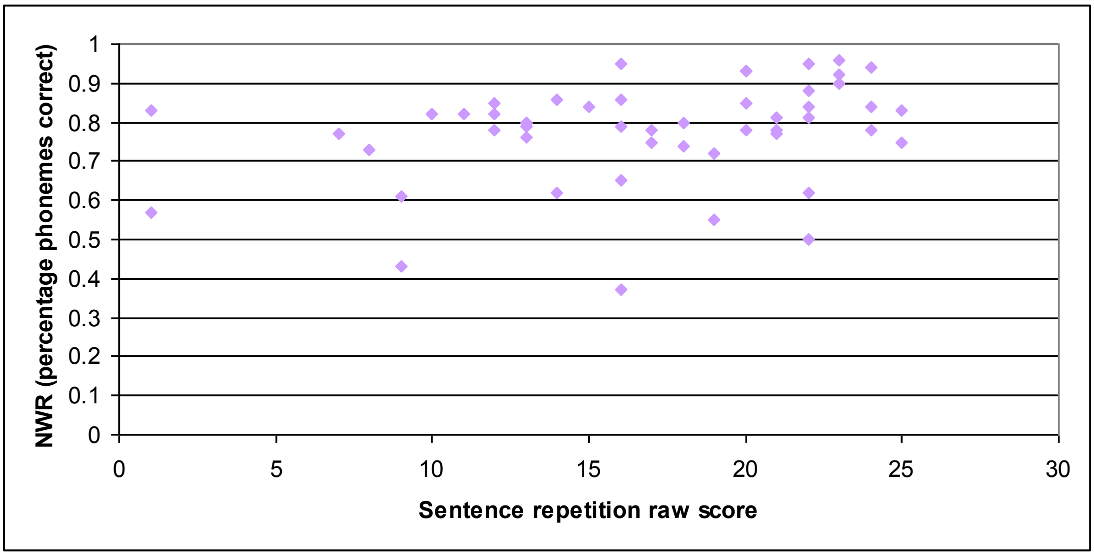
Non-parametric tests (calculated due to non-normality of the distribution of scores) revealed the same results as parametric tests. Therefore only results of the parametric tests are given.

7.3.1. Does NWR predict sentence repetition predictively and concurrently?

<u>Predictive correlations</u>

Following speech error correction, there was a significant correlation between NWR performance at T1 and scores on the SR task at T2: (r(49)=0.30, p=0.035).

Figure 7-1 Speech-corrected NWR scores at T1 plotted against SR scores

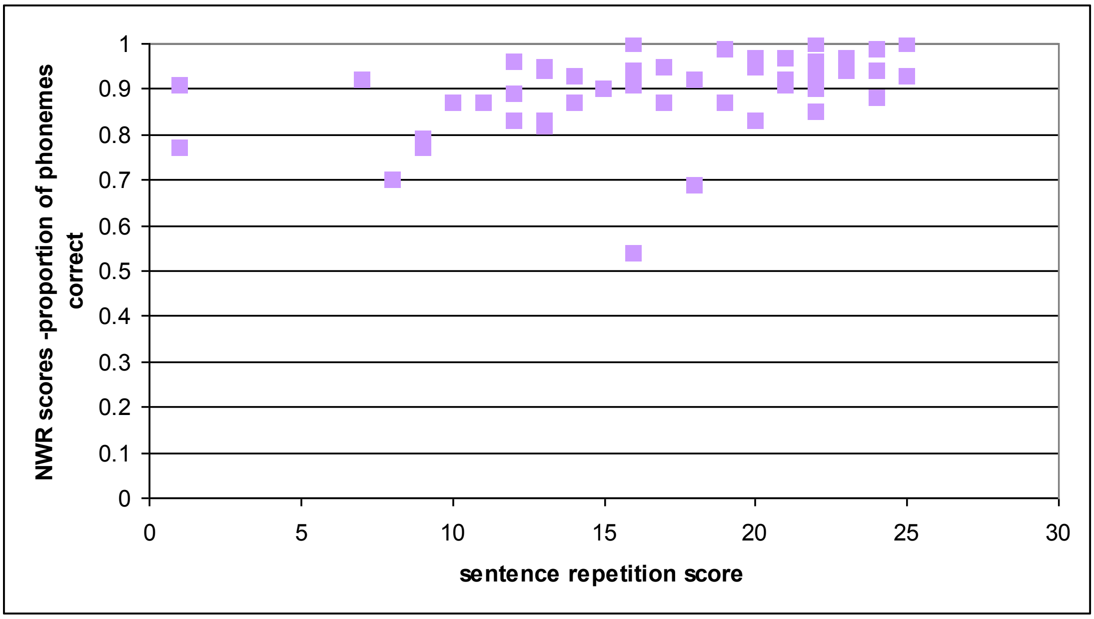


When Weak Syllable Deletion was not corrected from the NWR score, the correlation between non-word repetition at T1 and SR at T2 was also significant: r(49)=0.35, p=0.012. Therefore the correlation was significant under both type of analysis, and was medium in size (Field, 2009).

Concurrent correlations

Following speech error correction, there was a significant correlation between NWR performance at T2 and SR: r(49)=0.38, p=0.006 (medium in size, Field, 2009).

Figure 7-2 Speech-corrected NWR scores at T2 plotted against SR scores



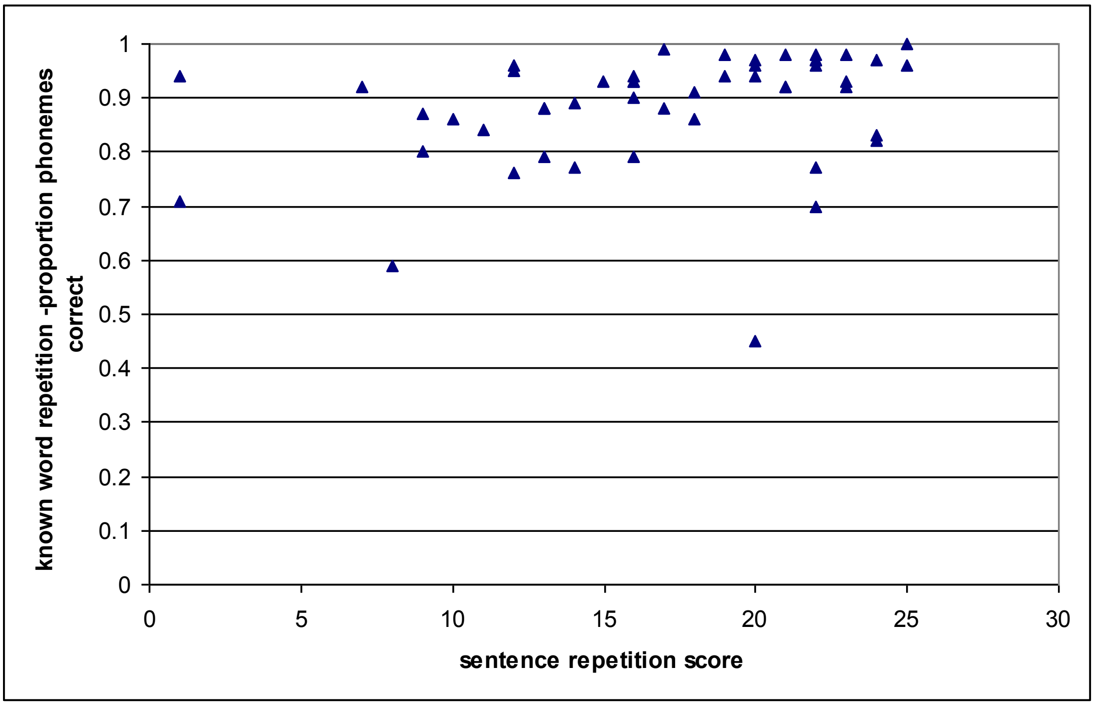
Therefore, overall NWR correlated with SR both predictively and concurrently. Children scoring high on NWR at T1 tended also to score high on SR at T2. Moreover, children scoring high on NWR at T2 also tended to score high on SR at T2.

7.3.2. Does WR predict SR predictively and concurrently?

Predictive correlations

Following speech error correction, there was a significant correlation between speech corrected scores on the word repetition task (known words) at T1 and sentence repetition scores at T2 (r(49)=0.40, p=0.004) (medium-large in size, Field, 2009). Again, those children scoring high on the word repetition task as T1 tended to score high on the SR task at T2.

Figure 7-3 Speech-corrected WR scores at T1 plotted against SR scores



<u>Concurrent correlations</u>

Following speech error correction, there was no significant correlation between word repetition at T2 and scores on the sentence repetition task at T2 (R(49)=0.114, p=0.429, NS). This is very likely due to the marked ceiling effects on the word repetition task at this time-point, as can be seen in tables 7 and 8, appendix H.

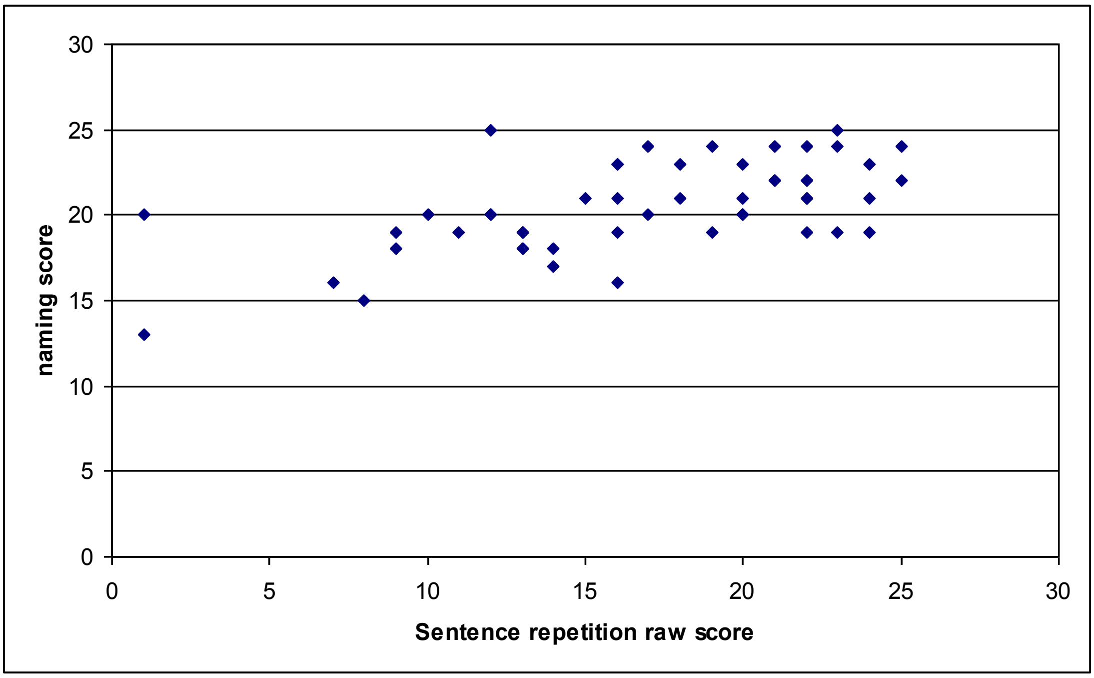
Therefore, overall there is evidence that children scoring high on WR at T1 also score high on SR at T2, but there is no evidence that children scoring high at WR at T2 also score high on SR at T2.

7.3.3. If NWR is a combined measure of word knowledge and PSTM, and WR is a measure of word knowledge (see chapter 4), then do these skills (PSTM and word knowledge) explain the relationship between NWR and SR?

7.3.3.1. Naming task

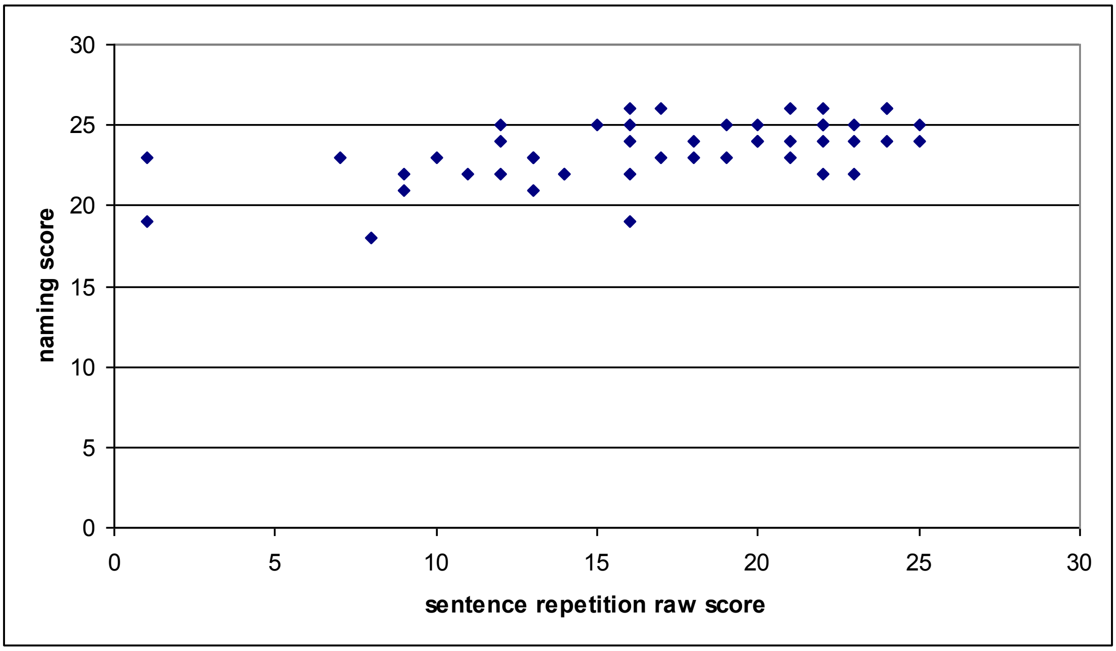
Analysis of the data revealed a significant correlation between scores on the naming task at T1 and SR at T2: $r(49)=0.599$, $p<0.001$. Those children who scored high on the naming task at age 3 years also tended to score high on the sentence imitation task a year later. The correlation size is large (Field, 2009). The relationship between the tasks is illustrated in figure 7-4.

Figure 7-4 Scores on the naming task at T1 plotted against SR scores



At T2, the correlation between performance on the naming task and SR task was significant: $R(49)=0.580$, $p<0.001$. The size of the correlation is large (Field, 2009), although this is subject to ceiling effects. This relationship is illustrated in the scatterplot in figure 7-5.

Figure 7-5 Scores on the naming task at T2 plotted against SR scores

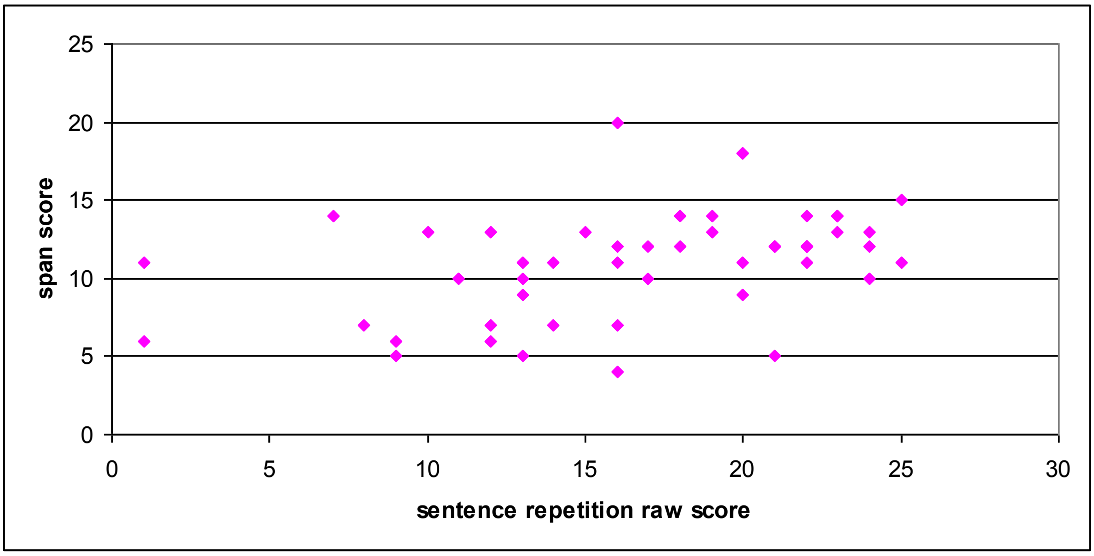


Overall, the data suggest a relationship between early word knowledge and later sentence repetition. The data also suggest a relationship between concurrent word knowledge and sentence repetition.

7.3.1.2. Span task

There was a significant correlation, showing a large effect, between performance on the span task at T1 and performance on the SR task at T2: $r(43)=0.540$, $p<0.001$. Children obtaining higher scores on NWR at age 3 years also tended to obtain higher scores on SR at age 4 years. A scatterplot illustrating this relationship is provided in figure 7-6.

Figure 7-6 Scores on the span task at T1 plotted against SR scores



There was also significant correlation between performance on the span task at T2 and performance on the sentence repetition task: R(49)=0.409, p=0.003. This represents a medium-large correlation (Field, 2009). This is also reported in chapter 6, and the data is illustrated in figure 6-1.

7.3.4. Regression

In order to determine the strongest predictor(s) of SR both predictively and concurrently, multiple regression modelling was performed. Four predictors were inserted into the model at each time-point. As the sample size was 54 at T1 and 52 at T2 and because where there is a small sample size, random data can seem to show a large effect (Field, 2009), the results of the model should be considered with caution. However, between 10-15 cases are generally considered to be sufficient for each predictor (Field, 2009) and the number of cases in the present data is consistent with this. The following scores from T1 were entered into the model: speech-corrected word and non-word repetition, naming and span. This revealed a significant model: F=11.60, p<0.001, explaining 49.7% of the variance in SR scores. Scores on the naming task (t=3.71, p<0.001) (Beta=0.481) and span task (t=2.80, p=0.008) (Beta=0.391) were independent significant predictors in the model. NWR and WR were not significant independent predictors.

Multiple regression modelling at T2, where speech-corrected non-word repetition, span and naming were inserted into the model, revealed a significant model: F=9.23, p<0.001, explaining 33.5% of the variance in SR scores. Naming (t=3.29, p=0.002) (Beta=0.490) was the only significant independent predictor.

Therefore the results suggest that NWR and WR at 3 years predict SR at 4 years only due to the influence of word knowledge and PSTM. The results indicate that NWR but not WR predicts SR concurrently at 4 years and that this can be explained by the influence of word knowledge. This latter finding is however interpreted with caution due to ceiling effects across the WR, NWR and naming data.

7.4. Discussion

This study aimed to explore the relationships between children's ability to repeat words and non-words and their later and concurrent ability to repeat sentences. It aimed to pick apart some of the factors influencing word and non-word repetition to better understand the common skills underlying NWR, WR and SR. This would have the clinical benefit of identifying potentially appropriate therapy targets.

7.4.1. Do children's NWR and WR scores at age 3 predict SR a year later?

The results from the longitudinal study suggest that there is a relationship between children's NWR and WR at 3 years old and their later performance on a SR task. The results were consistent with the findings of the study by Chiat and Roy (2008). Chiat and Roy (2008) found that children's WR and NWR performance at age 2 ½ years to 3 ½ years predicted their later ability to repeat function words in sentences. They found a correlation of r=0.41 between the children's combined WR and NWR scores and their function word scores 18 months later. This compares to correlations in the present study of r=0.30 (correcting for WSD) and r=0.35 (not correcting for WSD) for NWR and r=0.40 for WR. Chiat and Roy's sample was large: 163 children. Similar to their findings, the present smaller scale study found a correlation between children's early single item repetition skills and their later sentence repetition skills.

Previously the present study (chapters 4 and 5) found that word repetition draws extensively on previous knowledge of the words. It is possible therefore that the same underlying skills could be tapped by a naming task. NWR is known to be related to PSTM (see chapter 4), and is also influenced by lexical factors (see chapter 4 and 5) so the present study sought to investigate whether these factors explained the relationship between the early and later repetition skills.

The results from the present study show that children's naming and PSTM best predict SR a year later. While significant correlations were found between both WR and NWR when the children were 3 years old and sentence repetition a year later, this was explained entirely by the combined influence of their ability to name pictures representing early words of 1-4 syllables (an arguable approximation to their vocabulary) and their PSTM measured using a span task.

Chiat and Roy (2008) interpreted the correlation in their study as support for the theoretical hypothesis that early phonological skills predict later morphosyntactic skills. The authors define *phonological skills* as *memory and processing*. The results of the present study suggest some limited support for this hypothesis. It suggests that NWR and WR are predictors only to the extent that they measure the two underlying skills of span and word knowledge. These two underlying skills might fit loosely with Chiat and Roy's definition of *phonological skills* and therefore the results lend support for the theoretical position to a degree.

There were several differences between the study published by Chiat and Roy (2008) and the present study, which make the consistent findings between the studies more striking. First, Chiat and Roy recruited a larger sample of children and the children in their study covered a broader age-range. Second, Chiat and Roy combined scores on the WR and NWR tasks, rather than investigating the contribution of each of these factors independently. By contrast, the present study had separate scores for repetition of known words and repetition of non-words. Third, the word and non-word stimuli used in the present study differed from those used in Chiat and Roy's study. Fourth, the present study scored children's overall sentence repetition performance, rather than their function word scores. Fifth, the present study scored the WR and NWR tasks for percentage phonemes correct, while that by Chiat and Roy (2008) scored for whole items correct. Finally, while Chiat and Roy corrected the children's repetition scores for typical developmental speech errors, they did not have a means to assess for disordered speech errors. Chiat and Roy (2008) acknowledge this in the discussion of their findings: "Some speech production difficulties may lead to repetition errors that are not due to limitations in phonological processing and memory" (p643).

7.4.2. Do children's WR and NWR predict their SR concurrently?

The present study also explored concurrent relationships between NWR and SR and between WR and SR. It found that NWR was a significant predictor of SR, but WR was not. The results are also consistent with other studies that have found a relationship between NWR and SR, but that did not measure repetition of words (e.g. Willis and Gathercole, 2001; Moll et al. 2015; Riches 2012). The correlation between NWR and SR (r=0.38) is similar, though slightly lower in the present study compared to that of Moll et al., (2015): r=0.43 for controls and r=0.54 for children with dyslexia.

If SR is primarily a measure of grammatical skills (see chapter 6) these results are also consistent with the study by Dispaldro et al. (2011). They found a significant correlation between WR and NWR and grammatical skills in Italian but not in English. They measured grammatical skills using sentence completion tasks, whereby participants had to correctly provide third person singular and past tense morphology. Similar to their study, the present study found a significant relationship between NWR and grammatical skills (if SR is primarily a measure of this) and no concurrent relationship with WR.

Conversely, the studies by Dispaldro and colleagues (see also Dispaldro et al. 2009) found a stronger correlation between repetition of late acquired real words and grammatical skill compared to non-words in Italian for children whose language was developing typically. In English, however, NWR correlated more highly with grammatical skills. Dispaldro et al. (2011) explain the different cross-linguistic findings as being due to grammatical morphology encompassing more meaning in Italian than in English and therefore being more connected to the task of repeating meaningful words rather than nonsense words. However, the authors also suggest that the differences might be explained by the different word and non-word stimuli used or the different stimuli used in the grammatical tasks for the Italian and the English-speaking children.

The results of the studies by Dispaldro et al. (2009, 2011, 2013a,b) are confounded by the fact that no measure was taken of whether the children knew the words that they were repeating. As discussed in chapters 5 and 6, it is consequently difficult to determine exactly what is being measured by the WR task in these studies.

While the present study found NWR to correlate with SR performance at age 4 years, it found that this was entirely explained by its relationship with the naming task when entered into a regression model. This suggests that the skills underlying both NWR and SR accuracy may be word knowledge. However, if this is the case then it would be expected that WR and SR would be more closely related than NWR and SR; WR being more closely associated with word knowledge (see chapters 4 and 5).

It is interesting that the regression analyses, entering scores on the tasks at different time points led to different relationships on SR. It might be that there are different influences on sentence repetition at different ages. SR has been suggested to be a complex task in terms of which skills it taps (e.g. Riches et al., 2010). If SR draws upon PSTM and language knowledge, it might be that these two skills develop at different rates over the course of early development, and that they exert their influence on SR to the extent that they are developed. This is consistent with theories of skill development specificity that Chiat and Roy (2008) also discuss in their paper. This suggestion would only be confirmed however if SR had also been assessed at both time-points in the study. Furthermore, the different findings at each time-point might in fact be best explained by the ceiling effects and lack of variability in the scores evident across the WR, NWR and naming tasks at T2 (see 7.4.5.). This, being a limitation of the present study, precludes further speculation about the results at T2.

7.4.3. NWR as a predictor of language skills more generally

In a follow-up study, Chiat and Roy (2013) revisited 108 of the participants from their study 7 years after their original visit (when the children were aged 9-11 years) and found that NWR ability during the pre-school years was not a good predictor of later language skills at age 11 years. If NWR is partially a measure of PSTM then this finding is consistent with a study by Gathercole, Tiffany, Briscoe and Thorn (2005). This latter study showed children who had PSTM difficulties in their early development, but whose Working Memory skills were appropriate did not have any language difficulties by age 8 years. Gathercole et al. (2005) showed however that children with the opposite pattern of difficulties (impaired working memory skills, but preserved PSTM skills) went on to experience ongoing language problems.

One limitation of the present study and others upon which the study is based that use a correlation design is that this does not permit conclusions to be formed about the nature of relationships between factors. Correlations can only detect relationships between variables, rather than the nature of the relationship. For example, the results do not determine whether ability to name pictures (and the inference that this is an indication of vocabulary size) causes improved repetition of sentences, or whether these two tasks both tap a further underlying skill.

A stronger argument in support of the relationship between vocabulary and sentence repetition and PSTM and sentence repetition would be to recruit four groups of pre-school children: one sample set that showed strong PSTM skills but poor vocabulary skills, a second sample set who presented with strong PSTM and good vocabulary, a third sample with weak PSTM and low vocabulary and finally a set with weak PSTM and good vocabulary. The samples would further need not to differ on non-verbal measures or age. The study would compare performance by each of the groups on a sentence repetition measure that was carefully designed to balance sentence length, vocabulary and grammatical complexity (see chapter 6 for rationale). The study would look for dissociations in performance. Alloway and Gathercole (2005) went some way to designing such a task, although they looked only at PSTM skills, and did not control for or assess the children's potentially variable language skills. The proposed study is also similar in design to that of Moll et al. (2015), though their study focussed primarily on children with diagnoses of dyslexia and the children in the study were older than in the proposed study.

The above proposed experimental design, while theoretically strong, would likely be very difficult or impossible to recruit appropriate participants to. There may be more hope of designing such a task with single case studies. While it is questionable how generalisable the findings of such a study would be, Vance and Clegg (2012) explain that, where well-designed, these types of studies do have an important place in research into speech, language and communication needs, where there remains a limited evidence-base.

An alternative approach to confirming the findings of the study would be to recruit young children with equivalent vocabulary and PSTM skills and to split these children into three groups: those that received training on PSTM skills, those that received vocabulary therapy and those who did not receive any therapy. The participants would be assessed on their ability to repeat sentences pre- and post-intervention to determine the effect of each skill. Therapy studies of this kind have the added ethical benefit of potentially improving the skills of the participants who give their time to the research. This type of intervention design is also encouraged by Melby-Lervåg et al (2012). They warn of the dangers of interpreting causality from longitudinal studies that focus only on assessment. Instead they emphasise that in order to determine that relationships between tasks/abilities are causal, therapy studies need to be designed. These intervention studies would train one of the tasks and look for positive outcomes in the other task. They further report that no such therapy study has yet shown promising results (although see chapter 8, section 7).

7.4.4. Further considerations relating to longitudinal studies

A further consideration about the design of the present study is that half of the sample was known to have speech and/or language needs at T1 and they were in receipt of therapy during the study. This was also the case for Chiat and Roy's (2008) study, upon which this one was based. An optimistic view of the outcome of therapy would be that some of the children's speech and/or language difficulties resolved or changed over the course of the study as a result of the intervention they received. It is also possible that some of the children's needs may have changed or resolved due to maturation, or as a consequence of other external or internal influences. In support of this possibility, studies by Conti-Ramsden and Botting (1999) and by Conti-Ramsden, Botting, Simkin and Knox (2001) found that the language and communication needs of many children attending language units in the pre-school years changed over time. Some of the children's needs appeared to normalise entirely, while other children moved from one category of language and communication needs to another.

While the time period between data collection points was less in the present study compared with those reported by Conti-Ramsden et al. (1999, 2001), it might be that similar changes occurred for the children in the present study. These children might be especially likely to experience such changes due to their young age and potential for neural plasticity.

7.4.5. Limitations of the study

Some limitations of the study have already been discussed. First, there was a complication of ceiling effects at T2 on the WR task. According to the theoretical model used in the present study (see figure 1-5, chapter 1), this is unsurprising: once a word is stored lexically, it is this representation that is drawn upon during single word repetition. Other studies that have used measures of WR have not reported ceiling effects, but results from the first part of this study (chapters 4 and 5) suggest that this may be because the children did not know the words and were therefore treating them as non-words. There was also limited variability in the scores on the naming task at T2 and these scores also approached ceiling. This was due to the design and original rationale for the naming task. A stronger argument in support of vocabulary being an independent measure of sentence repetition would be established if a standardised vocabulary measure had been used. An example of an appropriate measure is the British Picture Vocabulary Scale: 3rd Edition (Dunn et al. 2009), which is a receptive vocabulary measure, or the Renfrew Word Finding Test (Renfrew, 1995) which looks at young children's ability to name pictures. As was the case for the measure used in the present study, these tests have the disadvantage of focussing exclusively on nouns. It might be that ability to label other word categories might be a better measure to use. Verbs might be a particular candidate for such an exploration, as children with language disorders are known to find these particularly troublesome (e.g. van der Lely, 1993).

7.4.6. Summary of findings

Results from the study show that in a heterogeneous group of children aged 3 years, ability to name pictures and word span are the best predictors of later ability to repeat sentences when compared with WR and NWR. They show that performance at 4 years old, their ability to name pictures gives the best estimate of their ability to repeat sentences at the same age.

Chapter 8

General Discussion

8.1. Overall study aims and findings

The focus of the present study was on repetition tasks and their relationship to language development and language difficulties. It sought to understand the contribution of PSTM and language knowledge to repetition tasks. In the case of language knowledge it investigated specifically whether knowledge of the words influenced performance on WR and whether grammatical skills influenced SR accuracy. The study also sought to understand the relationship between NWR, WR and later SR; if single item repetition tasks (WR and NWR) predict later SR (Chiat and Roy, 2008), then which underlying skills do the tasks share?

The study found in the case of WR, that word knowledge influenced performance, and in the case of NWR, both word knowledge and PSTM contributed to accuracy. Grammatical (morphological) skills played the most significant role in sentence repetition accuracy. The study also found that the key skill underlying all three tasks commonly used in the literature in this field (NWR, WR and SR) was language knowledge, rather than PSTM. That the tasks measure language knowledge makes them useful in SLT clinic.

The study was novel in its assessment of the children's knowledge of the words they were repeating. Previous studies had reported that children with language difficulties find repetition of real words as well as non-words difficult, but it was not clear whether the children in these studies knew the words they were repeating and therefore which skills were being assessed and how much this was influencing the results. The study was also novel in its methods for scoring speech errors.

Previous studies with young children had either managed scoring of children's speech errors by designing stimuli constructed only with early developing consonants (e.g. Dollaghan and Campbell, 1998) by correcting for developmental speech errors (e.g. Chiat and Roy, 2007, 2008; Roy and Chiat, 2004) or by looking for similar developmental speech errors in other repetition responses and correcting for these where consistent (Ellis Weismer et al., 2000). The present study, by contrast, used the children's naming data to gain a more accurate measure of the children's speech errors and to ensure that disordered speech errors were also considered in the scoring.

The longitudinal part of the study used the study by Chiat and Roy (2008) as a foundation. Their study had shown that WR and NWR accuracy predict later SR. The present study was novel in its attempts to identify whether PSTM, word knowledge or these skills combined were the skills that underlie this relationship. It further investigated the relationship both concurrently and predictively.

As discussed, the study was both cross-sectional and longitudinal in design. Chapters 4, 5 and 6 discussed findings from the children at fixed time-points. Chapter 7 discussed results from the longitudinal study, i.e. across the two time-points. This latter chapter also considered concurrent measurement of NWR, WR and SR to identify which skills these tasks share. A summary of the findings and implications from each of the chapters is provided in the next section (section 8.2.). Following this there is discussion about the implications for the role of existing language knowledge and PSTM in repetition tasks (sections 8.3. and 8.4.); implications for theoretical models (section 8.5.); a report on some limitations of the study (section 8.6.); brief discussion about intervention studies and conclusions from the study.

8.2. Summary of the findings from each part of the study and implications

In chapter 4 there were three main findings. First, an effect of word knowledge was identified: known words were repeated most accurately, whilst unknown real words and non-words were repeated with similar accuracy (although see T2 results). This finding suggested that while known words benefitted from a deep level of processing, being able to access the stored lexical (semantic and phonological) representations, unknown words were processed only shallowly, benefitting no more from sub-lexical processing than non-words did. The second finding concerned speech errors, which affected both words and non-words. This finding had implications for using the tasks in research studies, particularly where clinical populations are involved, as pronunciation errors might be difficult to identify. It highlighted the need for speech assessments to occur in conjunction with repetition tasks in clinical practice and in research studies. The third finding was in relation to PSTM. Length effects were found for non-words but not known words, highlighting the role for PSTM for these stimuli and not for known words. Furthermore, the span task was found to be a predictor of NWR but not WR in regression analyses, again highlighting the role for PSTM in NWR but not WR tasks. As ceiling effects were apparent for the WR task (where words were known) at both 3 and 4 years old, the study cautioned against using this as a clinical measure.

Chapter 5 explored differences in WR and NWR accuracy between children who had identified speech and language needs and those who had no such identified difficulties. It found that children with speech and language difficulties performed less well repeating both known words and non-words in comparison with their age-matched peers at age 3 and 4 years. Neither the clinical group nor the non-clinical group showed word-length effects in their repetition of known words, suggesting that previous studies' findings that these effects were present for children with language disorders (e.g. Chiat and Roy, 2007, 2008) were due to the children not knowing the words they were repeating. That they repeated words less accurately than their peers was consistent with previous studies (Casalini et al., 2007; Chiat and Roy, 2007, 2008; Dispaldro et al., 2011, 2013a,b). This finding did not however support the hypothesis that they consistently access long-term representations during the repetition of these items. Alternative interpretations were suggested. For the non-words, the clinical sample showed the same pattern of item-length effects as the non-clinical sample, but their scores for each syllable number was lower than their peers. This was consistent with findings by Dispaldro et al. (2013a,b). This indicates that while PSTM does contribute to performance for all the children, the clinical group's lower scores might be influenced more by impaired sub-lexical processing or access to this level of processing than by impaired PSTM.

Chapter 6 investigated the role of PSTM and grammatical (morphological) skills in SR. It found morphology to be an independent predictor of SR and found PSTM was not. This is consistent with the view that SR is a measure of language skills, rather than PSTM (e.g. Klem et al., 2015; Melby-Lervåg et al., 2012; Potter and Lombardi, 1990, 1998). An explanation relating to semantic and morphosyntactic priming emerged as most plausible (e.g. Potter and Lombardi, 1990, 1998; Riches, 2012). However this part of the study concluded that the findings may not generalise to all sentence types. For example, where word order is less fixed or where lists of adjectives or nouns are present in the sentence there may be a role for PSTM.

Chapter 7 investigated the reported relationship between NWR, WR and SR both predictively and concurrently. Interest had stemmed from a previous study (Chiat and Roy, 2008) which had identified that repetition of words and non-words by 2-3 year old children predicted later repetition of function words in a SR task. Furthermore, other studies (e.g. Conti-Ramsden et al., 2001; Riches, 2012) had identified concurrent correlations between these tasks. The present study sought to understand the nature of the relationship, i.e. which underlying skills did the tasks have in common. A predictive relationship was found between NWR and WR at age 3 years and SR a year later. Subsequent analysis revealed that this was explained by the combined influence of PSTM and word knowledge at 3 years. A concurrent relationship was found between NWR and SR at 4 years, which was consistent with previous studies (e.g. Conti-Ramsden et al., 2001; Riches, 2012). However, subsequent analysis revealed that this was explained by the variance both tasks shared with word knowledge. A concurrent relationship was not found between WR (known words) and sentence repetition, probably due to unavoidable ceiling effects in the case of repetition of known words. If sentence repetition is primarily a measure of grammar, then this finding was consistent with the findings from Dispaldro and colleagues that there was no relationship between WR and grammar in English children (Dispaldro et al. 2011) or in clinical samples of Italian children (Dispaldro et al., 2013b). However, it is inconsistent with the findings from Dispaldro et al. (2009; 2011) that word repetition is related to grammatical skills in Italian children who did not have language difficulties.

In summary then, the present study found that the key skill underlying all three tasks (NWR, WR and SR) was language knowledge.

8.3. Repetition tasks as a measure of children's underlying language skills

Overall, the study supports the view that existing language, rather than PSTM is the important factor in using repetition tasks clinically. There are three main pieces of evidence in support of this. Each is discussed in turn.

First, previous studies have indicated that WR tasks as well as NWR tasks identify children with language difficulties (Chiat and Roy, 2007; 2008; Dispaldro et al., 2013a, though see Dispaldro et al., 2011). The present study suggests that this is because the participants in these previous studies did not know some of the words that they are repeating, due to reduced vocabularies and less efficient sub-lexical processing. This leads them to process the unfamiliar words in the same way that they process non-words. It suggests that WR might identify children with language difficulties for two reasons: it uncovers the children's depleted lexicon; and the children are forced to rely on a less well developed sub-lexical processing system, possibly resulting from the depleted lexicon. Alternatively, it might be that the children have difficulty accessing these stored lexical and sub-lexical representations (see Leclercq et al., 2013, for a similar interpretation and Ramus and Szenkovits, 2008, for a similar interpretation in the case of dyslexia).

Second, while the study supports previous studies that have indicated that NWR taps PSTM as well as sub-lexical processing (e.g. Archibald and Gathercole, 2007), it is the latter which is likely to be impaired in the case of children with language difficulties. This is evidenced by the same pattern of performance on the NWR task by both groups (see figure 5-1). If PSTM was responsible for the clinical group's impaired performance then it would be expected that the children with language difficulties would show disproportionately worse performance on longer non-words compared to their peers, as these would require more resources from PSTM. This was not the case. Instead, the finding is consistent with the interpretation that the clinical group's general worse performance is due to their sub-lexical templates not being well developed or that there is impaired access to this level of processing. The former suggestion seems like the best argument, as templates are thought to be updated and better specified as vocabularies grow (Velleman and Vihman, 2002). Children with language difficulties, having reduced vocabularies, would therefore have less well-specified sub-lexical templates to support their PSTM. The clinical implication is that while targeting either PSTM or vocabulary might improve performance on a NWR task, any improvement in NWR is likely only to generalise to other language tasks by working on vocabulary, rather than PSTM.

Third, while a significant correlation was found between SR and the span task (which was the independent measure of PSTM), this task did not explain any significant additional variance when inserted with the grammar task into a regression equation. Conversely, the grammar based task did emerge as a significant predictor of SR. This highlights the role of existing language knowledge in this task too.

8.4. The role of PSTM in repetition tasks

There is however some evidence in support of the view that PSTM has a role in repetition tasks. First, it explained independently a significant amount of the variance in the NWR task. Second, the finding emerged (chapter 7) that the reason that NWR at 3 years old predicts later SR (at 4 years old) is due to the variance they both share with PSTM as well as existing word knowledge. It is difficult to interpret this latter finding when PSTM did not emerge as an independent predictor of SR. One explanation could be that the task used to assess use of grammatical morphemes (*Word Structure subtest* of the *CELF pre-school 2*) also tapped PSTM and so the inclusion of both the word structure task and the span task into the regression equation eliminated the effects of the latter (chapter 6). A second explanation might be that PSTM combined with vocabulary at an early age has a causal link to the development of grammatical skills. A third explanation is that the span test and naming tasks gave a better indication than the NWR and WR tasks of other factors influencing performance, e.g. motivation and attention. These extraneous factors might in fact be the best predictors of later scores on the SR task, although if so, it is unclear why these would not also be revealed by the NWR and WR tasks.

The second of the above explanations fits best with the existing literature, which suggests a relationship between vocabulary and PSTM (e.g. Bishop, 2006; Gathercole, 2006; Gathercole et al., 1992; Gray, 2006; Hoff, Core and Bridges, 2008) and also between vocabulary and grammar (e.g. Conboy and Thal, 2006; Devescovi, Caselli, Marchione, Pasqualetti and Reilly and Bates, 2005; Dixon and Marchman, 2007). Further studies have suggested a link between PSTM and grammar (Baddeley, Gathercole and Papagno, 1998, although see also Bishop, Adams and Norbury, 2006; Bishop, 2006). Furthermore, a study of foreign language learning found an association between all three skills (Martin and Ellis, 2012).

Whatever the reason for PSTM emerging as a predictive factor of SR, it is clear that this skill should be included in a model that explains NWR, and there is some indication that it should be included in a model that explains SR. Attention will now return to the model of Working Memory (Baddeley and Hitch, 1974; Baddeley 2000) introduced in chapter 1 and discussed throughout this thesis. As has been presented, this model has been proposed to explain NWR (e.g. Gathercole and Baddeley, 1990) and SR (e.g. Alloway, Gathercole, Willis and Gathercole, 2004; Baddeley et al., 2009). Discussion will follow as to the extent to which the model explains the findings of the present study.

8.5. Proposed models to explain repetition accuracy

8.5.1. Word and non-word repetition

Gathercole and Baddeley (1990) proposed that the phonological store component of the working memory model can account for performance on NWR tasks, particularly where the non-words are less word-like (Gathercole, 1995b). To recap, the phonological store forms part of the phonological loop ‘slave system’. As extensively discussed however, many subsequent studies have highlighted the role of language knowledge to successful repetition (e.g. Devescovi and Caselli, 2007; Klem et al., 2015; Moll et al., 2015; Potter and Lombardi, 1990, 1998; Riches, 2012; Seeff-Gabriel et al., 2010; Willis and Gathercole, 2001). The present study demonstrated that controlling for these language factors by matching words and non-words for phonotactic frequency, prosody and phoneme composition continued to result in word-length effects for non-words. Word length effects are generally considered to reflect the limited capacity of the phonological store. The fact that these effects were not present in the case of longer real words suggests that they cannot be explained by, for example, difficulties articulating the longer sequences of phonemes. Limitations in PSTM seem to be the most plausible explanation.

A theoretical model of NWR needs therefore to include the phonological store or some representation of verbal short-term memory as well as access to sub-lexical processing. If the model is to explain word repetition as well, then the model needs to demonstrate access also to stored lexical representations. Currently the Working Memory model does not specify very well how access to this long-term store of information is achieved. The Working Memory model describes an additional episodic buffer system that can mediate information between long-term and short-term memory, but it is not clear from the literature whether this episodic buffer is employed in the case of NWR. The model presented in chapter 1 (figure 1-5) suggested a different model that encompassed both PSTM and longer term language knowledge.

It is proposed that the present study's suggested model can explain the advantage for repetition of known words over unknown words and non-words, and that it can explain the non-word disadvantage by children with impaired language systems. These children show difficulties repeating unfamiliar lexical items most likely because they have depleted information available to draw upon in their sub-lexical store as do the language unimpaired children.

8.5.2. Sentence repetition

In the case of sentences, a satisfactory account has not yet been provided by the Working Memory model to explain better repetition of sentences over word lists. The Episodic Buffer has been tentatively put forward as a candidate to explain how lexical knowledge might be accessed to 'buffer' the phonological loop (Alloway et al., 2004; Baddeley, 2000; Baddeley et al. 2009), but it is unclear how this component of Working Memory achieves this and it is also unclear how to test it. One explanation of how long-term language knowledge supports PSTM is through a redintegration process, whereby decayed traces in PSTM are restored through matching to long-term stores (Hulme et al., 1991). However, there has also been suggestion in the literature that PSTM is not implicated in the repetition of sentences and that language processing systems, representing temporary activation of long-term language knowledge can explain performance on the task without the need for an additional short-term memory system (e.g. Acheson and MacDonald, 2009; Cowan, 1995; Klem et al., 2015; Potter and Lombardi 1990, 1998).

The present study is inconclusive with regard to the need for PSTM in the repetition of sentences. The results of the regression analyses investigating skills concurrent with sentence repetition provide no evidence for the role of PSTM.

However that PSTM at 3 years emerged as a predictor for SR at age 4 years indicates some role. The present study would have been better placed to answer this question had it manipulated the sentences to be remembered. It makes sense, for example that words in sentences that have optional word order would draw upon resources from PSTM, while other word components that have more mutually dependent syntactic relationships would draw upon existing syntactic and semantic knowledge, or *priming* as described by Potter and Lombardi (1990, 1998). In order to test this empirically, sentences could be designed which were equivalent in length (number of words or number of morphemes) but where words in one stimuli set were constrained by syntactic relationships and in the other set they were less constrained (e.g. sentences involving lists of nouns and adjectives). However, as highlighted by the carefully designed systematic study reported by Polišenská (2011) (chapter 6, section 6.1.2.) any manipulation of one aspect of the sentence necessarily affects another variable of the sentence, thus making it difficult to draw conclusions about observed effects.

One study that did manipulate sentence length and complexity and observed the effects on children with dyslexia was by Moll et al. (2015). Their findings are interesting to consider in relation to the present study's findings about NWR. While all the children (who were older than in the present study) showed greater difficulty repeating the longer sentences, suggesting the role of PSTM, the children with dyslexia were differentiated from the non-clinical group in terms of the sentence complexity. Again, this finding supports the view that while PSTM is involved in repetition tasks, it is the involvement of long-term language knowledge rather than PSTM that impairs performance by the children with identified language difficulties in comparison with their peers.

8.5.3. Proposed combined model of language and memory

The present study therefore proposes that the models set out in chapter 1 (figures 1-5 and 1-9) go some way to explaining performance on repetition tasks. The present study proposes that there is gain in drawing on both a cognitive and a psycholinguistic theoretical perspective and that aspects of these approaches can be understood in harmony, rather than in opposition. The models presented in this study take aspects from both the Working Memory model (Baddeley and Hitch, 1974) and in the case of WR and NWR also from theories of sub-lexical and lexical priming, and in the case of SR from semantic and syntactic priming (Potter and Lombardi, 1990, 1998). Models that incorporate both approaches are supported by findings from studies that find an influence both of PSTM and language experience in repetition. In the case of non-words, such studies include those by Archibald and Gathercole, 2006; Archibald and Gathercole, (2007), Casalini et al. (2007), Dispaldro et al., (2009, 2011, 2013a,b). In the case of sentences, studies in support of both PSTM and language knowledge include those by Alloway and Gathercole, (2005); Alloway et al. (2004); Baddeley et al. (2009); Moll et al., 2015; Willis and Gathercole, 2001. They are also supported by the suggestion of Bishop (2006) that children who have weak PSTM, or poor grammatical skills in isolation, may not manifest a language disorder, but that a combination of these difficulties may result in profound language difficulties.

The following account considers the models in the light of the data collected.

8.5.3.1. To what extent do the data support the model: words/non-words?

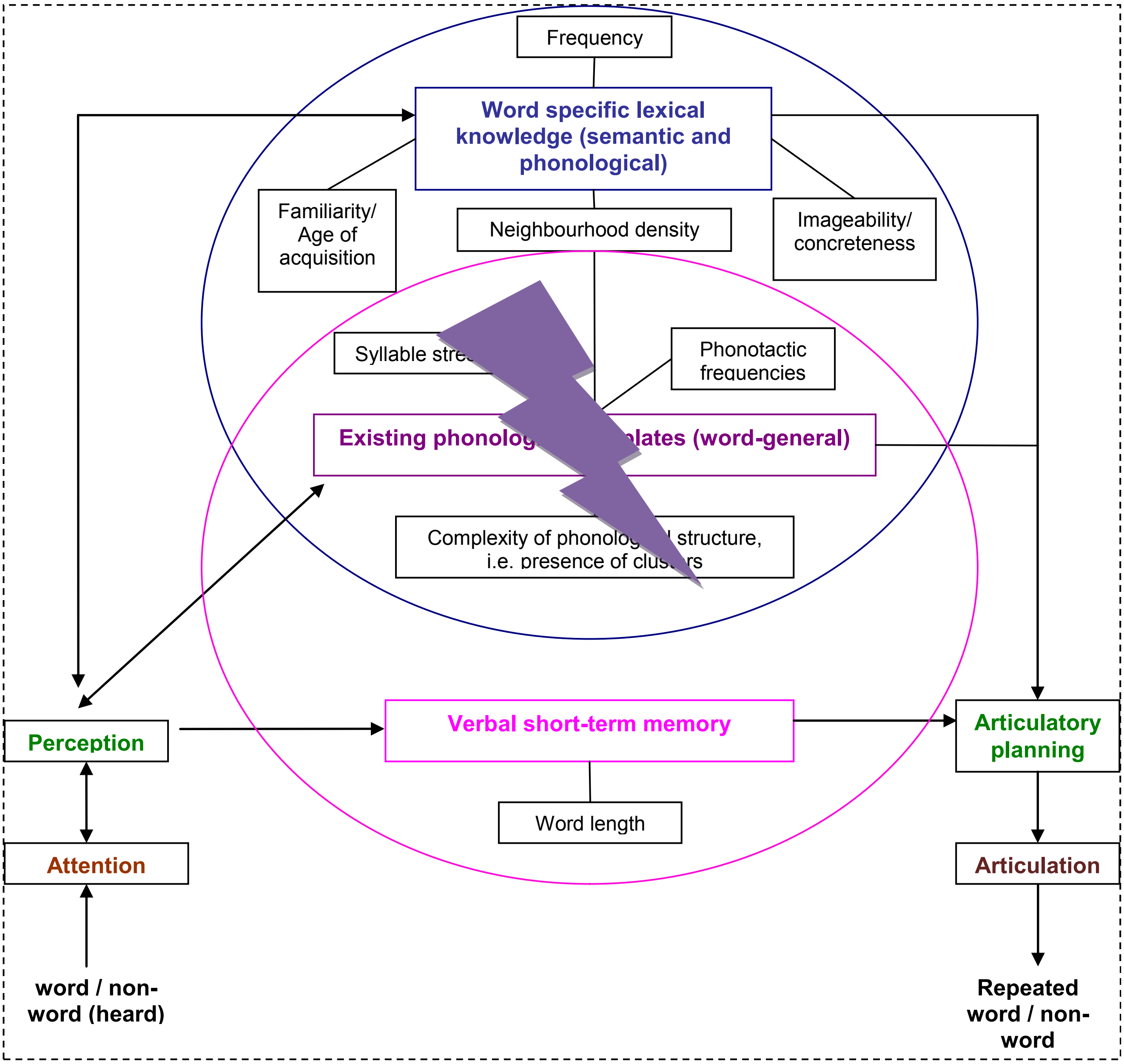
Results from the first part of the study suggest that there is no need for the activation of PSTM in the case of repeating words that are known. PSTM was included in the model for real word repetition (figure 1-5, although see also figure

1-7), based on previous findings that children with language difficulties repeat longer real words less accurately than shorter real words (e.g. Chiat and Roy, 2007). An alternative explanation was that the longer real words are more complex to articulate for children with language disorders. The present study however found no such word length effect for known words. Once speech errors had been corrected, children repeated known words with high accuracy regardless of their length. This was the case for children with and without speech and language difficulties (confirmed by the results of the mixed-design ANOVA in chapter 5 of the study). The findings suggest that the children in the previous studies did not know the longer words that they were repeating and were therefore treating them as non-words. The finding that PSTM is not implicated in the repetition of known words necessitates a change to the model. The bold arrow leading from the 'existing word knowledge' part of the model directly to articulation has therefore been inserted (figure 8-1).

There is evidence however to suggest that the model in its original form can explain the observed findings from the non-word repetition data. Both children with and without speech and language difficulties demonstrated clear length effects in their repetition of non-words. As the non-words increased in length, repetition accuracy worsened for both groups. Children with speech and language difficulties showed particular disadvantage for these stimuli, but the pattern of difficulty was identical for both groups (see figure 5-1). If the clinical group had shown increased difficulty repeating only longer items, this would have led to the conclusion that PSTM was driving the repetition impairment for these children. This however was not the case. Instead the sample of children with speech and language difficulties performed more poorly across non-words of all lengths. This was confirmed by the absence of a 3-way interaction in the mixed-design ANOVA. The findings point therefore to a problem at the level of the phonological templates, rather than

PSTM, for these children. This is illustrated in the revised model below by a purple lightning bolt.

Figure 8-1: Revised model of word and non-word repetition (indicating likely locus of difficulty for children with speech and language difficulties)



The weaker phonological templates likely result from a smaller overall vocabulary in the case of children with language delay and disorder, i.e. phonological templates would become more specified as conscious word knowledge increases. This would be partially confirmed by comparing the naming data by the clinical and non-clinical groups. It would further be supported by comparing repetition data by a clinical sample and a younger sample who were matched to the clinical sample by vocabulary.

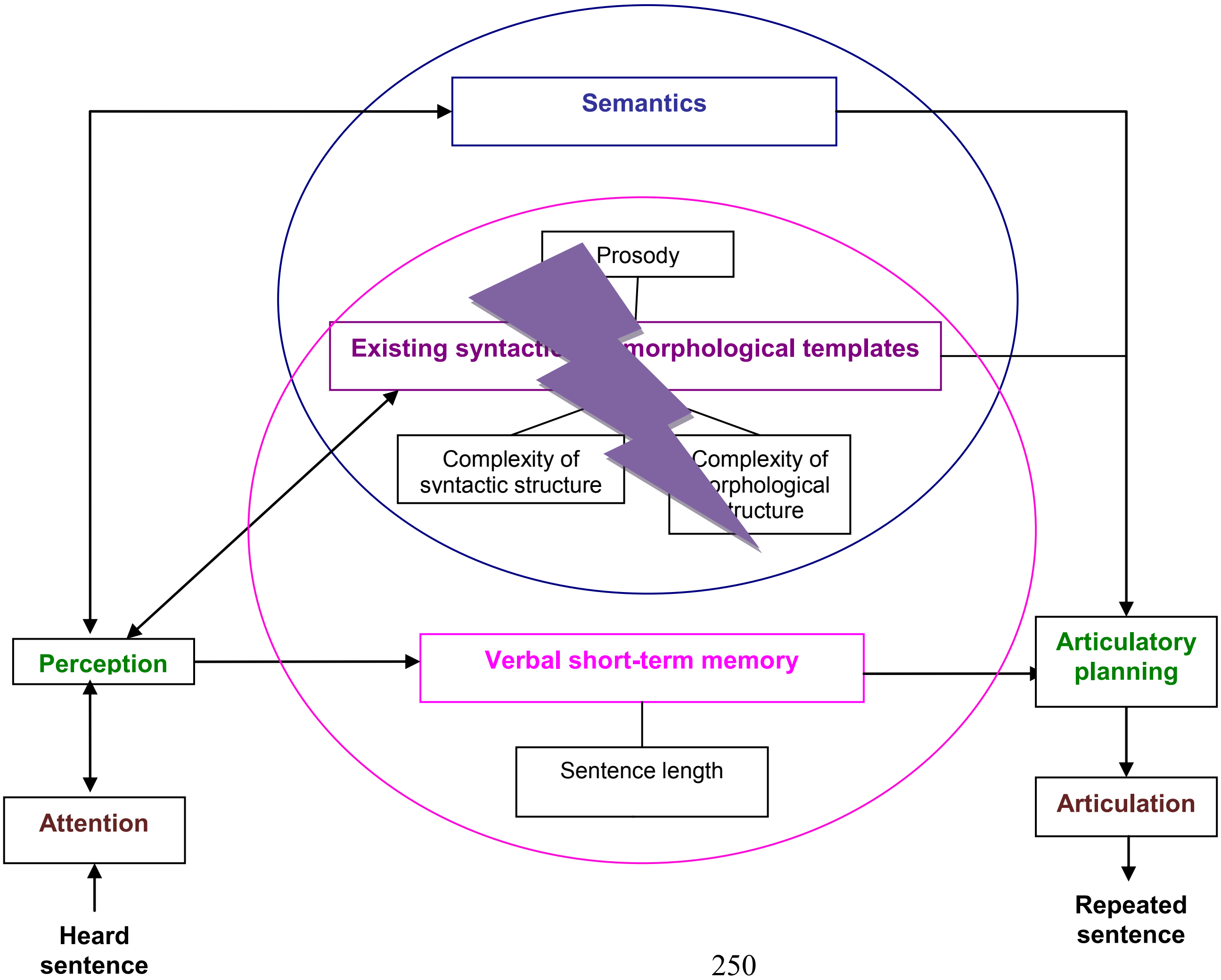
The present study cannot provide conclusive evidence for or against the inclusion of the individual factors included at each level of the model (e.g. syllable stress, phonotactic frequencies, imageability etc). This is because the study attempted to control for these factors in the design of the word and non-word stimuli, rather than to manipulate these factors. Future studies could further elucidate the role of such factors by systematically manipulating each factor, while controlling for the other factors in designing word and non-word stimuli.

8.5.3.2. To what extent do the data support the model: sentences?

The present findings support the view that existing grammatical (morphological) templates are implicated in the repetition of sentences. Those children who performed well on the test that tapped use of morphemes also tended to achieve better scores on the sentence repetition task, and the results of the regression analysis indicate that the tasks tap (a) common underlying skills(s). No group comparison (clinical versus non-clinical) was made. However, the findings suggest that where children exhibit grammatical difficulties, as is typically the case for children with language delay and disorder, they will also perform less well on the sentence repetition task. The locus of difficulty in the model of sentence repetition is therefore comparable to that of single word/non-word processing. As before, this is indicated in the model below as a purple lightning bolt.

The results of the present study also suggest that PSTM is not implicated in the repetition of the sentences used in this study, leading perhaps to an arrow leading directly from the blue area (existing language knowledge) of the model to articulation (inserted as a faint arrow in figure 8-2, below), or indeed to the complete disappearance of the pink area (verbal short term memory). However, this finding needs to be considered with caution. While the present study did require the children to repeat sentences of different lengths, the sentences did not vary in length by a large number of words. Furthermore, it considered only their overall scores for the SR test in the analysis. A stronger claim in support of there being no role for PSTM in the repetition of sentences would be to match carefully the sentences of different lengths for grammatical complexity. An equal proportion of errors at each sentence length would confirm that PSTM was not activated at all for these sentences. This type of analysis could compare the clinical and non-clinical group data to enable better specification the locus of impairment in the case of children with speech and language difficulties.

Figure 8-2: Revised model of sentence repetition (including likely locus of imparitment for children with speech and language difficulties)



Furthermore, the present study cannot prove or disprove the role of semantics in the repetition of sentences, as this was not a factor that was explored.

8.6. Methodological limitations of the study

Some limitations of the study have been discussed already in the individual chapters and these should be considered in the interpretation of the findings of the study. Additional potential limitations are reviewed in the following summaries.

8.6.1. Task order and stimulus order

As discussed in the methods (section 2.4.), when completing the WR and NWR task, all the children completed the word stimuli before the non-word stimuli. In addition, the words and non-words were presented in a fixed order such that the 1 syllable items were presented first, followed by the two syllable items and so on. These approaches were consistent with other studies (e.g. Chiat and Roy, 2007, 2008; Roy and Chiat 2004). However it might mean that the children either performed less well on the NWR and/or the later (longer) stimuli due to fatigue effects. Alternatively, effects of lexicality and/or item length may have been reduced as a result of practice effects. This is particularly problematic if the participants were affected differentially by fatigue or practice effects, which might be the case, given the heterogeneous sample included in the sample. While this possibility cannot be ruled out and any possible effects cannot be assessed, the two tasks were quick in duration and few children appeared to be affected by either practice effects or fatigue effects.

In the design of the study, an attempt was made to counter-balance the repetition and naming tasks (see 2.4.3.). This was important because the same word stimuli were used in both tasks, thus leading to possible practice or priming effects for the individual word stimuli. Counter-balancing the task would have enabled analysis to ensue to determine whether any such effects were observed. However, unfortunately few of the children were compliant in completing the repetition task before the naming task. It cannot therefore be ruled out that some of the children's WR scores may have been enhanced by the previous experience of the same words in the naming task. Attempts were made however to induce forgetting of the stimuli by including the non-verbal assessment tasks between the naming and repetition tasks, and thus to minimise any such effects.

8.6.2. Consistency of assessment conditions for the children

A consideration when comparing the present findings to other studies that have investigated PSTM in young children is the way in which the stimuli were presented. For measures of PSTM it is generally recommended that the assessor's mouth is covered so that the participant cannot benefit from any visual cues in encoding, storing or recalling the stimuli. To ensure a consistent stimulus is received by all the children, studies often further use pre-recorded stimuli presented through headphones. Consistent with other studies that have included children of this young age (e.g. Chiat and Roy, 2007, 2008; Roy and Chiat, 2004), the present study opted for a live presentation approach. This approach was deemed to be more appropriate for engaging these young children. The researcher therefore made attempts to be consistent in the style and speed of presentation with the children.

A second consideration regarding the assessment conditions for the children relates to the environment and time of day that testing took place. Assessment visits occurred according to the parent's choice. Some of the children were assessed in their homes, while others were visited in nursery or, at T2, at school. One child was assessed in the speech therapy clinic room. Furthermore, for some children their parent or key-worker was present while others were assessed individually. This might have influenced the children's performance or willingness to engage. Again, whether the child was assessed with a familiar adult present was the choice of the parent and/or the child. Furthermore, some children were assessed in the morning, while others in the afternoon. No child was assessed at the time of day that they would ordinarily have had a nap, but it is possible that fluctuations in alertness might have contributed to performance as a result of the time of day they were assessed. Again, this meant that not all the children received exactly identical conditions and it is not possible to assess how, if at all, these conditions might have affected their responses.

8.6.3. Heterogeneous sample

As discussed in chapter 3, the sample included in the present study represented a heterogeneous sample of children with and without a range of speech and language difficulties. Arguably this makes results from the study more generalisable to other clinical populations, which likely also include children with diverse speech and language needs. However, it makes it difficult to assess the extent to which the findings are generalisable to the general population.

Furthermore it is possible that where the sample was split to allow comparison of clinical and non-clinical groups (chapter 5), different results may have been found had the clinical sample been divided further into children with speech-only difficulties and children with language-only difficulties. The inclusion of children with speech-only difficulties might have diluted observed non-effects that might otherwise have been present. For example, while no 2-way group by item length interaction was found in the ANOVA, and no 3-way interaction was uncovered, the graph (figure 5-1) indicates that there is a tendency for the clinical group to repeat longer known words (as well as non-words) less accurately than shorter known words. This tendency is clearly not the case for the non-clinical group. This apparent 'tendency' did not yield significant results, precluding further analysis or discussion in chapter 5. However, it is possible that the exclusion of children presenting with speech delay/disorder in the absence of any language difficulties may have resulted in a different pattern of results. Further analysis of this type could ensue in the present study, by categorising participants based on their scores on the PLS-4. However this would further reduce the size of the groups, therefore resulting in compromises in statistical power.

8.6.4. Judgement of whether a given word was known

Chapter two (section 2.2.) described two methods that the present study designed to assess the children's knowledge of the words used in the repetition task. The first was the naming task and the second was the recognition task. The latter was however abandoned, as it yielded some unexpected results that were difficult to explain. Some of the children had demonstrated knowledge of the words by articulating these correctly during the naming task. However, they did not demonstrate their knowledge for the same items in the recognition task. As the recognition task occurred for all the children at the end of the first visit, it was concluded that the children were most likely demonstrating fatigue effects in this task and that this explains the surprising results.

As the recognition task was judged not to be a reliable measure of the children's word knowledge, attention focussed on the naming task only. The children's responses to the first naming attempt only were considered. Clear procedures were followed to assess whether the child knew the word or not (see section 2.3.2) using also the data available regarding their phonological systems, but some of the children's responses were ambiguous due to unusual phonological distortions. It is possible therefore that some of the stimuli were considered known or unknown when the reverse was true. It should be emphasised however that this represented a very small sample of the data.

8.6.5. Perceptual account of repetition difficulties

Some studies have emphasised the role of perception in children's repetition accuracy (e.g. Coady, Kluender and Evans, 2005) and in their grammatical development (Joanisse and Seidenberg, 1998). Speech perception difficulties have been associated with speech disorders and language disorders (e.g. Rvachew, Ohberg, Grawburg and Heyding, 2003; Ziegler, Pech-Georgel, George, Alario and Lorenzi, 2005). The present study did not explore the participants' speech perception abilities so cannot rule this out as an underlying influence in the children's repetition accuracy. Tests of speech perception that are suitable for the age-group of the participants are described by Vance, Rosen and Coleman (2009) and inclusion of such tasks would have enabled evaluation of the extent to which such difficulties may have influenced the children's performance on the repetition task.

8.6.6. Presence of ceiling effects

A further limitation of the present study, particularly at T2 was the presence of ceiling effects. Ceiling effects were especially apparent on the WR task and the naming task. There is good reason for this. In the case of the naming task, this was originally not designed as a test of the children's general vocabulary. Instead, the task was intended as a test of their knowledge of the specific words used in the word repetition test for in the first part of the study (see chapter 3, 4). Items in the test were deliberately chosen that young children were likely to know (see design for the stimuli, section 2.2.2.2). The extent to which the test can be used as an approximation of the children's word knowledge more generally is therefore questionable. In the case of repetition of known words, ceiling effects are also unavoidable due to the simplicity of repeating a short familiar phonological chunk.

8.6.7. Assumptions made by the study

In its attempts to unpick the skills underlying WR, NWR and SR, the present study made the assumptions about skills measured by other tasks. Discussion has already focussed on the questionable ability of the picture naming task to indicate the participants' vocabularies. The study also made the assumption that the word span task was an accurate measure of PSTM and that the *word structure* task measures morphological skills. However, no task can be a perfect measure of a skill and other variables such as the child's motivation, ability to focus attention and task comprehension might also influence performance on any of the tests.

8.6.8. Correlational design

A limitation in the design of the study was its use of correlation to determine relationships. Consequently it is not possible to conclude whether relationships between the variables are present due to another or several other factors, or which variable causes change in the other. For example, while the present study indicates that word knowledge and PSTM explain a significant amount of the variance in NWR, and that grammar explains a significant amount of the variance in sentence repetition, a more robust way to prove the relationships between these variables, as emphasised by Klem et al. (2015) and Melby-Lervåg et al. (2012), would be through intervention studies (see section 8.7.).

8.7. Further research

One way to confirm the role of sub-lexical representations in NWR, and morphology in SR would be to make direct measures of the underlying skills and compare group performance on NWR and SR. In the case of NWR, older children could be assessed on their phonological awareness skills and this would be considered a measure of their phonological templates. Two groups would be formed: a group with 'strong-phonological templates' and a group with 'weak-phonological templates' group based on their performance on the tasks. Performance on a NWR tasks for each of the groups would be compared and it would be predicted that the group with 'strong phonological templates' would perform better on NWR than the group with 'weak phonological templates'. In the case of sentences, children would be assigned in a similar way based on their scores on a test that taps awareness/use of morphemes. SR performance for each of the groups could then be compared. and it would be predicted that the 'high morphological group' would perform better than the 'low morphological group'.

Alternatively, intervention studies that targeted these aspects of language (phonological templates and morphological templates) could be designed. In the case of single 'word' processing, studies could target phonological templates through increasing awareness of syllabic and phonological structures of words (and non-words). In the case of sentences, studies could target awareness of morphological aspects of the sentence (e.g. tense markings, prepositions). The studies could monitor direct improvements in these factors, and could also use NWR or SR as the baseline and outcome measures. However, one problem with this approach is the finding that NWR and SR continue to be challenging for people with a history of language disorders, even when their difficulties appear to have resolved (Bishop et al., 1996; Conti-Ramsden et al., 2001; Moll et al., 2015). The extent to which any change in NWR or SR might be expected following targeted therapy, and therefore their usefulness as outcome measures is put into question.

Some studies have focussed on training working memory, reporting variable gains in the skills trained and these studies are reviewed by Melby-Lervåg and Hulme (2013) and Shipstead, Reddick and Engle (2012). The authors of these reviews caution against using WM approaches in the case of children with developmental disorders, arguing that studies have demonstrated limited evidence in support of any long-term measurable changes in WM skills. This is particularly the case for PSTM tasks, although one exception is a study by Henry, Messer and Nash (2014). They report improvements in word span, following an intervention that focussed on executive loaded phonological WM in 5 ½ to 8 ½ year old children without language difficulties. The effects were maintained 6 months post-intervention. Whether this effect would also be observed in children with identified language and/or PSTM difficulties, and whether it would generalise to improved accuracy on NWR however has not been tested.

Melby-Lervåg and Hulme (2012) emphasise the need for intervention studies with children with language difficulties to target language skills. Existing intervention studies with young children have focussed on developing vocabulary (e.g. Girolametto, Pearce and Weitzman, 1996; Kouri, 2005; Munro, Lee and Baker, 2008; Riches, Tomasello and Conti-Ramsden, 2005), phonological awareness (e.g. Munro et al., 2008) and grammar (e.g. Ebbels, 2007). No studies were however found that use NWR or SR as baseline and outcome measures in intervention studies.

8.8. Summary of the findings and conclusions

Overall, results of the present study support previous suggestions that the reason that repetition tasks are useful clinically is due to them revealing weaknesses in language processing systems (e.g. van der Lely and Howard, 1995). While there is evidence that NWR also taps PSTM, there is limited support for this being the reason for children with speech and language difficulties showing impairments on the task. Grammatical skills showed the greatest influence on the SR task included in this study, therefore emphasising its use in language assessment clinically.

In terms of theoretical models, the present study supports those that include both existing language knowledge and PSTM in their explanation of performance on NWR and SR tasks. It found PSTM redundant in the case of repeating known words. Confirmation of the role of these and other skills would be achieved through therapy studies, which would have the additional ethical benefit of potentially helping the children who participate.

Appendices

Appendix A: Consent form for the clinical group

Central London Community Healthcare NHS
NHS Trust

Barnet | Hammersmith and Fulham | Kensington and Chelsea | Westminster

[Clinic address has been removed]

Tel: [removed]

Site Number:
Patient Identification number:

Consent Form

Project Title: Assessing Children's Language Skills

Name of researcher: Hannah Hockey (Speech and Language Therapist)

- I have read and understood the information leaflet relating to this research project
- I have had the opportunity to ask questions about the research project and I have received satisfactory answers to these questions
- I understand that my child and I are participating voluntarily
- I am aware that I can withdraw my child from the research at any time and I will not have to give a reason. I am aware that this will not affect my child's medical care or legal rights
- I agree for my child's GP to be informed of his/her participation in the research
- I agree for the researcher to audio-record the sessions with my child
- I agree for the researcher to have access to my child's speech and language therapy case notes
- I understand that information relating to my child will be stored securely and anonymously
- I agree that if the researcher visits my home, my address will be shared with a second member of the research team
- I agree for my child to take part in this research

Child's name:……………………………………………..

Name of Parent/carer giving
consent:…………………………………………………………..

Signature:…………………………………………………
Date:…………………..

cc. *parent/carer, GP, research file, case file*

Appendix B Consent form for the non-clinical group

Central London Community Healthcare NHS
NHS Trust

Barnet | Hammersmith and Fulham | Kensington and Chelsea | Westminster

[Clinic address has been removed]

Tel: [removed]

Site Number:
Patient Identification number:

Consent Form

Project Title: Assessing Children's Language Skills

Name of researcher: Hannah Hockey (Speech and Language Therapist)

- I have read and understood the information leaflet relating to this research project
- I have had the opportunity to ask questions about the research project and I have received satisfactory answers to these questions
- I understand that my child and I are participating voluntarily
- I am aware that I can withdraw my child from the research at any time and I will not have to give a reason.
- I understand that information relating to my child will be stored securely and anonymously
- I agree for my child to take part in this research
- I agree that if the researcher visits my home, my address will be shared with a second member of the research team

Child's name:.……………………………………………..
Date:…………………..

Name of Parent/carer giving consent:…………………………………………………..

Signature:…………………………………………………
Date:…………………..

cc. parent/carer, research file

Appendix C Video consent form

Central London Community Healthcare NHS
NHS Trust

Barnet | Hammersmith and Fulham | Kensington and Chelsea | Westminster

[Address of NHS clinic has been removed]
Tel: [removed]

Site Number:
Patient Identification number:

Video Consent Form

Project Title: Assessing Children's Language Skills

Name of researcher: Hannah Hockey (Speech and Language Therapist)

Please indicate by circling the appropriate response to the following statements:

- I am willing for the researcher to video record the sessions with my child.

Yes / No

Recordings will be stored anonymously and securely and will be confidentially destroyed at the end of the study.

If you are willing for the researcher to record the sessions with your child, please indicate (by ticking the appropriate boxes) which of the following statements apply:

	Yes	No
The recording of the session with my child may be watched by the researcher after the session for the purposes of scoring my child on the assessment tasks		
The recordings of the session with my child may be watched by other members of the research team (e.g. research supervisors)		
The recordings of the sessions with my child may be used for educational purposes, e.g. during presentation of the research at seminars and conferences. Recordings will maintain their anonymity. At the end of the research and presentations the tapes will be confidentially destroyed.		

Child's name:……………………………………………………..

Name of Parent/carer giving consent:………………………………………………

Signature:…………………..…………………………………….

Date:…………………..

cc. parent/carer, research file

Appendix D Information sheet for the clinical group

All research proposals are reviewed by a Research Ethics Committee to ensure the safety of participants. The proposal for this research project has been accepted by the Barking and Havering Research Committee. It has also been accepted by the Senate Ethical Committee at City University.

If you would like to complain about any aspect of the study, City University London has established a complaints procedure via the Secretary to Senate Research Ethics Committee. To complain about the study, you need to phone 020 7040 3040. You can then ask to speak to the Secretary to Senate Research Ethics Committee and inform them that the name of the project is: Assessing Children's Language Skills

You could also write to the Secretary at:

[redacted]
Secretary to Senate Research Ethics Committee
Research Office, E214

City University London
Northampton Square
London
EC1V 0HB

Email: [redacted]

[redacted]

Contact details:

Speech and Language Therapist / Researcher:

Hannah Hockey
Speech and Language Therapy
[Clinic address has been removed]

Tel: [removed]

E-mail: [redacted]

Academic supervisor:

Prof. Shula Chiat
Department of Language and Communication Science
City University
Northampton Square
London
EC1V 0HB

Tel: [redacted]

Speech and Language Therapy manager:

Claire Withey
Speech and Language Therapy
[Clinic address has been removed]

Tel: [removed]

Central London Community Healthcare NHS
NHS Trust

Barnet | Hammersmith and Fulham | Kensington and Chelsea | Westminster

INFORMATION

FOR

PARENTS & CARERS

[redacted]

Assessing Children's Language Skills

Research Project

Thank you for expressing an interest in participating in this research project.

I am a speech and language therapist who works with pre-school children. I am doing some research, which will help speech and language therapists to better understand language development, and how to help children who are having difficulties with language. The research will look at children's early language skills, and will investigate which early language difficulties might predict problems that will persist as children get older.

Why has your child been selected?
Your child's communication skills have been assessed by a speech and language therapist, who feels that he/she would benefit from support with his/her language development. I am contacting all parents/carers of children whose main language is English, and who have been identified as requiring this type of support. I am aiming to gather information from about 40 children.

Do you have to take part?
No. It is entirely your choice. After discussing the project with me you can decide whether you would like your child to be included in the research project. If you decide that you would like your child to take part, you can withdraw that decision at any time.

Participation in the research will allow more information to be gained about your child's communication difficulties and therefore might be advantageous to your child. If, during the course of the research, we find out information that will help your child's speech and language therapist and you to better understand these difficulties, I will check with you that you are happy for me to contact your child's speech and language therapist to pass the information on.

Will your child's participation in the research affect the speech and language therapy that he/she receives?
No. Your child will continue to receive the same amount and type of speech and language therapy that he/she would have received.

What will happen if you agree to take part?
If you would like your child to take part in the research, we will arrange a meeting between you, your child and me, where you will have the opportunity to ask questions about the research. I will ask you to sign a consent form and we will complete a questionnaire together about your child's communication. I will ask you if you would be happy for me to video record the sessions with your child. It is useful for me to record the interaction so that I can devote my full attention to your child during the sessions. I will watch the video recordings afterwards in order to score your child on the assessment tasks. The video tape would be stored securely and would be destroyed at the end of the study. If you would prefer me not to video record the sessions, you can choose for this not to happen.

We will arrange a second meeting for you and your child, where I will assess some specific areas of your child's language development. The assessment tasks are designed to be fun for your child and they might be similar to tasks that your child has done with the speech and language therapist that originally assessed his/her communication skills.

I will contact you again 12 months after this session, when I will use some different tasks to assess your child's language development.

Where possible these meetings will take place in your child's nursery, school or your home so that you will not have to travel.

At all stages of your child's involvement you will have the opportunity to discuss any questions or concerns relating to the research with me.

At any stage, if you feel that your child is unhappy, we can stop the tasks.

What happens when the research stops?
If you and your child's speech and language therapist have agreed that your child still needs support with his/her language development, he/she will continue to receive this support from the relevant speech and language therapy team (community clinic or mainstream school).

The information about your child's language development gathered from the research will be analysed together with information from the other children involved. You will be sent a summary of the research findings.

If you are willing for your child to take part in the research:

- Your child's GP will be informed of your child's participation.
- Information relating to your child will be stored anonymously.
- All information will be treated confidentially. However, should any situation arise that compromises the well-being of your child, I will need to pass on this information as part of my duty of care.
- If at any stage of the research you decide that you no longer wish your child to take part, you can withdraw your consent and the information relating to your child will not be used.

If you would like to know more details about this research project, please do not hesitate to contact me (contact details overleaf).

Appendix E Information sheet for the non-clinical group

All research proposals are reviewed by a Research Ethics Committee to ensure the safety of participants. The proposal for this research project has been accepted by the Barking and Havering Research Committee. It has also been accepted by the Senate Ethical Committee at City University London.

If you would like to complain about any aspect of the study, City University London has established a complaints procedure via the Secretary to Senate Research Ethics Committee. To complain about the study, you need to phone 020 7040 3040. You can then ask to speak to the Secretary to Senate Research Ethics Committee and inform them that the name of the project is: Assessing Children's Language Skills

You could also write to the Secretary at:

[redacted]
Secretary to Senate Research Ethics Committee
Research Office, E214

City University London
Northampton Square
London
EC1V 0HB

Email: [redacted]

Central London Community Healthcare NHS
NHS Trust

Barnet | Hammersmith and Fulham | Kensington and Chelsea | Westminster

INFORMATION

FOR

PARENTS & CARERS

[redacted]

Assessing Children's Language Skills

Research Project

Thank you for expressing an interest in participating in this research project.

I am a speech and language therapist who works with pre-school children. I am doing some research that will help speech and language therapists to better understand language development and how to help children who are having difficulties.

Why has your child been selected?

Although the main part of the research will investigate skills of children who have language difficulties, it is also vital to gather information from children who do not have identified difficulties. I am aiming to gather information on 40 children whose language development has not been identified as being a concern.

Do you have to take part?

No, it is entirely your choice. If you do decide that you would like your child to take part, you can still withdraw at any time.

What will happen if you agree to take part?

If you agree to your child taking part in the research, we will arrange a meeting between you, your child and me, where you will have the opportunity to ask questions about the research. I will ask you to sign a consent form and we will complete a word list together about your child's talking. I will ask you if you would be happy for me to video record the sessions with your child. It is useful for me to record the interaction so that I can devote my full attention to your child during the sessions. I will watch the video recordings afterwards in order to score your child on the assessment tasks. The video tape would be stored securely and would be destroyed at the end of the study. If you would prefer me not to video record the sessions, you can choose for this not to happen.

We will then arrange a suitable time for me to visit your child in nursery or at home, when I will assess specific areas of your child's language in a 1:1 setting. You would also be very welcome to attend this session, which would take place in a quiet room. The assessment tasks are designed to be fun for your child and involve age-appropriate toys.

At all stages of your child's involvement you will have the opportunity to discuss any questions or concerns relating to the research with me.

The information gathered from the research about your child's language skills will be stored anonymously. It will be analysed together with information from the other children involved. If you are interested, I will send you a summary of the research findings.

If you have any questions relating to the research please do not hesitate to contact me:

Hannah Hockey
Speech and Language Therapy
[Clinic address has been removed]

Tel: [removed]

E-mail: [redacted]

Academic supervisor:

Prof. Shula Chiat
Department of Language and Communication Science
City University London
Northampton Square
London
EC1V 0HB

Tel: [redacted]

Speech and Language Therapy Manager:

Claire Withey
Speech and Language Therapy
[clinic address has been removed]

Tel: [redacted]

Appendix F: Background Questionnaire

Background Questionnaire

Thank you for choosing to take part in this study. I would first like to ask you some basic questions in order to obtain some information about you and your child. It is possible that some of the questions may not be applicable to you. In such a case, please mark the question with N/A. Please answer the questions as fully and honestly as possible. All information obtained will be kept strictly confidential.

Child Background Information

1. Child's Date Of Birth ____________

2. Child's Age ____________

3. Child's Gender: MALE FEMALE

4. Do you have any concerns about your child's health or development (including hearing)?
YES NO

If yes please state them here

____

____

Child's Language (This section is only for those children who have more than one language)

1. Child's First Language: ___________________________________

2. Other Languages: ________________________________

3. Which language do you feel is your child's stronger language?

4. Which language is your child mostly spoken to at home?

5. How often does your child use their First Language?
Everyday Weekly Fortnightly Monthly Yearly
Other ________________________

6. How often does your child use their second or other language(s)?
Everyday Weekly Fortnightly Monthly Yearly
Other ________________________

Family Background

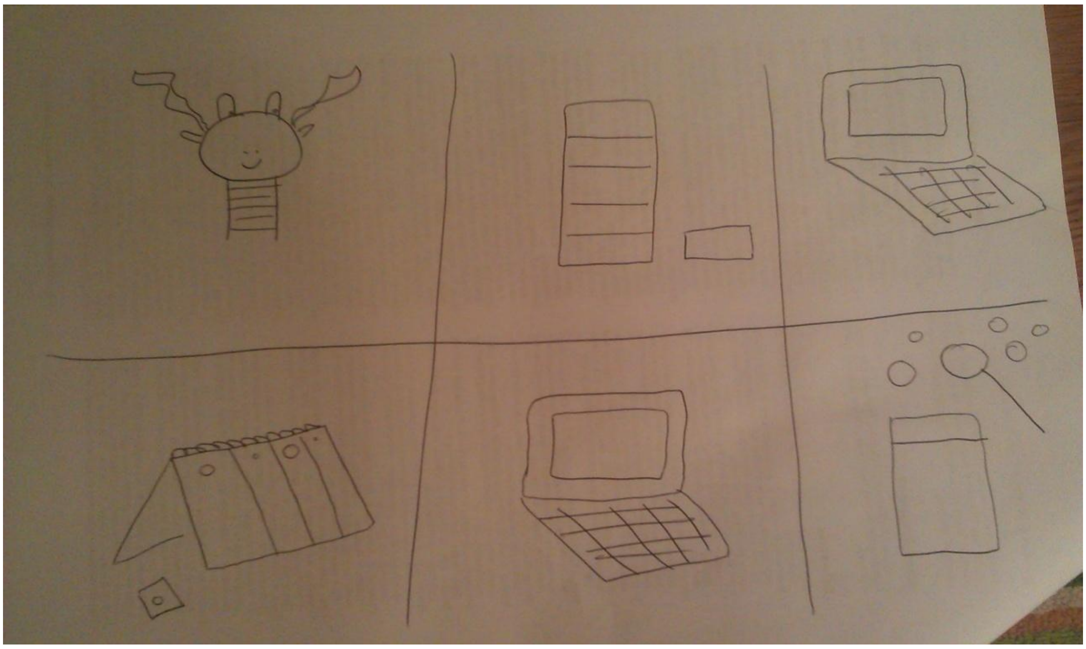
1. Mother's Highest level of Education Attained? ________________________________

2. Father's Highest level of Education Attained? ________________________________

3. Mother's Occupation: ___________________________

4. Father's Occupation: ___________________________

Appendix G: Example of visual timetable used during data collection



Appendix H: Tables of descriptive statistics

Table 1: PLS Auditory

	T1 clinical	T1 non-clinical	T1 overall	T2 clinical	T2 non-clinical	T2 overall
Mean	38.79	44.6	41.59	49.8	54.81	52.35
Standard deviation	6.15	5.08	6.33	6.53	4.31	6.01
Minimum	31	30	30	35	44	35
Maximum	52	55	55	60	61	61
No. participants	28	26	54	25	26	51

Table 2: PLS Expressive

	T1 clinical	T1 non-clinical	T1 overall	T2 clinical	T2 non-clinical	T2 overall
Mean	41.3	52.76	46.81	47.32	57.04	52.18
Standard deviation	6.19	6.2	8.43	4.78	3.71	6.48
Minimum	33	41	33	36	47	36
Maximum	55	64	64	54	66	66
No. participants	27	25	52	25	25	50

Table 3: BAS II Block Building

	T1 clinical	T1 non-clinical	T1 overall	T2 clinical	T2 non-clinical	T2 overall
Mean	4.04	4.65	4.33	9.56	10	9.78
Standard deviation	3.43	3.19	3.3	2.38	2.55	2.45
Minimum	0	1	0	6	6	6
Maximum	15	14	15	14	15	15
No. participants	28	26	54	25	26	51

Table 4: BAS II Picture Similarities

	T1 clinical	T1 non-clinical	T1 overall	T2 clinical	T2 non-clinical	T2 overall
Mean	16.46	16.11	16.3	20.2	21.9	21.04
Standard deviation	4.37	3.49	3.94	4.11	3.59	3.92
Minimum	10	8	8	12	15	12
Maximum	28	28	28	28	30	30
No. participants	28	26	54	26	26	52

Table 5: Naming task (scored for number of phonemes correct)

	T1 clinical	T1 non-clinical	T1 overall	T2 clinical	T2 non-clinical	T2 overall
Mean	64.5	96	79.7	89.5	110.4	100
Standard deviation	17.7	14	22.5	12.9	9.94	15.5
Minimum	34	68	34	65	93	65
Maximum	100	125	125	110	127	127
No.	28	26	54	26	26	52

participants						

Table 6: Naming task (number of items correct)

	T1 clinical	T1 non-clinical	T1 overall	T2 clinical	T2 non-clinical	T2 overall
Mean	19.2	21.5	20.3	22.8	24	23.4
Standard deviation	2.91	2.14	2.79	1.95	1.7	1.91
Minimum	13	18	13	18	19	18
Maximum	24	25	25	26	26	26
No. participants	28	26	54	26	26	52

Table 7: Known word repetition (number of phonemes correct)

	T1 clinical	T1 non-clinical	T1 overall	T2 clinical	T2 non-clinical	T2 overall
Mean	100.3	123.8	111.6	125.3	137.4	131.4
Standard deviation	26.9	10.7	23.7	17.3	5.42	14.1
Minimum	34	97	34	63	121	63
Maximum	138	136	138	144	148	148
No. participants	28	26	54	26	26	52

Table 8: Known word repetition (proportion of phonemes correct)

	T1 clinical	T1 non-clinical	T1 overall	T2 clinical	T2 non-clinical	T2 overall
Mean	0.72	0.854	0.784	0.942	0.982	0.962
Standard deviation	0.155	0.0711	0.139	0.0685	0.0212	0.543
Minimum	0.37	0.63	0.37	0.66	0.92	0.66
Maximum	0.95	0.93	0.95	1	1	1
No. participants	28	26	54	26	26	52

Table 9: Non-word repetition (number of phonemes correct)

	T1 clinical	T1 non-clinical	T1 overall	T2 clinical	T2 non-clinical	T2 overall
Mean	88	112.7	99.9	116.7	132.7	124.7
Standard deviation	31.2	18.7	28.6	17.2	6.13	15.1
Minimum	8	36	8	77	117	77
Maximum	139	130	139	147	148	148
No. participants	28	26	54	26	26	52

Table 10: Speech corrected non-word repetition (proportion of phonemes correct)

	T1 clinical	T1 non-clinical	T1 overall	T2 clinical	T2 non-clinical	T2 overall
Mean	0.674	0.831	0.75	0.846	0.938	0.892
Standard deviation	0.187	0.087	0.167	0.109	0.036	0.0924
Minimum	0.13	0.55	0.13	0.56	0.86	0.56
Maximum	0.95	0.96	0.96	1	1	1

No. participants	28	26	54	26	26	52

Table 11: Span task

	T1 clinical	T1 non-clinical	T1 overall	T2 clinical	T2 non-clinical	T2 overall
Mean	6.41	9.04	7.83	10.2	11.9	11
Standard deviation	4.25	2.88	3.77	4	3.01	3.61
Minimum	0	4	0	4	4	4
Maximum	16	14	16	20	18	20
No. participants	22	26	48	26	26	52

Table 12: Word Structure Task (assessed at Time 2 only)

	T2 clinical	T2 non-clinical	T2 overall
Mean	11.7	15.5	13.7
Standard deviation	5.41	4.5	5.27
Minimum	4	7	4
Maximum	22	22	22
No. participants	24	26	50

Table 13: Sentence repetition task (assessed at Time 2 only)

	T2 clinical	T2 non-clinical	T2 overall
Mean	14.6	19.1	16.8
Standard deviation	6.53	4.18	5.89
Minimum	1	12	1
Maximum	25	25	25
No. participants	25	25	50

Appendix I: Tests for normality

The Shapiro-Wilks test for normality was run for each task, first with the data from the full sample and then for the clinical and non-clinical groups separately. Tasks found to yield significantly non-normal data are given in red below.

Time 1:

Task	**Full sample**	**Clinical group**	**Non-clinical group**
PLS auditory	D(54)=0.973, p=0.253, NS	D(28)=0.912, p=0.220, NS	D(26)=0.956, p=0.318, NS
PLS expressive	D(52)=0.963, p=0.105, NS	D(27)=0.953, p=0.248, NS	D(25)=0.961, p=0.443, NS
BAS Blocks	D(54)=0.869, p<0.001	D(28)=0.858, p=0.001	D(26)=0.872, p=0.004
BAS Pictures	D(54)=0.892, p<0.001	D(28)=0.902, p=0.012	D(26)=0.848, p=0.001
Naming (scored for items correct)	D(54)=0.958, p=0.056, NS	D(28)=0.952, p=0.221, NS	D(26)=0.946, p=0.188, NS
Naming (scored for phonemes correct)	D(54)=0.967, p=0.143, NS	D(28)=0.979, p=0.819, NS	D(26)=0.944, p=0.167, NS
Word repetition (all item lengths)	D(54)=0.869, p<0.001	D(28)=0.951, p=0.209, NS	D(26)=0.882, p=0.206 NS
Non-word repetition (all item lengths)	D(54)=0.878, p<0.001	D(28)=0.953, p=0.232, NS	D(26)=0.716, p<0.001
Span	D(47)=0.136, p=0.029	D(21)=0.175, p=0.094, NS	D(26)=0.175, p=0.039
corrected 1 syllable word repetition	D(54)=0.763, p<0.001	D(28)=0.772, p<0.001	D(26)=0.714, p<0.001
corrected 1 syllable non-word repetition	D(54)=0.779, p<0.001	D(28)=0.845, p<0.001	D(26)=0.839, p<0.001
corrected 2 syllable word repetition	D(54)=0.696, p<0.001	D(28)=0.695, p<0.001	D(26)=0.919, p=0.042
corrected 2 syllable non-word repetition	D(54)=0.871, p<0.001	D(28)=0.921, p=0.037	D(26)=0.850, p<0.001
corrected 3 syllable word repetition	D(54)=0.708, p<0.001	D(28)=0.764, p<0.001	D(26)=0.873, p=0.004
corrected 3	D(54)=0.840,	D(28)=0.891,	D(26)=0.720, p<0.001

syllable non-word repetition	p<0.001	p=0.007	
corrected 4 syllable word repetition	D(54)=0.649, p<0.001	D(28)=0.780, p<0.001	D(26)=0.387, p<0.001
corrected 4 syllable non-word repetition	D(54)=0.889, p<0.001	D(28)=0.930, p=0.060, NS	D(26)=0.839, p=0.001

Time 2:

Task	Full sample	Clinical group	Non-clinical group
Sentence repetition	D(50)=0.936, p=0.010	D(25)=0.961, p=0.427 NS	D(25)=0.920, p=0.051 NS
Word structure	D(50)=0.959, p=0.081 NS	D(24)=0.955, p=0.354 NS	D(26)=0.947, p=0.199 NS
PLS auditory	D(51)=0.953, p=0.008	D(26)=0.956, p=0.236 NS	D(25)=0.931, p=0.089, NS
PLS expressive	D(50)=0.958, p=0.071, NS	D(25)=0.967, p=0.560 NS	D(25)=0.942, p=0.166 NS
BAS Blocks	D(51)=0.945, p=0.020	D(25)=0.929, p=0.084, NS	D(26)=0.954, p=0.285 NS
BAS Pictures	D(52)=0.937, p=0.009	D(26)=0.957, p=0.335 NS	D(26)=0.882, p=0.006
Naming (scored for items correct)	D(52)=0.912, p=0.001	D(26)=0.912, p=0.030	D(26)=0.877, p=0.005
Naming (scored for phonemes correct)	D(52)=0.964, p=0.119, NS	D(26)=0.974, p=0.720 NS	D(26)=0.938, p=0.123 NS
Word repetition (all item lengths)	D(52)=0.563, p<0.001	D(26)=0.610, p<0.001	D(26)=0.751, p<0.001
Non-word repetition (all item lengths)	D(52)=0.812, p=0.001	D(26)=0.894, p=0.012	D(26)=0.952, p=0.257 NS
Span	D(52)=0.944, p=0.045	D(26)=0.945, p=0.175 NS	D(26)=0.918, p=0.041
corrected 1 syllable word repetition	D(52)=0.488, p<0.001	D(26)=0.593, p<0.001	D(26)=0.376, p<0.001
corrected 1 syllable non-word repetition	D(52)=0.674, p<0.001	D(26)=0.781, p=0.001	D(25)=0.754, p<0.001
corrected 2 syllable word repetition	D(52)=0.669, p<0.001	D(26)=0.756, p<0.001	D(26)=0.694, p<0.001
corrected 2 syllable non-word repetition	D(52)=0.868, p<0.001	D(26)=0.902, p=0.018	D(26)=0.852, p=0.012
corrected 3 syllable word repetition	D(50)=0.403, p<0.001	D(24)=0.471, p<0.001	D(26)=0.503, p<0.001

corrected 3 syllable non-word repetition	D(50)=0.819, p<0.001	D(26)=0.873, p=0.004	D(26)=0.841, p=0.001
corrected 4 syllable word repetition	D(52)=0.787, p<0.001	D(26)=0.800, p<0.001	D(26)=0.794, p<0.001
corrected 4 syllable non-word repetition	D(50)=0.791, p<0.001	D(26)=0.808, p<0.001	D(26)=0.868, p=0.003

Appendix J: <u>Score distributions and results of non-parametric (or parametric) tests for chapter 4 and 5</u>

Chapter 4:

T1 comparisons between word repetition and non-word repetition

WR: $D(54)=1.88$, $p<0.001$, skew= -1.32, kurtosis=1.48;

NWR: $D(54)=0.15$, $p=0.004$, skew= -1.35, kurtosis=1.58.

A Wilcoxon signed rank test confirmed that children repeated words (Mdn=119.5) more accurately than non-words (Mdn=108.0), $T=5.13$, $p<0.001$).

T1 comparisons between known words, unknown words and non-words

Known words: $D(53)=0.158$, $p=0.002$, skew= -2.05, kurtosis=5.09;

unknown words: $D(53)=0.154$, $p=0.003$, skew= -0.69, kurtosis= -0.52;

non-words: $D(53)=0.155$, $p=0.003$, skew= -1.41, kurtosis= 2.02.

A Friedman's ANOVA revealed a Chi square value of 39.95, $p<0.001$. Post hoc testing using Wilcoxon tests, and making a Bonferroni correction so that results were considered significant if $p<0.001$, confirmed that known words (Mdn=0.83) were repeated more accurately than non-words (Mdn=0.72), $T=5.82$, $p<0.001$. Known words were also repeated more accurately than unknown words (Mdn=0.76), $T=3.98$, $p<0.001$). There was no significant difference in children's repetition of unknown words (Mdn=0.76) compared to non-words (Mdn=0.72), $T=1.30$, $p=0.19$.

T2 comparisons between word repetition and non-word repetition

words: $D(52)=0.25$, $p<0.001$, skew= -3.99, kurtosis=19.11;

nonwords: $D(52)=0.17$, $p<0.001$, skew= -1.91, kurtosis=4.18.

A Wilcoxon signed-rank test was conducted. This revealed that children repeated words (Mdn=143.5) significantly more accurately than non-words (Mdn=137.0), $T=4.92$, $p<0.001$.

T2 comparisons between known words, unknown words and non-words (parametric test results, as non-parametric test results are reported in main text)

A repeated measures ANOVA revealed that repetition was affected by the stimulus type: $F(1.55, 78.78)=5.72$, $p<0.01$, $\eta_p^2=0.101$ (Greenhouse-Geisser statistics are reported due to Mauchly's test of sphericity being significant). Planned comparisons applying a Bonferroni correction revealed that scores on known words (M=0.89, SD=0.09) were repeated more accurately than non-words (M=0.84, SD=0.10) $p<0.001$. There was no difference between the unknown words (M=0.86, SD=0.14) and the non-words ($p=0.54$) or between known words and unknown words ($p=0.31$).

T1 comparisons between speech corrected and uncorrected data for WR and NWR

Scores on the word repetition task were significantly higher following speech error correction (corrected Mdn=129.0; uncorrected Mdn=119.5), $T=6.40$, $p<0.001$. Scores were also enhanced for non-words following speech error correction (corrected Mdn=114.5; uncorrected Mdn=108.0), $T=6.33$, $p<0.001$. Consistent with the results reported using parametric tests, the results of the non-parametric test revealed that following speech error correction, words continued to be repeated more accurately than non-words, $T=5.50$, $p<0.001$.

T2 comparisons between speech corrected and uncorrected data for WR and NWR

Scores on the word repetition task were significantly higher following speech error correction (uncorrected Mdn=135.0; corrected Mdn=143.5), T=6.18, p<0.001. Scores were also enhanced for non-words following speech error correction (uncorrected Mdn=134.8; corrected Mdn=142.0), T=6.20, p<0.001. Consistent with the results reported using parametric tests, the results of the non-parametric test revealed that following speech error correction, words continued to be repeated more accurately than non-words, T=4.92, p<0.001.

T1 comparisons of words and non-words by item length

A Friedman's ANOVA was used to explore effects of length for word repetition at T1. There was a significant effect of length, χ^2 (3)=17.74, p=0.001. Wilcoxon tests were used to explore the significant effect. A Bonferroni correction was applied (i.e. 0.05 divided by the number of comparisons made, which was 3) so all effects are reported at a 0.0167 level of significance. Wilcoxon tests revealed that 1 syllable words (Mdn=0.93) were repeated more accurately than 2 syllable words (Mdn=0.90). No other comparisons were significant.

A Friedman's ANOVA was also used to explore effects of length for non-word repetition. There was a significant effect of length χ^2 (3)=63.84, p<0.001. Again, Wilcoxon tests were used to explore the significant effect. A Bonferroni correction was applied so tests were considered significant if p<0.0001. Wilcoxon tests revealed that 1 syllable non-words (*Mdn*=0.92) were repeated more accurately than 2 syllable non-words (*Mdn*=0.81) T=12, p<0.001, which were repeated more accurately than 3 syllable non-words (*Mdn*=0.76) T=17.3, p<0.001, which were repeated more accurately than 4 syllable non-words (*Mdn*=0.68) T=25.5, p<0.001.

T2 comparisons of words and non-words by item length

A friedman's ANOVA revealed a significant effect of length χ^2 (3)=14.06, p<0.005. Wilcoxon signed-rank tests were used to explore the significant effect. A Bonferroni correction was applied so test results are considered significant if p<0.001. Wilcoxon tests revealed that none of the comparisons were significant at p<0.001. Furthermore, median values for 2 syllable, 3 syllable and 4 syllable words were all 1.00 highlighting the problematic ceiling effects.

A Friedman's ANOVA was also used to explore effects of length for non-word repetition at T2. The results were consistent with the parametric tests. There was a significant effect of length χ^2 (3)=37.97, p<0.001. Again, Wilcoxon tests were used to explore the significant effect. A Bonferroni correction was applied so results are considered signficant if p<0.001. The tests revealed that 3 syllable non-words (*Mdn*=0.95) were repeated more accurately than 4 syllable non-words (*Mdn*=0.86) T=22.1, p<0.001. No other tests were significant.

Chapter 5: Comparisons between the groups (clinical and non-clinical) on repetition of words and non-words

Time 1: Consistent with the parametric tests, there was a significant difference between the groups (clinical and non-clinical) on both word repetition and non-word repetition at p<0.001. Wilcoxon signed rank tests were used to explore differences between the groups on their repetition of words and non-words. These revealed that the non-clinical group repeated words (Mdn=0.945) more accurately than non-words (Mdn=0.838). The clinical group also repeated words (Mdn=0.821) more accurately than non-words (Mdn=0.728). No further analyses exploring lengths were conducted.

Time 2: Consistent with the parametric tests, there was a significant difference between the groups (clinical and non-clinical) on both word repetition and non-word repetition at $p<0.001$. Wilcoxon signed rank tests were used to explore differences between the groups on their repetition of words and non-words. These revealed that the non-clinical group repeated words (Mdn=0.991) more accurately than non-words (Mdn=0.941). The clinical group also repeated words (Mdn=0.967) more accurately than non-words (Mdn=0.865). No further analyses exploring lengths were conducted.

Appendix K: Results for chapter 4 when speech errors were not corrected

Time 1 length comparisons before speech correction, using parametric tests (results that do not support a 'word' length effect are given in bold)

Length comparison	Known Words	Non-words
1 syllable and 2 syllable	1 syll > 2 syll: t(49)=7.96, p<0.001	1 syll > 2 syll: t(49)=8.11, p<0.001
1 syllable and 3 syllable	1 syll > 3 syll: t(49)=3.42, p=0.001	1 syll > 3 syll: t(49)=7.26, p<0.001
1 syllable and 4 syllable	**1 syll = 4 syll: t(49)=0.62, p=0.54**	1 syll > 4 syll: t(49)=7.74, p<0.001
2 syllable and 3 syllable	**2 syll < 3 syll: t(49)= - 4.74, p<0.001**	**2 syll = 3 syll: t(49)=2.01, p=0.050**
2 syllable and 4 syllable	**2 syll < 4 syll: t(49)= - 4.46, p<0.001**	2 syll > 4 syll: t(49)=3.99, p<0.001
3 syllable and 4 syllable	**3 syll = 4 syll: t(49)= -1.80, p=0.078, NS**	3 syll > 4 syll: t(49)=3.43, p=0.001

Time 1 speech uncorrected known word score and non-word score correlations (r, where N=47) with span:

	Known WR	NWR	Span
Known WR	1		
NWR	0.77 (p<0.001)	1	
Span	0.24 (p=0.11, NS)	0.48 (p=0.001)	1

When the variance shared with WR was partialled out of the correlation between NWR and span, this continued to be significant: r(44)=0.47, p<0.001. This was not the case when variance shared with NWR was partialled out of the correlation between WR and span: r(44)= -0.18, p=0.23, NS.

Regression modelling, where NWR was the dependent variable and WR and span were inserted as predictors resulted in a significant model (F(2,44)=35.77, p<0.001), explaining 60.2% of the variance. Both WR (t=6.67, p<0.001) and span (t=3.48, p=0.001) were significant predictors.

Regression modelling, where WR was the dependent variable and NWR and span were inserted as predictors resulted in a significant model (F(2,44)=24.85, p<0.001), explaining 50.8% of the variance. NWR (t=6.67, p<0.001) was a significant predictor, but span was not (t= -1.23, p=0.23, NS).

Time 2 length comparisons before speech correction (results that do not support a 'word' length effect are given in bold)

Length comparison	Known Words	Non-words
1 syllable and 2 syllable	1 syll > 2 syll: t(50)=11.95, p<0.005	1 syll > 2 syll: t(50)=7.99, p<0.001
1 syllable and 3 syllable	1 syll > 3 syll: t(50)=3.14, p=0.003 NS	1 syll > 3 syll: t(50)=6.96, p<0.001
1 syllable and 4 syllable	**1 syll = 4 syll: t(50)= -1.31, p=0.198**	1 syll > 4 syll: t(49)=6.67, p<0.001
2 syllable and 3 syllable	**2 syll < 3 syll: t(50)= -10.35, p<0.001**	**2 syll = 3 syll: t(50)=0.75, p=0.293**
2 syllable and 4 syllable	**2 syll = 4 syll: t(50)= -11.84, p<0.001**	**2 syll = 4 syll: t(50)=0.18, p=0.857**
3 syllable and 4 syllable	**3 syll = 4 syll: t(50)= -4.42, NS**	**3 syll = 4 syll: t(50)= -0.48, p=0.634**

Time 2 speech uncorrected known word score and non-word score correlations (r, where N=52) with span:

	Known WR	NWR	span
Known WR	1		
NWR	0.72, p<0.001	1	
Span	0.29, p=0.03	0.43, p<0.001	1

When the variance shared with WR was partialled out of the correlation between NWR and span, this continued to be significant: r(49)=0.34, p=0.016. This was not the case when variance shared with NWR was partialled out of the correlation between WR and span: r(49)= -0.03, p=0.82, NS.

Regression modelling, where NWR was the dependent variable and WR and span were inserted as predictors resulted in a significant model (F(2,44)=33.75, p<0.001), explaining 56.2% of the variance. Both WR (t=6.75, p<0.001) and span (t=2.49, p=0.016) were significant predictors.

Regression modelling, where WR was the dependent variable and NWR and span were inserted as predictors resulted in a significant model (F(2,49)=27.26, p<0.001), explaining 50.7% of the variance. NWR (t=6.75, p<0.001) was a significant predictor, but span was not (t= -0.23, p=0.82, NS).

Appendix L: Difficulty analysis of the word stimuli

Item	No. of syllables	Presence of consonant clusters?	Typical/ atypical stress pattern	No. /proportion of children unable to label picture at phase one		No. /proportion of children unable to label picture at phase two	
				No.	proportion	No.	proportion
Toe	1	No	-	22	0.41	13	0.25
Glove	1	**Yes**	-	13	0.24	3	0.06
Tiger	2	No	Typical	14	0.26	5	0.10
Tractor	2	**Yes,** word initially and word medially across syllables	Typical	11	0.20	6	0.12
Princess	2	**Yes,** word initially and word medially across syllables	**Atypical***	20	0.37	5	0.10
Guitar	2	No	**Atypical**	16	0.30	13	0.25
Dinosaur	**3**	No	Typical	13	0.24	2	0.04
Kangaroo	**3**	No	**Atypical**	26	0.48	3	0.06
Trampoline	**3**	**Yes,** word initially and word medially, across syllables	**Atypical**	16	0.30	4	0.08
Caterpillar	**4**	No	Typical	18	0.33	5	0.10
Helicopter	**4**	**Yes** (word medially, across syllables)	Typical	10	0.19	1	0.02
Binoculars	**4**	**Yes,** word medially, within syllable	**Atypical**	46	0.85	27	0.52
Harmonica	**4**	No	**Atypical**	54	1	52	1
Macaroni	**4**	No	**Atypical**	53	0.98	50	0.96
Avocado	**4**	No	**Atypical**	48	0.89	43	0.83

(* *Princess* generally has an atypical (iambic) stress pattern when articulated as a word in isolation)

Appendix M: Results for chapter 7 when speech errors were not corrected

Correlations between NWR and WR and SR when speech errors have not been corrected are presented in the tables below:

At T1, N=50.

T1	Known WR	NWR	Sentence repetition
Known WR	1		
NWR	0.78, p<0.001	1	
Sentence repetition	0.28, p=0.05	0.36, p=0.009	1

At T2, N=50

T2	Known WR	NWR	Sentence repetition
Known WR	1		
NWR	0.73, p<0.001	1	
Sentence repetition	0.12, p=0.40, NS	0.37, p=0.008	1

References

Acheson, D. J. and MacDonald, M. C. (2009). Twisting tongues and memories: explorations of the relationship between language production and verbal short-term memory. *Journal of Memory and Language*, **60**, 329-350.

Adams, A-M. and Gathercole, S. E. (1995). Phonological working-memory and speech production in pre-school children. *Journal of Speech and Hearing Research*, **38** (2) 403-414.

Alloway, T. P. (2007). Investigating the roles of phonological and semantic memory in sentence recall. *Memory*, **15** (6), 605-615.

Alloway, T. P. and Gathercole, S. E. (2005). Working memory and short-term sentence recall in young children. *European Journal of Cognitive Psychology*, **17** (2), 207-220.

Alloway, T. P., Gathercole, S. E., Willis, C. and Adams, A-M. (2004). A structural analysis of working memory and related cognitive skills in young children. *Journal of Experimental Child Psychology*, **87**, 85-106.

Archibald, L. M. D. and Gathercole, S. E. (2006). Non-word repetition: a comparison of tests. *Journal of Speech, Language and Hearing Research*, **49**, 970-983.

Archibald, L. M. D. and Gathercole, S. E. (2007a). Nonword repetition and serial recall: equivalent measures of verbal short-term memory? *Applied Psycholinguistics*, **28**, 587-606.

Archibald . M. D. and Gathercole, S. E. (2007b). Nonword repetition in specific Language Impairment: more than a short-term memory deficit? *Psychonomic Bulletin and Review*, **14** (5), 919-924.

Baddeley, A. D. (1986). Working memory. Oxford: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, **4**, 417–423.

Baddeley, A. D. and Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), The psychology of learning and motivation, (Vol. 8, pp. 47-90). New York: Academic Press.

Baddeley, A. D., Hitch, G. J. and Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language*, **61**, 438-456.

Baddeley, A.D., Gathercole, S.E., and Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, **105**, 158–173.

Baddeley, A.D., Thomson, N. and Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, **14**, 575-589.

Bird, H., Franklin, S. and Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behaviour Research Methods Instruments and Computers*, **33** (1), 73-79.

Bishop, D. V. M., Adams, C. V., and Norbury, C. F. (2006). Distinct genetic influences on grammar and phonological short-term memory deficits: Evidence from 6-year-old twins. *Genes, Brain and Behavior*, **5**, 158–169.

Bishop, D.V.M. (2003). Test for reception of grammar (2nd edition). London: Psychological Corporation

Bishop, D. V. M., (2006). Beyond words: phonological short-term memory and syntactic impairment in specific language impairment. *Applied Psycholinguistics*, **27** (4), 545-547.

Bishop, D.V.M., (2014). Ten questions about terminology for children with unexplained language problems*. International Journal of Language and Communication Disorders*, **49** (4), 381-415.

Bishop, D. and Byng, S. (1984). Assessing semantic comprehension – methodological considerations, and a new clinical test. *Cognitive Neuropsychology,* **1** (3), 133-143.

Bishop, D. V. M., North, T., and Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: evidence from a twin study. *Journal of Child Psychology and Psychiatry*, ***37*** (4), 391-403.

Bishop, D.V.M. and Snowling, M.J. (2004). Developmental dyslexia and specific language impairment: same or different? *Psychological Bulletin*, **130** (6), 885-886.

Botting, N. and Conti-Ramsden, G. (2001). Non-word repetition and language development in children with specific language impairment. *Journal of Language and Communication Disorders*, **36** (4), 421-432.

Bolderson, S., Dosanjh, C., Milligan, C., Pring, T. and Chiat, S. (2011). Colourful semantics: a clinical investigation. *Child Language Teaching and Therapy*, **27** (3), 344-353.

Bowey, J. A. (2006). Clarifying the phonological processing account of nonword repetition. *Applied Psycholinguistics*, 27 (4), 548-554.

Boyle, M. K. and Gerken, L. (1997). The influence of lexical familiarity on children's function morpheme omissions: a nonmetrical effect? *Journal of Memory and Language*, **36**, 117-128.

Broomfield, and Dodd, B. (2004). Children with speech and language disability: caseload characteristics. *International Journal of Language and Communication Disorders,* **39** (3), 303-324.

Burt, L., Holm, H. and Dodd, B. (1999). Phonological awareness skills of 4-year-old British children: an assessment and developmental data. *International Journal of Language and Communication Disorders*, **34** (3), 311-335.

Carter, A. and Gerken, L. (2003). Similarities in weak syllable omissions between children with specific language impairment and normally developing language: a preliminary report. *Journal of Communication Disorders*, **36**, 165-179.

Casalini, C., Brizzolara, D., Chilosi, A., Cipriani, P., Marcolini, S., Pecini, C., Roncoli, S. and Burani, C. (2007). Non-word repetition in children with specific language impairment: a deficit in phonological working memory or in long-term verbal knowledge? *Cortex,* **43** (6), 769-776.

Cherry, E. C. (1953). Some experiments upon the recognition of speech, with one and with two ears. *Journal of the Acoustical Society of America*, **25**, 975-979.

Chiat, S. (2006). The developmental trajectory of non-word repetition. *Applied Psycholinguistics*, **27** (4), 552-556.

Chiat, S. and Hunt, J. (1993). Connections between phonology and semantics: an exploration of lexical processing in a language-impaired child. *Child Language Teaching and Therapy*, **9**, 200-213.

Chiat, S. and Roy, P. (2007). The Pre-school Repetition Test: an evaluation of performance in typically developing and clinically referred children. *Journal of Speech, Language and Hearing Research*, **50**, 429-443.

Chiat, S. and Roy, P. (2008). Early phonological and socio-cognitive skills as predictors of later language and social communication outcomes. *Journal of Child Psychology and Psychiatry,* **49** (6), 635-645.

Clegg, J. Hollis, C., Mawhood, L. and Rutter, M. (2005). Developmental language disorders a follow-up in later life. Cognitive, Language and psychosocial outcomes. *Journal of Child Psychology and Psychiatry*, **46** (2), 128-149.

Coady, J. A. and Evans, J. L. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairment. *International Journal of Language and Communication Disorders*, **43** (1), 1-40.

Coady, J. A., Kluender, K. R. and Evans, J. L. (2005). Categorical perception of speech by children with specific language impairment. *Journal of Speech, Language and Hearing Research,* **48** (4), 944-959.

Coady, J. A., Mainela-Arnold, E. and Evans, J. L. (2013). Phonological and lexical effects in verbal recall by children with specific language impairments. *International Journal of Language and Communication Disorders*, **48** (2), 144-159

Cohen, J. (1988). Statistical power analysis for the behavioural sciences (2nd edition). New York: Academic Press

Colle, H. A. and Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, **15**, 17-32.

Conboy, B. T. and Thal, D. J. (2006). Ties between the lexicon and grammar: cross-sectional and longitudinal studies of bilingual toddlers. *Child Development*, **77** (3), 712-735.

Conrad, R. and Hull, A. J. (1964). Information, acoustic confusion, and memory span. *British Journal of Psychology*, **55**, 429-432.

Conti-Ramsden, G. (2003). Processing and linguistic markers in young children with specific language impairment (SLI). *Journal of Speech, Language and Hearing Research*, **46** (5), 1029-1037.

Conti-Ramsden, G. and Botting, N. (1999). Classification of children with language impairment: longitudinal considerations. *Journal of Speech, Language and Hearing Research*, **42**, 1195-1204.

Conti-Ramsden, G., Botting, N. and Faragher, B. (2001). Psycholinguistic markers for specific language impairment. *Journal of Psychology and Psychiatry*, **42** (6), 741-748.

Conti-Ramsden, G., Botting, N., Simkin, Z. and Knox, E. (2001). Follow-up of children attending language units: outcomes at 11 years of age. *International Journal of Language and Communication Disorders*, **31** (2), 207-219.

Conti-Ramsden, G., Mok, P.L.H., Pickles, A. and Durkin, K. (2013). Adolescents with a history of specific language impairment (SLI): strengths and difficulties in social, emotional and behavioural functioning. *Research in Developmental Disabilities*, **34** (11), 4161-4169.

Cowan, N. (1995). Attention and memory: An integrated framework. New York, NY, US: Oxford University Press.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research,* **169**, 323-338.

Craik, F.I.M. and Lockhart, R.S. (1972). Levels of processing –framework for memory research. *Journal of Verbal Learning and Verbal Behaviour*, **11** (6), 671-684.

Dahlin, E., Nyberg, L., Bäckman, L., and Neely, A. (2008). Plasticity of executive functioning in young and older adults: Immediate training gains, transfer, and long-term maintenance. *Psychology of Aging*, ***23***, 720–730.

De Bree, E., Rispens, J. and Gerrits, E. (2007). Non-word repetition in Dutch children with (a risk of) dyslexia and SLI. *Clinical Linguistics and Phonetics*, **21** (11-12), 935-944.

Devescovi, A. and Caselli, C. M. (2007). Sentence repetition as a measure of early grammatical development in Italian. *International Journal of Language and Communication Disorders*, **2**, 187-208.

Devescovi, A., Caselli, M. C., Marchione, D., Pasqualetti, P., Reilly, J. and Bates, E. (2005). The cross-linguistic study of the relationship between grammar and lexical development. *Journal of Child Language*, **32**, 759-786.

Dispaldro, M., Benelli, B., Marcolini, S. and Giacomo, S. (2009). Real-word repetition as a predictor of grammatical competence in Italian children with typical language development. *International Journal of Language and Communication Disorders*, **44** (6), 941-961.

Dispaldro, M., Deevy, P., Altoé, G., Benelli, B. and Leonard, L. (2011). A cross-linguistic study of grammatical competence in children with typical language development. *International Journal of Language and Communication Disorders*, **46** (5), 564-578.

Dispaldro, M., Leonard, L. and Deevy, P. (2013a). Clinical markers in Italian-speaking children with and without specific language impairment: a study of non-word and real word repetition as predictors of grammatical ability. *International Journal of Language and Communication Disorders*, **48** (5), 554-564.

Dispaldro, M., Leonard, L. and Deevy, P. (2013b). Real-word and nonword repetition in Italian-speaking children with specific language impairment: a study of diagnostic accuracy. *Journal of Speech, Language and Hearing Research*, **56**, 323-336.

Dixon, J. A. and Marchman, V. A. (2007). Grammar and the lexicon: developmental ordering in language acquisition. *Child Development,* **78** (1), 190-212.

Dollaghan, C., Biber, M. E., and Campbell, T. F. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics*, **38** (4), 864-875.

Dollaghan, C. and Campbell, T.F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language and Hearing Research*, **41**, 1136-1146.

Dodd, B. and Bradford, A. (2000). A comparison of three therapy methods for children with different types of developmental phonological disorder. *International Journal of Language and Communication Disorders*, **35** (2), 189-209.

Dodd, B., Hua, Z., Crosbie, S., Holm, A. and Ozanne, A. (2002). Diagnostic Evaluation of Articulation and Phonology. London: Pearson

Dunn, L. M., Dunn, D. M., Styles, B., and Sewell, J. (2009). <u>The British Picture Vocabulary Scale III – 3rd Edition</u>. London: GL Assessment

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., and Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? Journal of Experimental Psychology: *Human Perception and Performance*, **25** (6), 1568-1578.

Ebbels, S. H. (2007). Teaching grammar to school-aged children with specific language impairment using shape coding. *Child Language Teaching and Therapy*, **23**, 67–93.

Edwards, J., Beckman, M. E. and Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language and Hearing Research*, **47**, 421-436.

Elliot, C. D., Smith, P. and McCullough, K. (1996). <u>British Ability Scales II</u>. NFER Nelson Publishing Company Ltd: Berkshire, UK.

Ellis Weismer, S., Tomblin, J. B., Zhang, X., Buckwalter, P., Chynoweth, J. G. and Jones, M. (2000). Nonword repetition performance in school-age children with and without language impairment. *Journal of Speech and Hearing Research*, **43** (3), 865-878.

Fallon, A. B., Groves K. and Tehan, G. (1999). Phonological similarity and trace degradation in the serial recall task: When CAT helps RAT but not MAN. *International Journal of Psychology*, **34** (5-6), 301-307.

Field, A. (2009). <u>Discovering Statistics Using SPSS (3rd Edition) (and sex and drugs and rock 'n' roll.</u> London: Sage

Gallon, N., Harris, J. and van der Lely, H. (2007). Non-word repetition: an investigation of phonological complexity in children with grammatical SLI. *Clinical Linguistics and Phonetics*, **21** (6), 435-455.

Gardner, H., Froud, K., McClelland, A. and van der Lely, H. K. J. (2006). Development of the Grammar and Phonology Screening (GAPS) test to assess key markers of specific language and literacy difficulties in young children. *International Journal of Language and Communication Disorders,* **41** (5), 513-540.

Gathercole, S. E. (1995a). Non-word repetition: more than just a phonological output task. *Cognitive Neurospychology*, **12** (8), 857-861.

Gathercole, S. E. (1995b). Is nonword repetition a test of phonological memory or long-term knowledge –it all depends on the non-words. *Memory and Cognition*, **23** (1), 83-94.

Gathercole, S. E. (2006). Non-word repetition and word learning: the nature of the relationship. *Applied Psycholinguistics,* **27**, 513-543.

Gathercole, S. E. and Adams, A-M. (1993). Phonological working memory in very young children. *Developmental Psychology,* **29** (4), 770-778.

Gathercole, S. E. and Adams, A-M. (1994). Children's phonological working memory: contributions of long-term knowledge and rehearsal. *Journal of Memory and Language*, **33**, 672-688.

Gathercole, S. E. and Baddeley, A. D. (1990a). Phonological deficits in language disordered children: is there a causal connection? *Journal of Memory and Language*, **29**, 336-360.

Gathercole, S. E. and Baddeley, A. D. (1990b). The role of phonological memory in vocabulary acquisition: a study of young children learning new words. *British Journal of Psychology*, **81**, 439-454.

Gathercole, S. E. and Baddeley, A. D. (1995). Working Memory and Language, Lawrence Erlbaum Associates Ltd.: East Sussex

Gathercole, S. E., Pickering, S. J., Ambridge, B. and Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology*, **40** (2), 177-190.

Gathercole, S. E., Pickering, S. J., Hall, M. and Peaker, S. J. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *Quarterly Journal of Experimental Psychology*, **54A** (1), 1-30.

Gathercole, S. E., Tiffany, C., Briscoe, J. and Thorn, A. (2005). Developmental consequences of poor phonological short-term memory function in childhood: a longitudinal study. *Journal of Psychology and Psychiatry*, **46** (6), 598-611.

Gathercole, S. E., Willis, C. S., Emslie, H. and Baddeley, A. D. (1992). Phonological memory and vocabulary development in the early school years: a longitudinal study. *Developmental Psychology,* **28** (5), 887-898.

Gathercole, S. E., Willis, C., Emslie, H., and Baddeley, A. D. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory*, **2**, 103-127.

Gerken, L. A. (1994). Young children's representation of prosodic phonology: evidence from English-speakers' weak-syllable productions. *Journal of Memory and Language*, **33**, 19-38.

Gerken, L. A. (1996). Prosodic structure in young children's language production. *Language*, **72**, (4), 683-712.

Gilhooly, K. J. and Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity and ambiguity measures for 1,944 words. *Behaviour Research Methods and Instrumentation*, **12** (4), 395-427.

Girbau, D. and Schwartz, R. G. (2007). Non-word repetition in Spanish-speaking children with specific language impairment (SLI). *International Journal of Language and Communication Disorders*, **42** (1), 59-75.

Girolametto, L., Pearce, P. S. and Weitzman, E. (1996). Interactive focused stimulation for toddlers with expressive vocabulary delays. *Journal of Speech and Hearing Research*, **39** (6), 1274-1283.

Goldinger, S. D., Luce, P. A. and Pisoni, D. B. (1989). Priming lexical neighbours of spoken words –effects of competition and inhibition. *Journal of Memory and Language,* **28** (5), 501-518.

Goto, H. (1971). Auditory perception by normal Japanese adults of sounds l and r. *Neuropsychologia*, **9** (3), 317-323.

Graf Estes, K. G., Evans, J. L. and Else-Quest, N. M. (2007). Differences in the non-word repetition performance of children with and without specific language impairment: a meta-analysis. *Journal of Speech, Language and Hearing Research*, **50**, 177-195.

Gray, S. (2003). Diagnostic accuracy and test-retest reliability of nonword repetition and digit span tasks administered to preschool children with specific language impairment. *Journal of Communication Disorders*, **36** (2), 129-151.

Gray, S. (2004). Word learning by preschoolers with specific language impairment: predictors and poor learners. *Journal of Speech, Language and Hearing Research*, **47**, 1117-1132.

Gray, S. (2005). Word learning by preschoolers with specific language impairment: effects of phonological and semantic cues. *Journal of Speech, Language and Hearing Research*, **48**, 1452-1467.

Gray, S. (2006). The relationship between phonological memory, receptive vocabulary and fast-mapping in young children with specific language impairment. *Journal of Speech, Language and Hearing Research*, **49**, 955-969.

Hanten, G., and Martin, R.C. (2000). Contributions of phonological and semantic short-term memory to sentence processing: Evidence from two cases of closed head injury in children. *Journal of Memory and Language*, ***43***, 335–361.

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

Henry, L. A. (1991). Development of auditory memory span - the role of rehearsal. *British Journal of Developmental Psychology*, **9** (4), 493-511.

Henry, L. A., Messer, D. J. and Nash, G. (2014). Testing for near and far transfer effects with a short, face-to-face adaptive working memory training intervention in typical children. *Infant and Child Development*, **23**, 84-103.

Hoff, E., Core, C. and Bridges, K. (2008). Non-word repetition assesses phonological memory and is related to vocabulary development in 20-24 month olds. *Journal of Child Language,* **35,** 903-916.

Hulme, C., Maughan, S., and Brown, G.D.A. (1991). Memory for familiar and unfamiliar words: Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language*, ***30***, 685–701.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S. and Stuart, G. (1997). Word-frequency effects on short-term memory tasks: evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology –Learning, Memory and Cognition*, **23** (5), 1217-1232.

Hulme, C., Thomson, N., Muir, C. and Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology*, **38**, 241 -253.

Jarvella, R.J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behaviour*, **10**, 409-416.

Jefferies, E., Lambon Ralph, M. A., and Baddeley, A. D. (2004). Automatic and controlled processing in recall: The role of long-term and working memory. *Journal of Memory and Language*, **51**, 623–643.

Joanisse, M. F. and Seidenberg, M. S. (1998). Specific language impairment: a deficit in grammar or processing? *Trends in Cognitive Sciences*, **2** (7), 240-247.

Kalnak, N., Peyrard-Janvid, M., Forssberg, H. and Sahlen, B. (2014). Nonword repetition – a clinical marker for specific language impairment in Swedish associated with parents' language-related problems. *Plos One,* **9** (2), e89544.

Kapalková, S., Polišenská, K. and Vicenová, Z. (2013). Non-word repetition performance in Slovak-speaking children with and without SLI: novel scoring methods. *International Journal of Language and Communication Disorders*, **48** (1), 78-89.

Kay, J., Lesser, R. and Coltheart, M. (1992). Psycholinguistic Assessment of Language Processing in Aphasia. Psychology Press.

Klem, M., Melby-Lervag, M., Hagtvet, B., Lyster, S-A. H., Gustafsson, J-E. and Hulme, C. (2015). Sentence repetition is a measure of children's language skills rather than working memory limitations. *Developmental Science*, **18**, 146-154

Knowles, W. and Masidlover, M. (1982). The Derbyshire Language Scheme. Derbyshire County Council.

Kouri, T. A. (2005). Lexical training through modeling and elicitation procedures with late talkers who have specific language impairment and developmental delays. *Journal of Speech, Language and Hearing Research*, **48** (1), 157-171.

Leclercq, A-L., Maillart, C. and Majerus, S. (2013). Non-word repetition problems in children with specific language impairment: a deficit in accessing long-term linguistic representations? *Topics in Language Disorders*, **33** (3), 238-254.

Loftus, E. F. and Loftus, G. R. (1980). On the permanence of stored information in the human brain. *American Psychologist*, **35** (5), 409-420.

Lous, J., Burton, M. J., Felding, J. U., Ovesen, T., Rovers, M. M. and Williamson, I. (2005). Grommets (ventilations tubes) for hearing loss associated with otitis media with effusion in children. The Cochrane Database of Systematic Reviews, **1**, CD001801

Luce, P. A. and Pisoni, D. B. (1998). Recognising spoken words: the neighbourhood activation model. *Ear and Hearing*, **19** (1), 1-36.

Majerus, S. and van der Linden, M. (2003). Long-term memory effects on verbal short-term memory: a replication study. *British Journal of Developmental Psychology*, **21** (2), 303-310.

McCool, S. and Stevens, S.I. (2011). Identifying speech, language and communication needs among children and young people in residential care. *International Journal of Language and Communication Disorders*, **46** (6), 665-674.

McGregor, K. K. and Leonard, L. B. (1994). Subject pronoun and article omissions in the speech of children with specific language impairment: a phonological interpretation. *Journal of Speech and Hearing Research*, **37**, 171-181.

McKean, C. (2009). Investigating the development of a developmental disorder: Mapping the trajectory of lexical development in Specific Language Impairment. Unpublished PhD thesis, Newcastle University.

Marshall, C. R., Harris, J. and van der Lely, H. K. J. (2003). The nature of phonological representations in children with Grammatical-Specific Language Impairment (G-SLI). In D. Hall, T. Markopoulos, A. Salamoura and S. Skoufaki (eds.) *Proceedings of the University of Cambridge First Postgraduate Conference in Language Research*, **1,** 511-517.

Marshall, C. R. and van der Lely, H. K. J. (2009). Effects of word position and stress on onset cluster production: evidence from typical development, specific language impairment and dyslexia. *Language*, **85** (1), 39-57.

Martin, R. C. (1993). Short-term memory and sentence processing: Evidence from neuropsychology. *Memory and Cognition*, **21**, 76–183.

Martin, K. I. and Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, **34** (3), 379-413.

Martin, R.C., Lesch, M.F., and Bartha, M.C. (1999). Independence of input and output phonology in word processing and short-term memory. *Journal of Memory and Language*, ***41***, 3–29.

Marton, K. and Schwartz, R. G. (2003). Working memory capacity and language processes in children with specific language impairment. *Journal of Speech, Language and Hearing Research,* **46**, 1138-1153.

Maw, R., Wilks, J., Harvey, I., Peters, T. J. and Golding, J. (1999). Early surgery compared to watchful waiting for glue ear and effect on language development in preschool children: a randomized control trial. *Lancet*, **353** (9157), 960-963.

Melby-Lervåg, M. and Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, **49** (2), 270-291.

Melby-Lervåg, M., Lervåg, A., Lyster, S-A. H., Klem, M., Hagtvet B. and Hulme, C. (2012). Nonword-repetition does not appear to be a causal influence on children's vocabulary development, *Psychological Science*, **23** (10), 1092-1098.

Metsala, J. L. and Chisholm, G. M. (2010). The influence of lexical status and neighbourhood density on children's non-word repetition. *Applied Psycholinguistics*, **31**, 489-506.

Miller, G. A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review*, **63**, 81-97.

Miyawaki, K., Liberman, A. M., Strange, W., Verbrugge, R., Jenkins, J. J. and Fujimura, O. (1975). Effect of linguistic experience –discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics*, **18** (5), 331-340.

Moll, K. Hulme, C., Nag, S. and Snowling, M. (2015). Sentence repetition as a marker of language skills in children with dyslexia. *Applied Psycholinguistics*, 36, 203-22.

Munro, N., Lee, K. and 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.

Munson, B., Edwards, J. and Beckman, M. E. (2005). Relationships between non-word repetition accuracy and other measures of linguistic development in children with phonological disorders. *Journal of Speech, Language and Hearing Research*, **48**, 61-78.

Newcomer, P. L. and Hamill, D. D. (2008). Test of Language Development –primary (4th Edition). Austin, Tx: Pro-Ed.

Noble, K.G., McCandliss, B.D., & Farah, M.J. (2007). Socioeconomic gradients predict individual differences in neurocognitive abilities. *Developmental Science*, **10**, 464-480.

Papagno, R. R. (2010). Understanding Statistics in the Behavioural Sciences. CA, USA: Wandsworth Censage Learning

Peperkamp, S., Vendelin, I. and Dupoux, E. (2010). Perception of predictable stress: A cross-linguistic investigation. *Journal of Phonetics*, **38** (3), 422-430.

Pickering, M. J. and Ferreira, V. S., (2008), Structural priming: a critical review. *Psychological Bulletin*, **134** (3), 427–459.

Poirier, M. and Saint-Aubin, J. (1995). Memory for related and unrelated words – further evidence on the influence of semantic factors in immediate serial recall. *Quarterly Journal of Experimental Psychology, Section A –Human Experimental Psychology*, **48** (2), 384-404.

Polišenská, K., (2011). The Influence of Linguistic Structure on Memory Span: Repetition Tasks as a Measure of Language Ability. (Unpublished Doctoral thesis, City University London)

Potter, M.C., and Lombardi, L. (1990). Regeneration in the short-term recall of sentences. *Journal of Memory and Language*, **29**, 633-654.

Potter, M.C., and Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, ***38***, 265–282.

Ramus, F. and Szenkovits, G. (2008). What phonological deficit? *Quarterly Journal of Experimental Psychology*, **61** (1), 129-141.

Reilly, S., Bishop, D.V.M., and Tomblin, B. (2014). Terminological debate over language impairment in children: forward movement and sticking points. *International Journal of Language and Communication Disorders*, **49** (4) 452-462.

Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F.K., Morgan, A., Goldfeld, S., Nicholson, J.M. and Wake, M. (2014). Specific Language Impairment: a convenient label for whom? *International Journal of Language and Communication Disorders*, **49** (4) 416-451.

Renfrew, C. (1995). Renfrew Word Finding Test. UK: Speechmark.

Riches, N. G. (2012). Sentence repetition in children with specific language impairment: an investigation of underlying mechanisms. *International Journal of Language and Communication Disorders*, **47** (5), 499-510.

Riches, N. G., Loucas, T., Baird, G., Charman, T. and Simonoff, E. (2010). Sentence repetition in adolescents with specific language impairments and autism: an investigation of complex syntax. *International Journal of Language and Communication Disorders*, **45** (1), 47-60.

Riches, N. G., Tomasello, M. and Conti-Ramsden, G. (2005). Verb learning in children with SLI: frequency and spacing effects. *Journal of Speech, Language and Hearing Research*, **48** (6), 1397-1411.

Robinson, R. J. (1991). Causes and associations of severe and persistent specific language disorders in children. *Developmental Medicine and Child Neurology*, **33** (11), 943-962.

Rock, I., (1983). The logic of perception. Cambridge, Mass.: MIT Press

Roodenrys, S. and Hinton, M. (2002). Sublexical or lexical effects on serial recall of non-words? *Journal of Experimental Psychology –Learning Memory and Cognition*, **28** (1), 29-33.

Roodenrys, S., Hulme, C., Alban, J., Ellis, A. W., and Brown, G. D. A. (1994). Effects of word frequency and age of acquisition on short-term memory span. *Memory and Cognition*, **22** (6), 695-701.

Roy, P. and Chiat, S. (2004). A prosodically controlled word and non-word repetition task for 2- to 4-year olds: evidence from typically developing children. *Journal of Speech, Language and Hearing Research*, **47**, 223-234.

Roy, P. and Chiat, S. (2013). Teasing apart disadvantage from disorder: The case of poor language. In Marshall, C. (Ed.). Current Issues in Developmental Disorders. London: Psychology Press.

Rutter, M., Thorpe, K., Greenwood, R., Northstone, K. and Golding, J. (2003). Twins as a natural experiment to study the causes of mild language delay: I: Design; twin-singleton differences in language, and obstetric risks. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, **44** (3), 326-341.

Rvachew, S., Ohberg, A., Grawburg, M. and Heyding, J. (2003). Phonological awareness and phonemic perception in 4-year-old children with delayed expressive phonology skills. *American Journal of Speech-Language Pathology*, **12** (4), 463-471.

Saffran, E. M., and Martin, O. S. M. (1975). Immediate memory for lists of words and sentences in a patient with deficient auditory short-term memory. *Brain and Language*, **2**, 420–433.

Sahlen, B., Wagner, C. R., Nettelbladt, U. and Radeborg, K. (1999). Language comprehension and non-word repetition in children with language impairment. *Clinical Linguistics and Phonetics*, **13** (5), 369-380.

Saint-Aubin, J. and Poirier, M. (1999). Semantic similarity and immediate serial recall: is there a detrimental effect on order information? *Quarterly Journal of Experimental Psychology, Section A –Human Experimental Psychology*, **52** (2), 367-394.

Salamé, P. and Baddeley, A. D. (1982). Disruption of memory by unattended speech: implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior,* **21**, 150-164.

Schmiedek, F., Lövdén, M., and Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience*, **2**, 1–10.

Schwanenflugel, P.J., Akin, C., and Luh, W. (1991). Context availability and the recall of abstract and concrete words. *Memory and Cognition*, 20, 96–104.

Seeff-Gabriel, B., Chiat, S. and Dodd, B. (2010). Sentence imitation as a tool in identifying expressive morphosyntactic difficulties in children with severe speech difficulties. *International Journal of Language and Communication Disorders*, **45** (6), 691-702.

Seeff-Gabriel, B., Chiat, S. and Roy, P. (2008). Early Repetition Battery. London: Pearson

Shipstead, Z., Redick, T. S., Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, **138** (4), 628-654.

Snowling, M., Chiat, C. and Hulme, C. (1991). Words, non-words and phonological processes: some comments on Gathercole, Baddeley, Willis and Emslie, *Applied Psycholinguistics*, **12**, 369-373.

Spencer, S., Clegg, J. and Stackhouse, J. (2012). Language and disadvantage: a comparison of the language abilities from two different socio-economic areas. *International Journal of Language and Communication Disorders*, **47** (3), 274-284.

Stackhouse, J. and Wells, B. (1997). Children's speech and literacy difficulties: a psycholinguistic framework. London: Whurr

Stokes, S. F., Wong, A. M-Y., Fletcher, P. and Leonard, L. B. (2006). Non-word repetition as clinical markers of specific language impairment: the case of Cantonese. *Journal of Speech, Language and Hearing Research*, **49**, 219-236.

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

Stuart, G. and Hulme, C. (2000). The effects of word co-occurrence on short-term memory: Associative links in the long-term memory affect short term memory performance. *Journal of Experimental Psychology -Learning, Memory and Cognition*, **26** (3), 796-802.

Sundström, S., Samuelsson, C. and Lyxell, B. (2014). Repetition of words and non-words in typically developing children: The role of prosody. *First Language*, **34**, 428-449.

Thomson, J. M., Richardson, U. and Goswami, U. (2005). Phonological similarity neighbourhoods and children's short-term memory: typical development and dyslexia. *Memory and Cognition*, **33** (7), 1210-1219.

Thordardottir, E. and Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, **46** (1), 1-16.

Thordardottir, E., Kehayia, E., Mazer, B., Lessard, N., Majnemer, A., Sutton, A., Trudeau, N. and Chilingaryan, G. (2011). Sensitivity and Specificity of French language and processing measures for the identification of primary language impairment at age 5. *Journal of Speech, Language and Hearing Research*, **54**, 580-597.

Thorpe, K., Rutter, M. and Greenwood, R. (2003). Twins as a natural experiment to study the causes of language delay: II: Family interaction risk factors. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, **44** (3), 342-355.

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X. Y., Smith, E., O'Brien, M. (1997). Prevalence of Specific Language Impairment in kindergarten children. *Journal of Speech, Language and Hearing Research,* **40** (6), 1245-1260.

Turner, J. E., Henry, L. A. and Smith, P. T. (2000). The development of the use of long-term knowledge to assist short-term recall. *Quarterly Journal of Experimental Psychology,* ***53A***, 457-478.

Turner, E., Henry, L.A., Smith, P.T. and Brown, P. (2004). Redintegration and lexicality effects in children: Do they depend upon the demands of the task? *Memory and Cognition,* **32**, 501-510.

van der Lely, H. K. J. (1993). Specific language impairment in children –research findings and their therapeutic implications. *European Journal of Disorders of Communication*, **28** (3), 247-261.

van der Lely, H. K. J., Gardner, H., Froud, K. and McClelland, A. (2007). Grammar and Phonology Screening (GAPS) test. St Mabyn: STASS publications

van der Lely, H. K. J. and Howard, D. (1993). Children with specific language impairment –linguistic impairment or short-term memory deficit? *Journal of Speech and Hearing Research*, **36** (6), 1193-1207.

Vance, M. and Clegg, J. (2012). Use of single case study research in child speech, language and communication interventions. *Child Language Teaching and Therapy*, **28** (3), 255-258.

Vance, M., Rosen, S. and Coleman, M. (2009). Assessing speech perception in young children and relationships with language skills. *International Journal of Audiology*, **48**, 708-717.

Vance, M., Stackhouse, J., Wells, B. (2005). Speech-production skills in children aged 3-7 years. *International Journal of Language and Communication Disorders*, **40** (1), 29-48.

Velleman, S. L. and Vihman, M. (2002). Whole-word phonology and templates: trap, bootstrap, or some of each. *Language, Speech and Hearing Services in Schools*, **33**, 9-23.

Vitevitch, M. S. and Luce, P. A. (1999). Probabilistic phonotactics and neighbourhood activation in spoken word recognition. *Journal of Memory and Language*, **40** (3), 374-408.

Walker, I. and Hulme, C. (1999). Concrete words are easier to recall than abstract words: evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology – Learning, Memory and Cognition*, **25** (5), 1256-1271.

Warren, R. M. (1970). Perceptual restorations of missing speech sounds. *Science*, **167**, 392-393.

Wiig, E. H., Secord, W. A., and Semel, E. (2004). Clinical evaluation of language fundamentals—Preschool, second edition (CELF Preschool-2). Toronto, Canada: The Psychological Corporation/A Harcourt Assessment Company.

Williams, N. and Chiat, S. (1993). Processing deficits in children with phonological disorder and delay –a comparison of responses to a series of output tasks. *Clinical Linguistics and Phonetics,* **7** (2), 145-160.

Willis, C. S. and Gathercole, S. E. (2001). Phonological short-term memory contributions to sentence processing in young children. *Memory*, **9** (4/5/6), 349-363.

Zimmerman, I. L., Steiner, V. G. and Pond, R. E. (2002). Pre-school Language Scales -4. London: Pearson.

Ziegler, J. C., Pech-Georgel, C., George, F., Alario, F. X. and Lorenzi, C. (2005). Deficits in speech perception predict language learning impairment. *Proceedings of the National Academy of Sciences of the United States of America*, **102** (39), 14110-14115.