Speech processing in typical and atypical language development: using nonwords to map the way

Susan-Sophia Gates

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
UCL

Department of Speech, Hearing and Phonetic Sciences

2010

Declaration

I, Susan-Sophia Gates, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgements

First and foremost I would like to thank my supervisors, Stuart Rosen and Maggie Vance for their support and encouragement, particularly in the final months of the project. Stuart has been an amazing source of technical and statistical information and Maggie has guided me through the minefield of clinical research. Both have provided valuable theoretical advice and given very welcome feedback. Many thanks to Stuart, too, for supporting my studentship application, without which my PhD would not have been possible.

There are many people in the Department of Speech, Hearing and Phonetic Sciences (SHaPS) who have contributed to this project and whom I would like to thank. Mark Huckvale provided valuable assistance with the speech morphing software and procedure, and willingly answered my many questions. Mike Coleman shared his SpaRedux and SipSLI software and updated both to meet the requirements of my experimental design. Gaston Hilkhuysen provided an essential, emergency tutorial in multiple regression, for which I am extremely grateful. Bronwen Evans and Katrin Skoruppa also helped in a statistical emergency.

SHaPS is supported by a wonderful technical team. I will be ever grateful to Dave Cushing, Steve Nevard, Steve Newton and Warwick Smith for helping me set up and calibrate my equipment and for coming to my rescue during some unpleasant IT disasters!

I would also like to thank the many people with whom I have discussed my work. To name but two, Chloe Marshall explained the intricacies of her nonword stimuli so that I could generate my own and Souhila Messaoud-Galusi shared her experiences of data collection and analysis on numerous occasions.

Data collection would not have been possible without the help of the staff and pupils at the following mainstream lower schools in Bedfordshire: Balliol, Broadmead, Church End and Slisoe; and schools and nurseries in London: Invicta School, UCL Day Nursery and Institute of Education Nursery. I would like to express particular gratitude to head teachers Brian Storey

(Church End Lower School) and Susan Purdue (Silsoe Lower School) for their personal support of my project and their commitment to random recruitment of large numbers of control participants. I am also extremely grateful to the staff and pupils at the following schools with language provisions: Applegarth Junior School, New Addington; Coleraine Park School, Tottenham; Green Street Green Primary School, Orpington; the Rainbow Centre at Egerton Rothesay School, Berkhamsted; and one of the ICan provisions: Meath School, Ottershaw.

I should not forget the people who brought my frogs to life: Vicki, Sam and the others whose recordings were essential but did not make it to the experimental tasks.

At a personal level, my biggest and most heartfelt ‘thank you’ goes to my long-suffering husband Michael, who has ridden the PhD roller-coaster by my side. Michael has been unwavering in his support, both emotional and technical, and has never ceased to encourage me despite my frequent grumpiness. I have appreciated his endless patience and ability to remain calm in a crisis.

My dear parents, too, have provided unconditional emotional support, not just over the course of my PhD, but throughout my life. They always value my achievements, whether large or small, and from a young age they encouraged me to be inquisitive and creative and to pursue my own ambitions. I thank them for the love they have given, the examples they have set and the lessons they have taught.

Similarly, I would like to express my thanks to my dear grandparents, who have been so generous with their love and attention. I am especially grateful to Grandpy for teaching me to appreciate the academic opportunities that were unavailable to him. His never ceasing zeal for learning and self-development is truly inspiring. All my grandparents are, or would have been, extremely proud that I have completed a PhD. Grandad, in particular, would have been overwhelmed that there could be a Dr. Gates in the family!

Finally, thank you to all the friends that have listened to my grumbles, sympathised with my dilemmas and assisted with my problems. Special thanks to Katharine Mair and Kayoko Yanagisawa for participating in my dramas and talking me out of my panics. Thanks, similarly, to everyone in 326, especially Chierh Cheng, Kota Hattori, Young Shin Kim, Jeong Sug Kyong, Angelos Lengeris, Melanie Pinet and Mark Wibrow. I would also like to thank Jan Eaton for offering much needed support before my Upgrade. I will remember her wise words.

This research was funded by a joint studentship from the Medical Research Council and the Economic and Social Research Council. I would like to thank these councils for the financial support which made this project possible.

Abstract

Accurate differential diagnosis of specific language impairment (SLI) is essential to determine the optimum form and content of treatment. It is therefore important to address the cognitive processes underpinning SLI and to evaluate potential clinical markers.

The experiments presented here were designed to investigate input and output phonological processing in typically developing (TD) children, children with SLI and children with SLI and a concomitant speech disorder (SLI+SSD). Participants carried out a battery of published assessments and three experimental tasks: nonword repetition (NWR), nonword discrimination (NWD) and categorical perception (CP). Each experimental task used the same nonword stimuli which had been created by manipulating the position and number of consonant clusters and reflected repetition errors previously observed in children with SLI and/or dyslexia (Marshall & van der Lely, 2009).

The results showed that NWD and NWR were highly correlated in TD children, implicating the same phonological processes in performance accuracy. Furthermore, none of the tasks was related to published linguistic assessments, including measures of receptive vocabulary and phonological short-term memory. It seemed that the experimental tasks tapped phonological representations at an unspecified stage of sub-lexical speech processing.

Subtle processing differences were found between the clinical groups. Children with SLI showed deficits in NWR and NWD but not CP. Children with SLI+SSD were as impaired as SLIs on NWD, but they showed a NWR deficit of greater magnitude and were additionally impaired on CP. It was proposed that the SLI deficit was related to the length and complexity of the nonword stimuli and that a bidirectional transfer of information between sub-lexical output and input phonological representations may explain SLI+SSD performance. Sentence recall was the most reliable marker of SLI. The clinical implications of the results were discussed and it was concluded that caution should be exercised when administering or interpreting NWR tasks.

Contents

Tables

Figures

1. Literature Review

It is estimated that 6-8% of children in the UK aged between 0 – 11 years have speech, language and communication needs (R.C.S.L.T., 2006). Such needs undoubtedly affect a child's personal and social development as well as his/her access to the curriculum. Childhood speech and language difficulties often persist into adolescence (Conti-Ramsden & Durkin, 2007; Conti-Ramsden, Durkin, Simkin, & Knox, 2009) and adulthood, when they are associated with social adaptation difficulties and increased risk of psychiatric disorder (Clegg, Hollis, Mawhood, & Rutter, 2005). Many children with developmental language disorders develop mental health problems (Clegg, Hollis, & Rutter, 1999) and social and behavioural difficulties (Conti-Ramsden & Botting, 2004). High levels of speech, language and communication difficulties have been found among the young offender population (Bryan, 2004). There are clear benefits, therefore, in developing increasingly effective methods of diagnosis and treatment of such difficulties. To this end, we need to learn more about the processes that underlie typical and atypical speech and language development, and to develop reliable markers for communication disorders.

1.1 Specific Language Impairment (SLI)

Specific Language Impairment (SLI) is a developmental disorder that results in a functional communication impairment (Enderby, et al., 2009). The prevalence of SLI has been estimated by several research groups whose findings differ depending on, for example, the criteria used to define a language impairment, the age of the population tested, and the sampling methods employed. The SLI literature often cites the American study carried out by Tomblin and colleagues which found that 7.4% of kindergarten children met the criteria for SLI (Tomblin, et al., 1997). Tomblin et al found the prevalence estimates for boys and girls to be 8% and 6% respectively. SLI is diagnosed - largely by exclusion - in children who have significant deficits in receptive and/or expressive language skills which are not directly attributable to neurological or anatomical abnormalities, sensory impairments, cognitive deficits, or environmental factors (Leonard, 1998). SLI is a heterogeneous disorder which presents with a wide range of individual profiles that change over developmental time under the influence of external factors such as intervention (Conti-Ramsden & Botting, 1999). A summary of the key deficits in children with SLI is presented below.

1.1.1 Linguistic deficits

The language of children with SLI is characterised by impairments in lexical, grammatical and morphological development (for review, see Leonard, 1998). Experiments have also shown that children with SLI have difficulty in acquiring the phonological forms of new words (Dollaghan, 1987; Gray, 2004; Weismer & Hesketh, 1996).

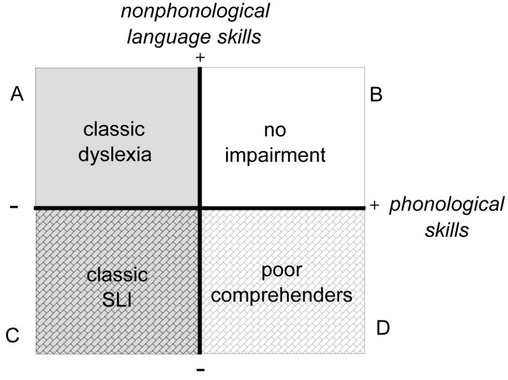
Although children with SLI seem to be late in acquiring their first words, their early lexical development matches that of younger typically developing (TD) children in terms of the types of words used. However once language impaired children begin joining words into sentences, their lexical abilities fall behind those of their peers and begin to deviate from the TD pattern. Verb production is especially impaired throughout the preschool years (for review, see Leonard, 1998). Longitudinal studies of ‘late talkers’ have shown that deficits in expressive vocabulary size at two years of age are not strongly predictive of a language deficit at six years of age (Girolametto, Wiigs, Smyth, Weitzman, & Pearce, 2001). Indeed, preschoolers with SLI are more likely to improve deficits in vocabulary size than problems with morphosyntax (Rice, Wexler, & Hershberger, 1998).

The lexical deficit in SLI is not restricted to vocabulary size. Language impaired children also have difficulty creating complex semantic relationships within and between words (McGregor, 1997) and are worse at incidental word learning than language- and age-matched peers (Rice, Buhr, & Nemeth, 1990). Research into the word finding difficulties (WFD) of children with SLI is relatively scarce (Messer & Dockrell, 2006) yet a population survey found that 25% of children with language impairments had WFD (Dockrell, Messer, George, & Wilson, 1998). There is evidence that WFD predict reading difficulties and poor educational performance (Wolf & Segal, 1992).

To acquire a new word, a child must identify the phonological form and correct meaning from linguistic experience, then store and organise the word’s phonological, syntactic and semantic information within the lexicon so that it can be accessed for expressive use (Brackenbury & Pye, 2005). Phonological skills are therefore crucial to word learning, lexical knowledge and expressive language. It is not particularly surprising to find that children impaired in their ability to acquire certain aspects of language, potentially have problems with phonology (Leonard, 1982).

A phonological awareness deficit has been found in some but not all language impaired children and this finding has recently been related to the presence or absence of a co-morbid literacy impairment. Bishop and Snowling (2004) proposed a two-dimensional model whereby SLI and dyslexia may occupy different sectors of a quadrant in which phonological skills are represented on the x-axis and non-phonological language skills are represented on the y-axis. Bishop and Snowling (2004) labelled children in the bottom right quadrant as ‘poor comprehenders’ as they have intact phonological skills but subtle language difficulties that may remain after an earlier, more significant impairment has resolved (Nation, Clarke, Marshall, & Durand, 2004).

Figure 1.1: Bishop & Snowling’s model of the relationship between SLI and dyslexia.



Source: (Bishop & Snowling, 2004)

Children in Bishop and Snowling’s (2004) bottom left quadrant have poor phonological and poor language skills and are labelled ‘classic SLI’. A child with SLI who has good reading skills may present with a profile that resembles that of a poor comprehender, with phonological skills that are not significantly impaired. A study by Catts et al. (2005) found that reading unimpaired children with SLI showed only mild deficits in phonological awareness (phoneme deletion) and nonword repetition (NWR) when compared with typically developing (TD) controls. School-aged children with dyslexia or dyslexia + SLI performed worse than those with single deficit SLI. The authors concluded that a phonological impairment does not seem to be a major factor in SLI when it occurs in the absence of dyslexia. Bishop, McDonald, Bird and Hayiou-Thomas (2009) replicated this finding, reporting that 9 year old children with SLI performed within normal limits on two phonological processing tasks, namely NWR and rapid serial naming (RSN). However Bishop et al. (2009) did find that the same SLI children had small but significant deficits on other measures that challenged phonological output (oromotor skills) and phonological memory (memory for names). The memory and nonword repetition skills of children with SLI will be described in more detail below.

A grammatical deficit in SLI has been well-documented. Children with SLI have difficulty producing and/or comprehending specific types of syntactic relationships, for example reversible passives, bound pronouns and reflexives (van der Lely, 1996; van der Lely & Stollwerck, 1997). These children tend to communicate in utterances that are shorter and less complex than those of TD peers (van der Lely, Rosen, & McClelland, 1998). The main focus of research interest in this area has been on the SLI impairment in grammatical morphology. Children with SLI are both delayed and less consistent in their use of grammatical morphemes that mark tense and agreement (such as third-person singular *–s* and past tense *–ed*), using fewer of these morphemes in obligatory contexts than younger, TD children with similar mean length of utterance (Rice & Wexler, 1996). The deficit in past tense morphology is so marked that it has been proposed as a differentiating clinical marker for SLI (Rice & Wexler, 1996).

The underlying cause of the grammatical deficit is still debated. Some researchers believe that SLI reflects anomalies in the neurobiological representation of grammar and that co-occurring non-linguistic impairments involve other causes or are a consequence of the primary linguistic deficit itself (van der Lely, et al., 1998). An alternative view holds that impaired speech processing is at the heart of the grammatical deficit, such that children with SLI are unable to adequately extract morphological markers from the speech signal (Joanisse & Seidenberg, 1998). The latter theory is intuitively appealing when one considers that there is typically a high degree of phonological similarity between a present and a past tense form (e.g. walk – walked). Furthermore, the regular past tense suffix is phonologically realised in one of three ways, depending on the verb stem: /-t/ (e.g. walked), /-d/ (e.g. sailed), or /-ɪd/ (e.g. waited). The phonological structure of the inflected and uninflected form must be analysed in order to acquire an adequate generalisation to a novel item. This type of analysis is also necessary in subregular forms (e.g. ring – rang) and across regular and irregular forms (e.g. sail*ed*, hel*d*) (Archibald, Joanisse, & Shepherd, 2008).

1.1.2 Speech perception deficits

As stated above, there are some who believe the primary deficit in SLI to be impaired speech processing. This theory gained momentum in the 1970s with the work of Tallal and colleagues who reported that children with SLI (whom the authors referred to as *developmental aphasics*), had difficulty perceiving a sequence of short non-speech tones in a temporal order judgement (TOJ) task (Tallal & Piercy, 1973). In a TOJ task, the listener is presented with pairs of auditory stimuli in one of four possible sequences (1-1, 1-2, 2-1, 2-2) and is required to identify and order the constituent stimuli. Stimulus pairs typically vary according to duration of the stimuli and the interstimulus interval (ISI). Tallal and Piercy's language impaired participants had a specific difficulty when the ISIs were short. This non-speech data was extrapolated to speech sounds and led to the rapid temporal processing theory (RTP) of SLI, which holds that children with SLI have difficulty perceiving rapidly changing acoustic properties of any nature, including those inherent to speech (such as formant transitions, Tallal & Piercy, 1973). Difficulties in rapid auditory processing have been found in some later studies of SLI (e.g. Frumkin & Rapin, 1980) but they have not been found in others (e.g. Norrelgen, Lacerda, & Forssberg, 2002).

It has been suggested that where a child with SLI shows an auditory processing deficit, the deficit is not characterised by poor temporal resolution of perceptual systems and may not be applicable to speech sounds (for a review, see Rosen, 2003). Rather, difficulties may arise from auditory maturity (Bishop, Adams, Nation, & Rosen, 2005) or as an artefact of task demands (Coady, Kluender, & Evans, 2005). Alternatively the heterogeneous nature of the SLI population may account for the conflicting evidence (McArthur & Bishop, 2001).

Many perception studies, like that of Tallal and Piercy (1973) report the performance of children with SLI on tasks involving non-speech sounds. Others have used speech stimuli to investigate the perceptual skills of children with SLI.

Categorical perception is a process of perceptual organisation. In speech, it refers to the organisation of perceived speech sounds into categories despite variation between tokens at the acoustic level. Categorical perception allows the listener to ‘ignore’ inter- and intra-speaker variation to extract the linguistic content of an utterance. The acquisition of this skill is a protracted and complex process for the developing child. Infants are born with excellent language-independent discrimination skills (Jusczyk & Luce, 2002). The ability to categorise speech sounds is developed throughout childhood and into the teenage years with children performing less consistently than adults in categorical perception tasks (Hazan & Barrett, 2000). Children with SLI have difficulty with categorical perception tasks, although task demands have been implicated in these findings, including the added processing demands of synthetic stimuli (Coady, et al., 2005). There is evidence that the categorical perception of children with SLI is impaired for synthesised speech stimuli (e.g. Thibodeau & Sussman, 1979), but not natural speech stimuli (Coady, et al., 2005).

1.1.3 Memory deficits

While evidence is emerging that memory deficits in SLI also involve the procedural and declarative memory systems (Lum, Gelgic, & Conti-Ramsden, 2010; Tomblin, Mainela-Arnold, & Zhang, 2007), there is a substantial body of literature describing phonological short-term memory (PSTM) and working memory (WM) deficits in children with SLI (Archibald & Gathercole, 2006b; Gathercole & Baddeley, 1990; Montgomery, 2003). It has been suggested that there is a causal relationship between phonological WM impairments and disordered language development (Gathercole & Baddeley, 1990), although more recent research suggests an additive rather than a causal connection (as deficits in language and WM do not always co-occur) (Archibald & Joanisse, 2009).

Research in this field has been dominated by the WM model first described by Baddeley and Hitch (1974). According to this domain-general model, WM is a three-component system comprised of a controlling ‘‘central executive’’ and two ‘slave’, modality-specific components: the phonological loop and the visuo-spatial sketchpad. The central executive is thought to organise and control communication within WM, and to regulate and control the retrieval of information from other memory systems as well as the processing and storage of information. The phonological loop includes a capacity-limited phonological short-term store and an articulatory control function which stores and refreshes verbal material, allowing it to be retained for longer in the loop. The phonological loop’s function is to store verbal input temporarily, while other cognitive tasks such as auditory comprehension take place. This applies to novel phonological input and allows the listener the opportunity to create long-term phonological representations of that material – a necessary process in learning new words (Baddeley, Gathercole, & Papagno, 1998). The visuo-spatial sketchpad, which will not be described in detail here, acts as a capacity-limited, short-term store for visual information.

The initial tripartite model was later amended by Baddeley (2000) to include a component called the ‘episodic buffer’. This component is assumed to be a capacity-limited, temporary storage system that integrates multimodal information by providing a temporary interface between the slave systems (the phonological loop and the visuospatial sketchpad) and long-term memory (LTM). The buffer is controlled by the central executive, which is capable of retrieving information from the store, reflecting on it and even manipulating and modifying it. In this way, it is proposed that the episodic buffer serves as a ‘modelling space’ (Baddeley, 2000, p. 421) that is important in long-term episodic learning.

An important distinction must be made between WM and PSTM. Both systems involve temporary storage but they may be distinguished by whether or not significant processing activity is required concurrently (Archibald & Gathercole, 2006b). PSTM tasks, such as digit or word recall, impose storage but minimal processing demands whereas WM tasks require significant information processing in addition to storage. WM tasks typically use complex memory span paradigms e.g. where listeners are required to make a meaning- or truth-based judgement about a sentence and also recall the last word of the sentence.

Gathercole and Baddeley (1990) first proposed that SLI may involve a specific deficit to the phonological loop component of working memory. They reported that children with SLI are poor at repeating polysyllabic nonwords[1] and recalling word lists. Their research sparked debate about the role of PSTM in SLI and the underlying skills tapped by nonword repetition (the latter of which will be discussed below). Gathercole and Baddeley could not explain their results in terms of speech output as the articulation rates for the SLI children and the language matched controls were very similar while the control children performed consistently better on the memory tasks. Moreover, restricted recall was found in the SLI group even when the children were required to point to a sequence of pictures and not make a verbal response. Nor could the authors explain their results in terms of an auditory perceptual deficit in the SLI children. They carried out word and nonword pair discrimination tasks and found SLI children performed worse but were not significantly impaired on discriminating nonword pairs ‘of the kind of speech stimuli’ (p. 353) used in the memory tasks. However, the nonword stimuli were not phonologically matched across the tasks, and the SLI group was very small (6 children). Gathercole and Baddeley concluded that a primary deficit in PSTM could be the basis of the language deficits seen in SLI. PSTM deficits, as measured by serial recall, have been found in subsequent studies (Briscoe & Rankin, 2009) and have prompted some to argue that grammatical deficits observed in SLI may be caused by a failure of PSTM to maintain adequate phonological representations of sentences while comprehending them (Joanisse & Seidenberg, 2003).

Finally, deficits in visuo-spatial short-term memory have been found in preschool (Hick, Botting, & Conti-Ramsden, 2005) and school-aged (Cowan, Donlan, Newton, & Lloyd, 2005) children with SLI. This evidence suggests a more general short-term memory deficit in SLI with performance potentially affected by a number of underlying factors such as use and efficacy of

---

[1] Gathercole and colleagues consider nonword repetition to index PSTM however this is a position that is not universally accepted. Nonword repetition and the skills it measures will be considered further below.

memory strategies, control of attention, speed of information processing, or rate of information decay (Cowan, et al., 2005).

1.1.4 A specific language impairment?

SLI is usually diagnosed in children who have significant deficits in expressive and/or receptive language which are not directly attributable to neurological or anatomical abnormalities, sensory impairments, cognitive deficits, or environmental factors (Leonard, 1998). However as the discussion on memory above shows, SLI often co-occurs with non-linguistic deficits and the specificity of the disorder is increasingly under question.

There is evidence of motor deficits in children with language impairments (for an extensive reveiw and a meta-anlaysis, see Hill, 2001; Rechetnikov & Maitra, 2009). These deficits are not confined to orofacial movements but have been found in fine and gross motor function as well as complex sequential and learned movements. Specifically, the evidence suggests that children with SLI have particular difficulty matching the performance of typically developing (TD) controls on motor tasks involving speeded, sequential and/or timed movements (e.g. Bishop, 2002). Moreover, in the studies reviewed by Hill (2001), between 40% and 90% of children with SLI had movement difficulties that met the diagnostic criteria for developmental coordination disorder (DCD, also known as ‘dyspraxia’). Movement difficulties interfere significantly with activities of daily living such as dressing and eating (*Diagnostic and Statistical Manual of Mental Disorders*, 1994). Leonard et al. (2007) suggested that such difficulties may be a result of processing limitations which affect motor skills and exacerbate language difficulties.

The clinical diagnostic criteria for SLI imply that general cognitive skills develop at the typical rate. Historically and despite claims that particular domains of cognition cannot develop in isolation (Karmiloff-Smith, 1998), researchers have assumed no underlying cognitive deficit in SLI and have focussed on investigating the precise nature of the language impairment itself. However there is mounting evidence to suggest that cognitive skills may not be spared in SLI. Johnston and Smith (1989) first demonstrated that children with SLI performed worse than TD controls on a task of non-verbal reasoning, a difference that could not be explained in terms of comprehension difficulties in the SLI group. Johnston and Smith concluded that the language impaired children were also ‘thought impaired’ (1989, p.38), with processing problems ultimately leading to difficulty in practical problem solving and the construction of basic cognitive schemes. Other cognitive deficits have been reported in SLI. For example, there is evidence that language impaired children have difficulty on tasks involving mental rotation (Johnston & Weismer, 1983; Miller, Kail, Leonard, & Tomblin, 2001), visuo-spatial short term memory (Cowan, et al., 2005; Hick, et al., 2005), speeded processing (Fazio, 1999) and deductive reasoning (Newton, Roberts, & Donlan, 2010). There is also evidence to doubt the persistence of age-appropriate non-verbal IQ through development (Botting, 2005).

While the precise cause or causes of SLI are still unknown, many theories have been proposed to account for the disorder. Most models fall into one of two distinct classes: those that treat language difficulties as secondary to non-linguistic processing deficits, and those that regard the impairment as specifically linguistic (Pennington & Bishop, 2009).

Some processing deficit hypotheses claim the problems are general in nature, such as a reduced cognitive processing rate or capacity limitations on processing (Kail, 1994; Miller, et al., 2001). Other theories argue for a deficit to a specific cognitive or processing mechanism, such as PSTM (see memory deficits in SLI above).

Extensive research has established that WM is not a single store but rather a system of interacting components. Theoretical accounts propose a domain-general processing system that controls domain-specific storage systems for each modality while others propose that verbal and visual information is stored and processed in entirely independent, themselves domain-specific, systems (from Alloway, Gathercole, & Pickering, 2006). Alloway et al (2006) explored the structure of short-term and working memory in children aged 4-11 years. In support of models such as that of Baddely and Hitch (1974), their analysis confirmed that storage aspects of WM tasks depended on domain-specific verbal and visuo-spatial resources, while the processing aspects of such tasks relied on a shared, executive resource.

Still other processing deficit hypotheses have argued that the impairments in SLI can be explained by a perceptual or temporal processing deficit. Of these theories, the rapid temporal processing (RTP) theory (Tallal & Piercy, 1973) has been most widely researched and holds that a central auditory processing deficit (a slowing of perception) is at the core of SLI (see description of speech perception deficits in SLI above). Nonlinguistic temporal processing was implicated in research by Miller et al. (2001) who found that language impaired children performed more slowly than controls on a range of linguistic tasks and nonverbal, non-linguistic cognitive tasks such as mental rotation (a slowing of cognition).

An alternative explanation for SLI is that it arises from deficits in specifically linguistic systems and is not a secondary result of a more general processing deficit. This viewpoint, too, has many ‘flavours’, but all agree that an impairment to the system responsible for the rule-governed combination of words into complex structures is at the core of SLI. As noted earlier, some proponents of such a specific grammatical impairment identify particular grammatical operations that are problematic, for example marking tense (Rice & Wexler, 1996). A further account claims that wide-ranging grammatical difficulties in many children with SLI can be explained by a representational deficit of grammatical relations (van der Lely, et al., 1998).

While processing accounts identify tasks that are difficult for children with SLI, they fail to make a clear connection between the grammatical deficits observed in SLI and the proposed processing or perceptual deficits (Tomblin, et al., 2007). Furthermore, while some children with SLI undoubtedly have deficits to central perceptual and/or cognitive systems, not all language impaired children show similar deficits. It seems more likely that, rather than a causal relationship with SLI, general processing deficits instead co-occur in some, but not all, children with SLI. An alternative perspective on SLI was advanced by Ullman and Pierpont (2005) in their Procedural Deficit Hypothesis (PDH). According to this account, SLI can be explained by abnormal development of brain structures responsible for the learning and execution of motor and cognitive skills. These deficits inhibit the implicit learning of rule-based behaviours, especially where sequence information is important, for example in complex motor activity or in grammatical aspects of language learning. Ullman and Pierpont suggest that their theory is a kind of ‘third way’, as it has the potential to explain the linguistic as well as the non-linguistic deficits seen in SLI and to account for the heterogeneity of the disorder.

Twin studies have shed some light on the various theoretical accounts of SLI. Bishop and colleagues studied samples of twin children and revealed that several deficits are implicated in SLI. These authors found that children with SLI performed worse than controls on an auditory processing task and a PSTM task (nonword repetition), but that some children with typically developing language also performed poorly on one of the tasks (Bishop, et al., 1999). A genetic analysis was carried out which suggested that the auditory deficit was environmental in origin but that the nonword repetition deficit was heritable, thereby implicating different genes and different causes. A later twin study compared the performance of language impaired children to that of controls in tests of productive verb morphology and nonword repetition (Bishop, Adams, & Norbury, 2006). Again, children with SLI performed worse than controls on both tasks but some children with typically developing language scored in the impaired range on one of the measures. Both deficits were found to be heritable, suggesting a genetic aetiology, although there was no evidence that common genes were implicated.

According to Pennington and Bishop (2009, p. 291), these twin studies raised some general points that could also be applied to other developmental disorders:

(1) SLI cannot be explained by any theory that postulates a single underlying deficit. Several distinct deficits are implicated, none of which is necessary or sufficient in isolation to cause SLI.

(2) Although the different deficits can be dissociated and appear to have distinct aetiologies, they tend to co-occur at above-chance levels in children with SLI.

(3) It is possible to have a single deficit e.g. in PSTM without showing a language impairment.

(4) Children with SLI typically show more than one deficit.

In recent years, genetic research has further helped to disentangle the various theories of SLI. There is strong evidence that SLI is “familial, moderately heritable and has several replicated linkages to specific chromosome locations" (Pennington & Bishop, 2009, p. 295). The FOXP2 gene has been associated with a specific speech and language disorder in a three-generational British family (for a review, see Fisher, 2005) and the gene CNTNAP2 has been associated with nonword repetition (regarded as a measure of PSTM) in children with SLI (Vernes, et al., 2008). It should be noted, however, that the heritability of the disorder is generally significantly less than 100%, so environmental factors must play a role in its development (Pennington & Bishop, 2009).

In summary, Tomblin (2009) notes that language emerges out of multiple interactive neuro-cognitive systems that are highly responsive to the biological and experiential environment. The same systems that give rise to 'normal' variation also contribute to 'abnormal' variation. This results in traits that are continuous from exceptionally high to exceptionally low. SLI represents a region of low language ability that shares the same aetiologic factors as those that contribute to normal variation; people with SLI are distinguished from other poor language learners by a set of common features such as worse grammar than vocabulary skills. Tomblin (2009) concludes that there may be no unique underlying cause of SLI; rather it may be a constellation of common factors arising out of the number of risk factors and the nature of their interactions.

1.1.5 A single heterogeneous disorder?

While several distinct deficits are implicated in SLI, none is necessary and therefore individual profiles of SLI vary greatly in terms of the characteristics and severity of the impairment. This has inevitably led some researchers to question whether SLI can be considered to be a single heterogeneous condition. Conti-Ramsden and colleagues identified six robust sub-groups of children with language difficulties (Conti-Ramsden, Crutchley, & Botting, 1997) and found that while the language strengths and weaknesses of individual children changed over time, the six sub-groups remained stable (Conti-Ramsden & Botting, 1999). After Rapin and Allen (1987), the sub-types were identified and labelled as: (1) lexical-syntactic deficit syndrome; (2) language "normal" group (i.e. children maintaining a pattern of typically developing language skills); (3) verbal dyspraxia; (4) phonologic programming deficit syndrome; (5) phonological-syntactic deficit syndrome and (6) semantic-pragmatic deficit syndrome.

There is much evidence that many, if not most individuals with SLI exhibit one or more non-linguistic deficits in addition to their language impairments (see previous sections). In cases where non-linguistic deficits have not been found (e.g. van der Lely, 1993; van der Lely, et al., 1998), it seems plausible that not all deficits have been probed for, or that the deficits were subtle and hard to detect (Ullman & Pierpont, 2005). For example, McArthur and Bishop (2005) suggested that age may be an important factor when combined with the age-sensitivity of psycho-acoustic tasks such as frequency discrimination. It is possible that on some measures, children with SLI perform at adult levels at a relatively young age, so that underlying deficits may not be detected in older participants. Moreover, children with obvious non-linguistic deficits would likely be excluded from studies of SLI despite a language profile that might be characteristic of the disorder (Ullman & Pierpont, 2005).

Given that the individual strengths and weaknesses of children with SLI change over time (Conti-Ramsden & Botting, 1999), it seems plausible that developmental disorders such as SLI are somehow ‘fluid’. The nature and level of the deficit(s) may change in response to factors such as the maturation of other cognitive systems, environmental changes, education (especially literacy development) and therapeutic intervention. When the level of an initial deficit improves sufficiently, it may no longer be detectable or may only manifest in particularly taxing tasks or situations.

1.1.6 Clinical markers (endophenotypes) for SLI

Currently, SLI is often diagnosed in children by exclusion rather than by positively identifying particular characteristics. There is an impetus within the health and education professions to develop diagnostic markers for communication disorders to aid clinical/educational decision making. The discovery of a reliable marker for SLI could have a significant effect on the screening, differential diagnosis and remediation of communication disorders as well as informing research about their underlying mechanisms. In recent years, several research groups have suggested clinical markers for SLI. The three most salient in the literature are finite verb morphology (Rice & Wexler, 1996), sentence recall (Conti-Ramsden, Botting, & Faragher, 2001) and nonword repetition (Bishop, North, & Donlan, 1996).

Clinical markers are also potentially significant in research into the heritability of communication disorders. For this reason they should be ‘subclinical’ in nature so that they identify family members who show no overt deficit but who have overcome a genetic predisposition to the impairment. Grammar related markers, e.g. past tense formation, do not reliably identify those with resolved difficulties (Botting & Conti-Ramsden, 2003). Moreover, children with speech sound disorder (SSD) who are not diagnosed with SLI have difficulty with finite verb morphology that cannot be explained in terms of their speech deficit (Rvachew, Gaines, Cloutier, & Blanchet, 2005). If SLI and SSD are indeed distinct disorders, this raises questions about the reliability of verb morphology as a marker for SLI (Mortimer & Rvachew, 2010).

Sentence recall has also been proposed as a marker for SLI (Conti-Ramsden, et al., 2001). This test discriminates between children with SLI and typically developing peers (Conti-Ramsden, et al., 2001) as well as between young adults with and without a history of language impairment (Tomblin, Freese, & Records, 1992). Sentence recall potentially involves multiple processes including memory, however it potentially taps long-term language knowledge in addition to short-term storage capacity (Conti-Ramsden, et al., 2001). Conti-Ramsden et al. (2001) contrast the demands of sentence recall with those of nonword repetition (NWR), which they state probably involves phonological short term memory (PSTM) to a greater extent than does sentence recall. These authors also suggest that NWR taps single word processing while sentence recall taps the underlying ‘language knowledge base’ (p.747). This proposition was supported by a stronger correlation between sentence recall and linguistic tense tasks than between the linguistic tasks and NWR (Conti-Ramsden, et al., 2001).

A NWR deficit is present in children with SLI across the full range of childhood years, ranging from the preschool period (Gray, 2003) through to adolescence; it is even evident in older children whose language impairment has apparently resolved (Bishop, et al., 1996; Stothard, Snowling, Bishop, Chipchase, & Kaplan, 1998). According to some researchers, the consistency and size of the NWR deficit in SLI make it a good candidate for a behavioural marker for the disorder (Gathercole, 2006). Other groups, however, have only found a weak association between SLI and NWR accuracy and have suggested that deficits in NWR are more closely linked to dyslexia (Catts, et al., 2005).

Several studies have evaluated the reliability of proposed clinical markers. Archibald and Joanisse (2009) assessed the sensitivity and specificity of both NWR and sentence recall. These authors concluded that the sentence recall, but not NWR, usefully pinpointed deficits in language and working memory in children with SLI. Conti-Ramsden et al. (2001) evaluated these markers and two verb morphology tasks (past tense and third person singular). The grammatical tasks proved the least reliable markers for SLI, while sentence recall was the most reliable.

1.2 Speech sound disorder (SSD)

Learning to articulate speech sounds is a long process which begins in infancy and is typically complete by the time a child is nine years of age (Smit, Hand, Freilinger, Bernthal, & Bird, 1990)[2]. In the absence of congenital or syndromic conditions, some children experience significant difficulty acquiring articulate speech sounds and such children are described as having a speech sound disorder (SSD; Enderby, et al., 2009). Children with SSDs produce speech that is characterised by a greater number of misarticulations and lower intelligibility than are expected for their age (Mortimer & Rvachew, 2008). The most common errors in children with SSD are substitutions of phonemes rather than phonetic distortions (Leonard, 1995). Prosody, speech rhythm, rate and intonation may also be affected (Enderby, et al., 2009). Some speech difficulties may be traced to structural or functional problems with the articulation system itself. Others, however, are of unknown aetiology. It has been suggested that deficits in the cognitive-linguistic domain underlie SSD. Studies have shown that some, if not all, young children with SSD have speech perception deficits (Edwards, Fox, & Rogers, 2002) and poor phonological awareness skills (Rvachew, Ohberg, Grawburg, & Heyding, 2003). Most researchers in this field agree that an adequate underlying knowledge of the sound system of the ambient language is essential for the acquisition of articulation. In the US, SSD has an

---

[2] Although the majority of a child's phonological system is acquired by the age of 4 – 4;6 years so that his/her speech is sufficiently intelligible to those outside his/her immediate social circle (Grunwell, 1992).

estimated prevalence of 3.8% in 6 year old children and is 1.5 times more prevalent in boys than in girls (Shriberg, Tomblin, & McSweeny, 1999). A British study estimated that SSD has a prevalence in the UK of 1.5% at 3-5 years, 4.6% at 5-7 years, 12.6% at 6-12 years and 7.3% at 12-14 years (Law, Boyle, Harris, Harkness, & Nye, 2000). SSD is the most prevalent communication disorder in early childhood (Shriberg, et al., 1999).

Childhood speech disorders were originally considered to be disorders of generating oral-motor programmes. They were described in articulatory terms and labels such as ‘functional articulation disorder’ reflected this approach (Bishop, 1997). In recent years there has been a paradigmatic shift from articulation to linguistic and psycholinguistic research. Some children undoubtedly do have a pure motor deficit, but it now seems unlikely that motor deficits can fully explain the disorder. (See below for a discussion of other deficits in SSD).

While other terms are used in a clinical context to reflect the precise nature of an individual child’s speech disorder (e.g. ‘phonological disorder’, ‘articulation disorder’), the term SSD is now favoured in the literature as a generic label for speech disorders arising from both the articulatory (sensorimotor) and phonological (cognitive-linguistic) domains.

A final distinction should be drawn between speech disorders of unknown aetiology and those arising from structural, physiological or neurological factors such as cleft palate or hearing loss. The term ‘SSD’ is not generally applied to the latter clinical groups unless individuals have co-occurring motor planning and/or phonological deficits. A summary of the key deficits in SSD is presented below.

1.2.1 Linguistic deficits

Children with SSD are known to be at risk for related or co-occurring linguistic deficits which may contribute to later language or literacy difficulties.

Early developmental studies reported that children with SSD used shorter, less complex sentences (e.g. Menyuk, 1964). Such evidence has been supported by more recent reports of grammatical deficits in SSD that cannot be explained in terms of expressive phonology.

Rvachew, Gaines, Cloutier and Blanchet (2005) examined the relationship between articulation skills and errors on the plural (-*s*), possessive (-*'s*) and regular present tense third person singular (-*s*) morphemes. It was found that the SSD children had difficulties with these suffixes that exceeded their difficulties producing word-final singleton and cluster /s/ or /z/. The SSD children had particular difficulty with the present tense third person singular morpheme. This finding was replicated in a study by Mortimer and Rvachew (2010), who found that poor finite verb morphology combined with low mean length of utterance (MLU) in pre-kindergarten resulted in limited language output two years later. These authors also found that pre-kindergarten children with SSD made fewer attempts at finite embedded clauses than typically developing (TD) children.

In addition to possible grammatical deficits, it seems increasingly likely that the majority of children with SSD have a linguistic impairment affecting phonological development (Pennington & Bishop, 2009). Studies have shown that children with SSD perform worse than typically developing (TD) controls on measures of phonological awareness such as rhyme judgement and sound blending (e.g. Bird, Bishop, & Freeman, 1995; Larrivee & Catts, 1999; Raitano, Pennington, Tunick, Boada, & Shriberg, 2004). A phonological deficit is associated with reading outcomes in children with SSD. For example, Rvachew (2007) found that speech perception, rime awareness and onset awareness measured at four years of age predicted nonword decoding skills two years later. Rvachew (2007) notes that despite small sample sizes, previous studies of reading outcomes in SSD consistently described poorer literacy in this group compared to TD controls (Bird, et al., 1995; Larrivee & Catts, 1999; Nathan, Stackhouse, Goulandris, & Snowling, 2004). Rvachew also comments that there is considerable variation in the effect sizes reported in hers and earlier studies of reading outcomes in SSD. She suggests that there must be variables mediating the relationship between speech accuracy and literacy development. One such variable may be rapid serial naming (RSN; Pennington & Bishop, 2009), which is unimpaired in some children with SSD (Raitano, et al., 2004). Despite poor phonological skills, children with SSD often learn to read well unless they are also poor at RSN (Peterson, Pennington, Shriberg, & Boada, 2009).

1.2.2 Speech perception deficits

Most researchers agree that an adequate underlying knowledge of the sound system of the ambient language is essential for the acquisition of articulation. Studies have shown that some, if not all, young children with SSD have speech perception deficits, performing poorly in tasks such as categorical perception (e.g. Edwards et al, 2002). Performance on auditory discrimination is not necessarily related to severity of speech impairment - rather there may be an effect of task type (Bird & Bishop, 1992).

It is thought that speech perception deficits in SSD relate to the accuracy and/or precision of the underlying representations of phoneme categories. Rvachew and Jamieson (1989) reported that some children with SSD lack a phonemic category for the sound /ʃ/. These children assimilated stimuli containing /ʃ/ to the /s/ category in perception and production tasks involving both phonemes. Hoffman, Stager and Daniloff (1983) found evidence that misarticulating children lack knowledge of fine phonetic details rather than contrastive phonological representations. These authors reported that children who misarticulated /r/ had representational knowledge of a /r/ – /w/ contrast, but defined these phoneme categories in terms of atypical acoustic cues. In Hoffman et al.'s (1983) study , children with SSD produced /r/ with a second formant frequency that was mid-way between that appropriate for /r/ and that appropriate for /w/. The same children were more likely to categorise words containing sounds using this mid-way second formant frequency as exemplars of /r/. Typically developing children and adults were more likely to categorise the words as containing exemplars of /w/.

1.2.3 Memory deficits

Many early studies of the verbal memory skills of children with SSD did not report the non-speech language skills or the cognitive status of participants. Reported deficits in sentence recall (Saxman & Miller, 1973) and word lists (Locke & Scott, 1979) were therefore confounded by the possible involvement of a language and/or cognitive impairment.

Raine, Hulme, Chadderton and Bailey (1991) controlled for non-verbal cognition and receptive vocabulary in a study of the memory skills of children with SSD. They found a relationship between speech rate and verbal short-term memory, with speech-disordered children performing poorer than controls on measures of both skills. In addition, Raine et al. (1991) tested short-term memory in a task that did not have a verbal input or require a verbal response. They reported that the SSD group showed a memory deficit that was not attributable to difficulties with speech output.

Later studies which also controlled for language and cognitive skills have replicated the relationship between verbal memory and expressive phonology in children with SSD. Adams and Gathercole (1995) found some evidence that three-year-old children with poorer phonological memory skills produced more phonological errors in their speech, however this difference did not reach significance. It should be noted that Adams and Gathercole (1995) included a nonword repetition (NWR) test in their assessment of phonological memory; NWR accuracy may also depend on speech processing skills (see below for a discussion of NWR). A further study by Adams and Gathercole (1996) reported a correlation between a composite measure of speech rate and a composite memory score (again including NWR) in an unselected sample of four- and five-year-old children.

1.2.4 A single heterogeneous disorder?

Like SLI, SSD is a heterogeneous disorder and no single underlying cause has been identified. There is mounting evidence of a genetic susceptibility for SSD (for a review see Lewis, Shriberg, et al., 2006). As is the case with SLI, research (including familial aggregation and twin studies) indicates that SSD is familial and heritable (Lewis, Freebairn, et al., 2006) and linked to specific chromosome regions (Stein, et al., 2006). Candidate genes, including FOXP2 have been proposed for speech and language (for a review, see Fisher, 2005). An alternative explanation for non-motor speech disorders is that they are a surface manifestation of an underlying language disorder that primarily affects phonological development (Pennington & Bishop, 2009). There is also evidence to suggest that children with SSD may be affected by non-linguistic deficits, such as motor impairments. Bishop (2002) presented data from a twin study that suggested that motor impairment is part of the phenotype of heritable SLI, and that the link between communication impairments and motor difficulties is stronger for speech than for language impairment.

Attempts have been made to classify SSD into discrete subtypes. Shriberg and colleagues introduced a complex framework of seven putative subtypes of SSD (e.g. Shriberg, 1993; Shriberg, Austin, Lewis, McSweeny, & Wilson, 1997; Shriberg, et al., 2005).

- speech delay – genetic;
- speech delay – otitis media with effusion;
- speech delay – apraxia;
- speech delay – dysarthria;
- speech delay – developmental psychosocial involvement;
- two categories of speech errors limited to distortions of speech sounds (as listed in Lewis, Freebairn, et al., 2006).

There is some evidence to support this classification system (Hauner, Shriberg, Kwiatkowski, & Allen, 2005 for speech delay - psychosocial involvement) and to identify diagnostic markers for the subtypes (Shriberg, et al., 2005 for speech delay - genetic). However the practical (diagnostic) applicability of the system has been questioned on the grounds that superficially similar speech errors may be caused by different - unknown - aetiologies or underlying processing deficits (Fox, Dodd, & Howard, 2002).

An alternative classification system, based on surface phonology with linguistic underpinnings (rather than aetiology), was proposed by Dodd (1995). This system included four subtypes of SSD: phonological delay, consistent deviant phonological disorder, inconsistent deviant phonological disorder and articulation disorder. A further subtype, developmental verbal dyspraxia (DVD), was added to this system by Broomfield and Dodd (2004). However these authors found that while the symptoms of DVD appear to differ sufficiently from the other subtypes of speech disorder for a differential diagnosis to be made, cases of this subtype are rare at a population level.

A model with psycholinguistic origins was proposed by Stackhouse and Wells (1997). This model allows for deficits at various levels of input and output speech processing. Such models can be used to plan intervention as well as to identify specific levels of processing breakdown. They focus on a child as an individual with a potentially unique processing deficit and do not attempt to create groupings of children with superficially similar symptoms (Pascoe, Stackhouse, & Wells, 2005).

A clinically useful method of classifying children with SSD is to group them according to whether or not they have a co-occurring language disorder (Lewis, Freebairn, et al., 2006). This dichotomy has been validated by a large body of research that shows poorer outcomes for children with combined SLI and SSD than for children with isolated SSD (see below for a review).

1.3 Co-morbidity of SLI and SSD

Shriberg et al (1999) reported that approximately 11-15% of six year old children with a persisting speech delay also had SLI and that approximately 5 – 8% of six year old children with persisting SLI had speech delay. Conti-Ramsden et al (1997) found that 12% of a sample of 7-year olds with language impairment also performed poorly on measures of articulation accuracy and expressive syntax. This group corresponded to Rapin and Allen’s (1987) ‘verbal dyspraxia’ category. It seems that co-morbidity between SLI and SSD may vary with age, possibly because speech difficulties are more likely to resolve and leave a residual language impairment (Bishop & Edmundson, 1987).

Although SSD and SLI are more likely to occur independently of each other (Shriberg, et al., 1999), a double deficit may place a child at higher risk for poor reading and spelling and (s)he may have special educational needs (Lewis & Freebairn, 1992). This double deficit may also increase the risk of other co-morbid developmental disorders. For example, children with both SLI and SSD show increased rates of attention deficit hyperactivity disorder (ADHD) symptomatology compared to controls and children with only SSD (McGrath, et al., 2008).

A phonological deficit may be a key feature of SLI, SSD and dyslexia, yet its specific manifestation will depend on the presence of other deficits. It seems that a shared underlying phonological deficit may be the factor that associates these developmental disorders. Most children with language impairments also have reading impairments (Bishop, et al., 2009), but children with an isolated speech disorder may not develop literacy difficulties unless they have a co-morbid language impairment (Peterson, et al., 2009). Within the single deficit SLI population, Bishop and Snowling (2004) identified the presence of a phonological impairment as the factor separating children with co-morbid dyslexia from those who do not have a specific reading difficulty. Similarly, reading outcomes are predicted by early phonological awareness skills in children with single deficit SSD (Rvachew, 2007).

Lewis and colleagues have carried out a number of follow-up studies showing that children identified with SLI+SSD in early childhood perform more poorly on later measures of phonological awareness, language, reading and spelling than children originally identified with isolated SSD (e.g. Lewis, Freebairn, & Taylor, 2000; Lewis, Shriberg, et al., 2006). Worse outcomes are also predicted in adolescence and adulthood for preschoolers with SLI+SSD (Lewis & Freebairn, 1992).

In terms of grammatical performance, finite verb morphology is worse in children with SLI+SSD than in children with SLI only (Haskill & Tyler, 2007). Haskill and Tyler (2007) reported that both groups had lower language scores than TD controls, but only the double deficit group had significant difficulty producing the third person singular *–s*, past tense *–ed* and auxiliary and copular ‘be’.

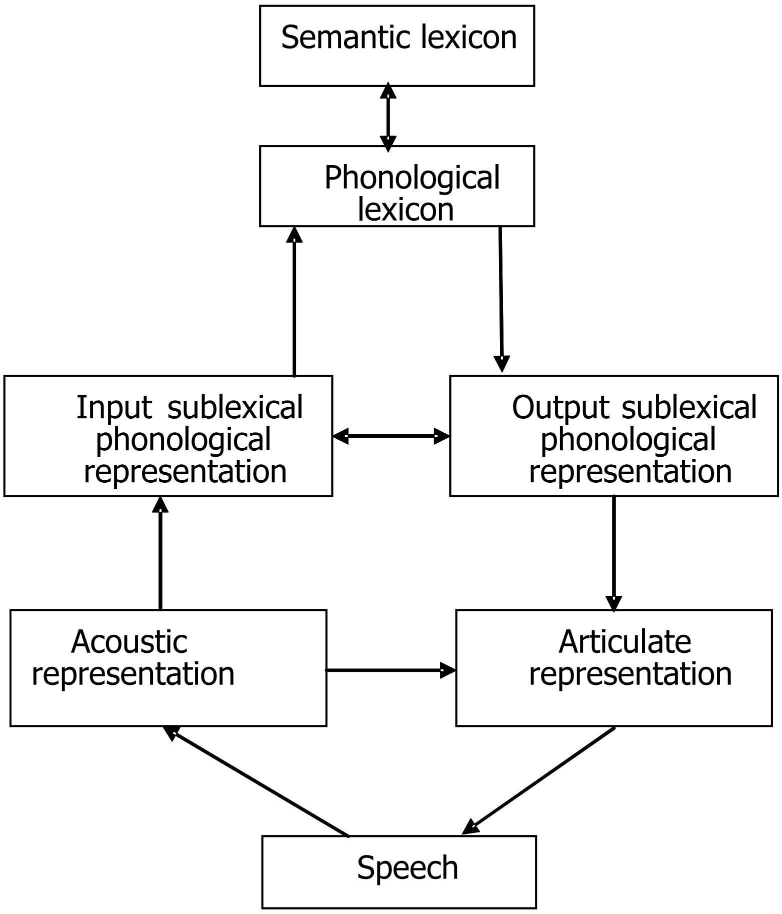
Tomblin (2009) defines co-morbidity as an overlap in two disorders that is not due to random occurrence, sampling bias or an overlap in diagnostic criteria. If both disorders exist, both would occur in some individuals by chance alone; this should not be considered co-morbidity. Children with a double deficit are likely to have significant difficulties which will result in help-seeking and the convergence of individuals with similar profiles in clinical populations. It is difficult to state with certainty, then, that SLI and SSD are ever truly co-morbid or whether they simply co-occur in some children.

1.4 Psycholinguistic models of speech processing

Psycholinguistic models are used to describe typical speech and language development as well as to investigate impairment to these systems. Such models hypothesise different levels of speech processing involved in the perception and production of speech (Vance, Stackhouse, & Wells, 2005), with links between input and output processing.

Within models such as that depicted in Figure 1.2 (adapted from Szenkovits & Ramus, 2005) nonwords tap phonological processing below the level of the semantic and the phonological lexicons. Nonword discrimination (NWD) tasks require input processing of a novel speech token (and a non-verbal response). Nonword repetition (NWR) tasks require input processing of the stimulus as well as output processing of the response. As such, NWR is a powerful tool which can be used to assess speech processing at all levels below the lexicon. NWD, on the other hand, potentially targets only input processes and will not reveal deficits to phonological systems involved in speech production.

Figure 1.2: Model of speech perception and production (adapted from Szenkovits & Ramus, 2005)



Nonword perception deficits have not been found in many studies of SLI, even in the presence of NWR deficits (Marton & Schwartz, 2003). These results suggest that the phonological deficit in SLI is at the output level. However, most studies have used separate stimuli for the NWD and NWR tasks. Furthermore, Szenkovits & Ramus (2005) used the same stimuli for NWD and NWR tasks and did find perception deficits in adults with dyslexia. Given that SLI and dyslexia share deficits to phonological systems, it seems there is a need to similarly investigate nonword perception and repetition deficits in SLI.

NWR is widely used to investigate language acquisition in both typically developing and clinical populations (Coady & Evans, 2008). Indeed, while the underlying reason is still unknown, children with SLI and dyslexia perform so poorly on NWR as compared to typically developing peers, that this task has been proposed as a marker for both disorders (Bishop, et al., 1996).

There have been many studies with typically developing children that focus on the relationship between NWR and other linguistic skills (for a complete review, see Coady & Evans, 2008). Correlations between NWR accuracy and receptive (but not expressive) vocabulary have been variously reported (e.g. Bowey, 2001; Gathercole & Baddeley, 1989). NWR skills are also correlated with measures of phonological working memory (Gathercole & Adams, 1994) and they have been found to predict narrative recall, grammatical complexity of utterances (Adams & Gathercole, 1996) and the ability learn new words (Michas & Henry, 1994).

The relationships between NWR and other language measures have been explained in terms of the wide range of underlying skills tapped by the NWR paradigm. NWR tasks model the phonological component of word learning (Coady & Evans, 2008). When presented with a nonword to repeat (equivalent to a novel real word), the child does not have an existing lexical representation or an existing motor programme to use to re-produce the stimulus (Vance, et al., 2005). NWR instead involves speech perception, segmentation of the acoustic signal into phonological units, short-term storage then retrieval of this phonological representation, followed by the formulation of a motor plan that assembles the appropriate phonological units and finally, articulation.

Few experiments have been carried out that use both NWD and NWR tasks to isolate speech processing deficits. Szenkovits & Ramus, (2005) used this task paradigm to investigate the phonological deficit in dyslexia. It appears that most experiments investigating the interaction between NWR and phonological skills (including phonological memory, phonological encoding and phonological awareness) use different tasks with different stimuli to measure each skill. It seems likely that a task battery controlling for the effects of stimulus variation and task demands would target speech perception and phonological skills more directly.

1.5 Speech processing and phonological development

During the first year of life and before the emergence of first words, an infant's speech perception skills develop significantly. Perceptual reorganisation occurs so that (s)he learns to pay selective attention to speech sound contrasts in the ambient language. The boundaries between phonemes are adjusted to reflect the ambient categories. By 12 months of age an infant can no longer discriminate between non-native phonetic contrasts even though (s)he could do so at the age of three months (Werker & Tees, 1984). This perceptual fine-tuning continues gradually throughout childhood, most likely as a result of repeated exposure to words, and adult phoneme boundaries are not fully established until well into the second decade of life (Hazan & Barrett, 2000).

When a child starts to generate phonological representations of lexical items, those phonological representations are underspecified in terms of phonetic detail (Werker, Fennell, Corcoran, & Stager, 2002). This affects the young child's ability to recognise familiar words (Edwards, et al., 2002) or to learn new ones (Werker, et al., 2002). There appears to be a reciprocal relationship between perceptual encoding and vocabulary growth. Word learning and speech perception are correlated with vocabulary size in young children (Edwards, et al., 2002; Werker, et al., 2002) however the directionality of the relationship is not clear. It may be that children with a larger vocabulary are better at word learning because they can attend to fine phonetic detail (bottom-up influence). Alternatively it may be that having a larger vocabulary results in improved speech perception (top-down influence) (Edwards, et al., 2002).

The development of perceptual skills occurs in tandem with changes to the structure of phonological representations. A child's representations become increasingly segmental over time. There is evidence that infants start segmenting monosyllabic words from fluent speech between the ages of six and 7.5 months (for a review, see Jusczyk & Luce, 2002). Even between the ages of three and five years, children are more sensitive than adults to acoustic cues associated within the syllable rather than individual segments (Nittrouer & Studdert-Kennedy, 1987). Phonological awareness tasks have shed light on the increasingly segmental nature of the child's phonological representations. Children are able to segment words into increasingly smaller units as they get older, from word and syllable segmentation in the preschool years, through onset-rime segmentation at nursery age, progressing to the segmentation of individual phonemes during the early school years (e.g. Treiman & Zukowski, 1996).

Improvements in expressive phonology interact with a child's developing perceptual knowledge about phonemes. Edwards, Fox and Rogers (2002) found a small but significant effect of articulation accuracy when predicting word discrimination scores. This effect was independent of vocabulary size, and showed that children who produced more sounds correctly in more word positions performed better on a discrimination task. Edwards et al. (2002) suggest that speech perception and speech production have a 'symbiotic' (p.241) relationship with both skills building on each other.

1.6 The perception-production link

The idea that there is an innate link between speech perception and speech production has been of interest for decades. The motor theory of speech perception originated in the 1950s with the work of Liberman and colleagues. Liberman (1957) presented evidence from categorical perception experiments in which listeners perceived acoustically dissimilar speech events as belonging to a single phoneme category. He claimed that articulatory gestures and their sensory effects must mediate between the acoustic features of speech on the one hand, and speech perception on the other.

This theory has been the focus of a considerable body of research, with evidence suggesting that two of its main claims stand up to scrutiny (see Galantucci, Fowler, & Turvey, 2006 for a full review of this theory). Firstly, it seems plausible that the perception of speech involves the perception of gestures. Perhaps the most widely-known evidence for this claim is the phenomenon known as the McGurk Effect (McGurk & Macdonald, 1976). This describes the integration of auditory and visual information in speech perception, whereby a listener's perception of a syllable is affected simply by watching a speaker produce a different syllable to that which is heard. Alternative explanations for this effect, such as auditory coding of visual information, have not been explored.

Secondly, the so-called mirror neuron system could play a fundamental role in perception and learning. Mirror neurons have been studied in macaque monkeys and there is evidence that audiovisual mirror neurons in the premotor cortex discharge when a monkey performs a specific action, when it observes the same action being performed by another monkey or a human and – of particular significance to the current discussion - when it hears the related sound (Kohler, et al., 2002). Evidence is emerging from imaging (fMRI, MEG), EEG and TMS studies that the motor system in humans has similar mirror properties. Hauk et al (2004) found that when participants read action words associated with the face, arm or leg, areas of the motor strip adjacent to or overlapping with those activated in movement of these body parts were activated. Furthermore, motor circuits invoked during articulation of speech sounds are activated when listeners hear these sounds (Pulvermuller, et al., 2006).

1.7 Consonant clusters

Consonant clusters occur in many of the world's languages, and are common in English. It is reported that a third of English monosyllabic words begin with a consonant cluster (McLeod, van Doorn, & Reed, 2001).

The acquisition of consonant clusters is a protracted process that challenges even typically developing children. Word initial clusters are acquired later than those occurring in word final position and it has been proposed that they are more complex in terms of articulatory demands (Kirk & Demuth, 2005). Longitudinal studies in English have shown that word initial consonant clusters emerge at around two and a half years of age. The phonemic inventory of a child this age will typically contain simplified word-initial clusters that are not permissible in English (/pw/ and /bw/ derived from /pl/ and /bl/). By the age of three years, the permissible clusters /st/ /sp/ and /pl/ may be produced appropriately in word initial position (Watson & Scukanec, 1997). It is some years before a full inventory of consonant clusters is established, with children as old as nine years of age still learning to master two- and three-element s-clusters (Smit, et al., 1990). It is unsurprising, then, that the development of consonant clusters is particularly vulnerable to impairment. Most children with speech sound impairments are reported to have difficulty producing appropriate consonant clusters (e.g. Dodd & Iacano, 1989).

Clearly consonant clusters are linguistically important, late acquired and vulnerable to impairment, yet very little attention has been paid to them in the field of speech perception.

Gathercole and Baddeley (Gathercole & Baddeley, 1990) found no selective impairment in the repetition of nonwords containing clustered rather than single consonants even though the articulatory ouput procedures required for the clusters are considerably more complex. Their articulatory rate was no slower than that of control children who had superior memory abilities. Therefore, disturbance of articulatory procedures and associated subvocal rehearsal processes can be discounted as a likely basis for the poor memory performance of the SLI children in their experiments.

1.8 Polysyllabic words

Polysyllabic words usually contain more phonological constituents than monosyllabic or disyllabic words (e.g. more phonemes and levels of stress) and as a consequence they may provide insight into speech production and phonological processing (James, Van Doorn, & McLeod, 2008). More production errors are observed in polysyllabic vs. shorter words produced by school-aged children and such errors persist to an older age (James, 2006).

James (2006) identified five stages of polysyllabic word acquisition which are summarised in Table 1.1. James suggests that polysyllabic word production that falls outside these broad criteria may indicate speech impairment and in turn may provide insight into the quality of a child's phonological representations and phonological awareness.

Table 1.1: Adaptation of James' 2006 model of polysyllabic word acquisition

Stage	Age (years)	Key features matching ambient language
1	1 – 2;3	Stressed syllables + duration of word
2	2;4 – 3;11	No. of syllables + prosody
3	4 – 6;11	No. of syllables + accuracy of phonemes outranking prosody
4	7 – 10;11	No. of syllables + phonemes + prosody
5	11+	Normal production

Experiments have been carried out to test repetition of polysyllabic words. Typically developing children aged between three and seven years are significantly worse at repeating bi- and polysyllabic nonwords than those with a single syllable (Vance, et al., 2005). However, polysyllabic words have not been used in categorical perception experiments.

1.9 Nonword repetition (NWR)

1.9.1 What does NWR measure?

The ability to repeat multisyllabic nonwords is subject to a high degree of variation during childhood and NWR is a task that has been proposed as a reliable clinical marker for SLI (Bishop, et al., 1996). Despite a considerable body of evidence showing that NWR predicts language learning, there is still debate about which skills NWR measures.

Gathercole and Baddeley (1990) first proposed that NWR measured phonological short term memory (PSTM) and identified a deficit of this nature in children with SLI. This explanation of NWR has been supported by many studies, including those finding significant associations between NWR and other PSTM tasks such as word or digit span (Metsala, 1999).

Most researchers now agree that while PSTM capacity inevitably influences NWR accuracy, NWR must also reflect phonological processing skills. When a nonword is presented, a child does not have an existing lexical representation or motor program to use to repeat the stimulus. Even before the nonword can be stored in PSTM, acoustic-phonetic information must be perceptually encoded as phonemes and syllables (input processing stages). The resulting representation must be retrieved from PSTM then used to plan and execute a novel motor programme (output processing stages; Vance, et al., 2005). Deficits in any or all of these operations could adversely affect NWR accuracy.

Ceponiene et al (1999) investigated the role of input processing in NWR. They divided typically developing children aged seven to nine years into two groups according to how accurate they were on a NWR task. Participants in the group with high NWR scores were significantly better than those in the low-NWR group at discriminating between CV syllables (‘ba-ka’ and ‘ba-ga’) and non-speech tones in same/different tasks. Ceponiene et al. (1999) concluded that the ability to discriminate fine acoustic differences seemed to affect the generation of phonological representations of the nonword stimuli.

A number of other researchers have described the additional output processing stages that may be involved in NWR. Leitao et al. (1997) considered NWR to measure the phonological encoding skills of children with and without speech and language impairments. They suggested that NWR depended on accurate perception and analysis of nonword stimuli so that an appropriate articulatory programme could be compiled. Leitao et al. (1997) suggested that tasks such as NWR could be useful to screen children with clear speech for potential subtle language deficits.

Regardless of the precise behaviours measured by NWR, this task seems to be associated with language impairments. Children with SLI show a consistent NWR deficit of a magnitude that makes it a good candidate for a behavioural marker for SLI (Gathercole, 2006). There is also evidence that the CNTNAP2 gene is associated with both NWR and SLI (Vernes, et al., 2008). Finally, deficits on a particular measure of NWR, the CNRep (Gathercole & Baddeley, 1996a), in children with SLI have been found to have a strong genetic basis (D. V. M. Bishop, et al., 1999; D. V. M. Bishop, et al., 1996).

1.9.2 NWR and word learning

If children are unable to generate sufficiently accurate phonological representations to support repetition of nonwords, they will have difficulty generating representations of novel real words. There are two main accounts of the relationship between nonword repetition (NWR) and word learning as indexed by receptive vocabulary (Gathercole, 2006). The first account holds that NWR accuracy depends on phonological working memory capacity. In support of this view, there is evidence from typically developing (TD) children that NWR accuracy is closely associated with scores on digit span, a widely used measure of phonological short term memory (Gathercole, Willis, Baddeley, & Emslie, 1994). Furthermore, stimulus length affects NWR performance in young TD children (Vance, et al., 2005) and children with SLI (Bishop, et al., 1996; Dollaghan & Campbell, 1998). The NWR deficit in SLI can be explained at least in part by an impairment of phonological storage as children with SLI also perform poorly on measures of serial recall (Archibald & Gathercole, 2006a).

A number of findings support the alternative account that NWR is supported by long-term lexical knowledge. Firstly, nonwords that have been rated higher on adult ratings of wordlikeness are repeated more accurately by young TD children than nonwords that are rated low in wordlikeness (Gathercole, 1995). Secondly, nonwords with stressed syllables corresponding to real words are repeated more accurately by TD children (Dollaghan, Biber, & Campbell, 1993; Dollaghan, Biber, & Campbell, 1995). Furthermore, real word syllables contained in nonwords are repeated more accurately than nonword syllables in stimuli of four (but not two or three) syllables (Metsala & Chisholm, 2010). Thirdly, TD children repeat nonwords more accurately if the stimuli contain sequences that are high versus low in phonotactic probability (Edwards, Beckman, & Munson, 2004). Finally, syllables from dense versus sparse lexical neighbourhoods are repeated more accurately when constituting part of three or four syllable nonwords (Metsala & Chisholm, 2010).

Both accounts of the relationship between receptive vocabulary and NWR rely on the generation and storage of adequate phonological representations. Within the lexicality model, lexical knowledge may influence perceptual analysis and the construction of phonological representations to be stored in phonological working memory. Alternatively, nonword stimuli may partially overlap existing lexical phonological representations at input, and these representations are used to fill the gaps when incomplete output phonological representations are retrieved from phonological working memory (Gathercole, Frankish, Pickering, & Peaker, 1999). Within the phonological store model, variation in storage capacity between individuals may result from differences in perceptual encoding or in the stability of the representations over time.

1.9.3 Tests of NWR

Many studies of nonword repetition have used one of two tests: in the UK, the CNRep (Gathercole & Baddeley, 1996a) and in the US, the NRT (Dollaghan & Campbell, 1998). The two assessments differ in several important ways.

Firstly, the CNRep contains 40 nonwords of between two and five syllables in length. Some of the stimuli contain consonant clusters (e.g. /ˈprɪndl/) and most nonwords include a weak syllable with a reduced vowel (e.g. /ˈhæm*pənt*/). Many of the stimuli contain real words or morphemes (e.g. ‘ing’ in /ˈblɒntəˈsteɪp*ɪŋ*/). In ‘live’ presentation, the administrator is instructed to present the nonwords with natural English prosody; stimuli may also be presented from a recorded tape. Responses should be scored online as correct/incorrect, and deviations consistent with a strong regional accent should be ignored.

In contrast, the NRT contains only 16 nonwords ranging from one to four syllables in length. None of the nonwords includes consonant clusters, and all constituent phonemes are late acquired and acoustically salient. No constituent parts of the NRT stimuli correspond to lexical items. There are no weak vowels in the nonwords so they should be presented with equal stress on each syllable. Responses are scored after the test has been administered, by calculating the percentage of phonemes correctly repeated in appropriate positions.

The differences between the CNRep and the NRT are significant because most have been implicated in the NWR deficit in SLI (see below, and Archibald & Gathercole, 2006a). This means that a language impaired child will likely perform differently on the two tests. Similarly, direct comparison of children who have completed different NWR tests is invalid.

The following section describes the factors that affect performance on NWR tests and which are significant to comparison of the CNRep and NRT.

1.9.4 Factors affecting NWR accuracy

NWR accuracy in children, particularly those with SLI, is largely dependent on the structure of the nonword stimuli. An effect of stimulus length has been identified in typically developing children (Vance, et al., 2005) and in children with SLI (Bishop, et al., 1996; Dollaghan & Campbell, 1998), with longer nonwords being the most challenging to both populations. The CNRep may be harder to complete than the NRT on the grounds that it contains longer stimuli (five syllables vs. the four syllable nonwords in the NRT). The CNRep also includes no items of one syllable, unlike the NRT.

Consonant clusters are acquired late in typical development (Smit, et al., 1990). In addition, most children with speech sound impairments have difficulty producing appropriate consonant clusters (e.g. Dodd & Iacano, 1989) and there is evidence that children with SLI have a selective impairment in their repetition (Gathercole & Baddeley, 1990). The NWR accuracy of children with SLI is also affected by the position of a consonant cluster and whether or not it is stressed (Marshall & van der Lely, 2009). The CNRep (but not the NRT) contains consonant clusters but their position and stress is not controlled so some test items may be more difficult to repeat than others.

The NWR accuracy of both typically developing (Dollaghan, et al., 1995) and language impaired children (Munson, Kurtz, & Windsor, 2005) is affected by the wordlikeness of the stimuli. This means that children are likely to perform better on a NWR test, like the CNRep, that includes nonwords containing sub-parts resembling real words or morphemes.

Finally, attention should be paid to the administration instructions when carrying out and scoring a test of NWR. Wells (1995) notes that in order for standardisation to be valid, the tester must aware of what constitutes a NWR error. For example many tests, like the CNRep, exclude errors that represent dialectal variation.

1.9.5 Summary

"Understanding the processes involved in nonword repetition is… not only important for theoretical analysis of language learning in typically developing populations [and in adults], but may also hold the key to understanding developmental disorders of language learning" (Gathercole, 2006, p. 519).

2. Methods

This chapter presents a description of the screening tests, published assessments and experimental tasks that were carried out in Experiments 1 and 2 (reported in Chapters 3 and 4). The assessment battery for Experiment 1 was designed to yield a comprehensive profile of the verbal and non-verbal skills of typically developing (TD) children aged 5;0 to 9;11. Some minor changes were made to the battery for Experiment 2: (1) several tasks were eliminated so that the battery was appropriate for testing younger TD participants and those with speech and language disorders; (2) tasks were added so that reading could be assessed and phonological short term memory (PSTM) could be measured in more detail.

2.1 Equipment

The experimental tasks and most of the standardised assessments were administered on a Fujitsu-Siemens Amilo SI 1520 laptop running Windows Vista. Participants used Sennheiser HD 25-1 headphones for tasks with recorded stimuli that required a non-verbal response. Where tasks required a verbal response to a recorded stimulus, participants used a Sennheiser PC151 headset with integrated microphone. Any other verbal response was recorded via an Eagle G157B clip-on microphone.

2.2 Screens

2.2.1 Hearing screen

All participants were required to pass a pure-tone audiological screen at or better than 25dB HL at frequencies of 250Hz, 500Hz, 1kHz, 2kHz, 4kHz and 8kHz in both ears. The hearing screen was administered via a laptop connected to an external attenuator with tones presented via Telephonics TDH-39 headphones. The laptop/headphone combination was calibrated using a Brüel & Kjær type 4153 artificial ear.

2.2.2 Computer familiarisation screen

This task, originally devised by Vance, Rosen and Coleman (2009), was carried out to familiarise the participants with the laptop and the mouse. Each child was required to click on a series of cartoon babies which appeared on the screen in random locations. After each successful mouse click on a baby, the animation disappeared from the screen. Participants who had difficulty manipulating the mouse comfortably were given the opportunity to respond by pointing at the screen in computer-based tasks.

2.2.3 Informal reading screen (Experiment 2 only)

Anticipating that the youngest participants would be unable to complete the published test of nonword reading (see below), an informal reading screen was included to evaluate letter and word knowledge. The task was administered via laptop and headset with recorded instructions. Participants progressed from letter naming to decoding real word then nonword digraphs. The screen was scored out of a total of 31 points.

2.3 Published assessments

2.3.1 Non-verbal cognition

The Block Design subtest of the Wechsler Intelligence Scale for Children – Third Edition UK (WISC; Wechsler, 1992) was administered to assess the non-verbal cognitive skills of all participants in Experiment 1. For this task, participants were required to manipulate two, four then nine red and white blocks so that the tops of the blocks displayed the same geometric pattern as a stimulus picture. Each task item had to be completed accurately and within a time limit. Instructions for this task were given live by the experimenter and the stimulus pictures were presented via PowerPoint. The task yielded a raw score (consisting of points awarded for accuracy and speed) for each participant and a scaled score for those over the age of 6 years.

The same subtest of the WISC was administered to Experiment 2 participants over the age of 6 years. Younger participants carried out the block design subtest of the Wechsler Preschool and Primary Scale of Intelligence - Revised UK Edition (WPPSI; Wechsler, 1990) which is standardised for children aged 2 years 11½ months to 6 years 11½ months. This test was administered in a similar way to the WISC (see above), but children were asked to reproduce geometric designs using three or four flat, red/white tiles, sometimes after an adult demonstration. The subtest was timed and participants were given bonus points on some items for speed. The WPPSI yielded raw and scaled scores for children below the age of 6 years.

2.3.2 Working memory

In Experiment 1, phonological short-term memory (PSTM) was assessed using the Digit Recall subtest of the Working Memory Test Battery for Children (WMTB-C Pickering & Gathercole, 2001). In this test, participants were required to listen to, then repeat, sequences of digits. There were three practice trials followed by blocks containing lists of increasing length. Each participant started the test items at a level dependant on their success in the trials, with lists of one, two or three digits. If a child responded correctly to four trials within any block, the test proceeded to the next block and the child was given credit for any omitted items. The test was discontinued after three errors had been made within any block. This task yielded a raw score, standard score and percentile rank for all participants.

In the first phase of Experiment 1, stimuli for the Digit Recall subtest were presented live by the experimenter in an even monotone and at the rate of one digit per second. Automated presentation of recorded stimuli was introduced in the latter stages of Experiment 1, after a comparison of live vs. recorded stimulus presentation revealed no effect of stimulus presentation type (see Supplementary Study below). The automated Digit Recall subtest was administered via laptop and headset using DMDX software (developed by Jonathan Forster, University of Arizona). All participants in Experiment 2 carried out the automated task.

In Experiment 2, the Word Recall subtest of the WMTB was also administered in order to gain a more precise picture of participants' PSTM. In this subtest, lists of words were presented in the same way as the digit task, via laptop and headphones. The task yielded a raw score, standard score and percentile rank for all participants.

2.3.3 Phonological awareness

In Experiment 1, phonological awareness was assessed using the Rhyme, Spoonerism and Naming Speed subtests of the Phonological Assessment Battery (PhAB; Fredrickson, Frith, & Reason, 1997). In the Rhyme task, three recorded words were presented via the laptop and headset and participants were required to repeat the two words that sounded the same at the end (e.g. “sail, boot, nail” gives “sail, nail”). The first three trials were practice items for which feedback was given; these were followed by 21 test trials. The number of correct responses was totalled to obtain a final raw score.

The Spoonerism subtest had two parts, both of which were presented via the laptop/headset. In the first part of the subtest, listeners were required to replace the initial phoneme of a word with a new consonant (e.g. “red with a [b] gives bed”). In the second part, a noun phrase consisting of two words was presented. Listeners were required to exchange the first sound of each word (e.g. “daisy log” gives “lazy dog”). For each part, feedback was provided for the first three (practice) trials and there were ten test items. As with the Rhyme subtest, the number of correct responses was totalled to obtain a final raw score.

The Naming Speed subtest assessed the speed of phonological production. There were two parts to the task: rapid picture naming and rapid digit naming. For the picture naming, line drawings of four familiar objects were repeated randomly in lines on a page. Participants had to label the pictures as quickly as they could with no mistakes. The time taken to complete this task was recorded in seconds. The digit naming task used digits instead of pictures and was again timed. The raw scores of the naming tasks consisted of number of seconds taken, such that a lower score indicated a better performance.

The phonological awareness tasks yielded raw scores for all participants and standard scores for children over the age of six years.

The Spoonerism and two Naming Speed tasks were eliminated from the Experiment 2 battery on the following grounds:

- The Spoonerism task was not engaging for participants in Experiment 1. Many children struggled with the practice trials and lost motivation for that task and any that followed. It was anticipated that the youngest children in Experiment 2 would have difficulty performing above floor level on this test.

- Data from the Rhyme and Spoonerisms tasks were correlated in Experiment 1 ($p < 0.05$), indicating that the either task alone adequately represented participants' phonological awareness. Spoonerisms could therefore be eliminated without affecting the evaluation of phonological awareness.
- It was predicted that the youngest TD children would have difficulty naming digits in the digit Naming Speed test. It was also predicted that accurate scoring of the picture naming task would be difficult for participants with speech production difficulties.

In addition to the Rhyme task, participants in Experiment 2 also carried out the Alliteration Test with Pictures and the Nonword Reading subtests of the PhAB. The Alliteration test was introduced because it did not require a verbal response, thereby eliminating potential difficulties scoring data from participants with speech difficulties. For this task, participants were shown pictures of three objects, listened to recordings of each object name, then pointed to the two objects that started with the same sound.

2.3.4 Speech sound production

The diagnostic screen of the Diagnostic Evaluation of Articulation and Phonology (DEAP; Dodd, Hua, Crosbie, Holm, & Ozanne, 2002) was administered to assess the speech sound production abilities of all participants. The diagnostic screen is a picture naming task designed to elicit the full range of English consonants and vowels. Instructions were given live by the experimenter and the stimulus pictures were presented via PowerPoint on a laptop. The consonant inventory of each child was compared to the published norms and scored as either a pass (presenting with an adult inventory or age appropriate error patterns) or a fail (presenting with developmentally unacceptable errors). All children who took part in Experiment 1 were typically developing, yet one participant failed the speech screen; his data were excluded from the analysis. Eight participants with language difficulties failed the DEAP screen in Experiment 2; they were assigned to the SLI+SSD experimental group.

2.3.5 Sentence comprehension

The Test for Reception of Grammar-2 (TROG; Bishop, 2003) was used to assess sentence comprehension. Participants were required to look at four on-screen pictures and click on the one that went best with a recorded sentence. There were two practice items, then 20 test blocks, each of which contained four items. Each block tested understanding of a particular type of grammatical structure e.g. negative, reversible passive. The blocks were arranged in order of increasing difficulty. A child passed a block if (s)he responded correctly to all four of its items. The test was discontinued when five consecutive blocks were failed. This task yielded a raw score, standard score, percentile rank and age equivalent for all participants.

2.3.6 Single word comprehension

Single word comprehension was assessed using the short form of the British Picture Vocabulary Scale II (BPVS; Dunn, Dunn, Whetton, & Burley, 1997). Four picture stimuli appeared on the screen and the participant was required to click on the one that corresponded to the word that he/she heard. Each participant began the test at a block determined by his or her age, and each began with four practice items. The vocabulary became increasingly difficult and the test was discontinued when the participant made eight or more errors within a single testing block of 12 trials. The task yielded a raw score, standard score and age equivalent for each participant.

2.3.7 Expressive vocabulary

The Renfrew Word Finding Test (RWFT; Renfrew, 1995) was used to assess expressive vocabulary. This assessment consists of a set of black and white line drawings which were presented to the child one at a time. For each item, the child was instructed to name the single object depicted (e.g. ‘snake’), a specific part of the picture (e.g. ‘sleeve’ and ‘cuff’ when presented in a picture of a shirt) or a common noun (e.g. ‘jewellery’ depicted as a group of items). The target vocabulary became progressively more complex. The test was discontinued when six progressive pictures had elicited no correct responses and the child was unable to name at least one picture when the following six pictures were presented as a group. This test yielded a raw score and an age equivalent for all participants.

2.3.8 Expressive language: grammar and information content

Expressive language was assessed using the Renfrew Action Picture Test (RAPT; Renfrew, 1997). This test uses 10 simple pictures to elicit samples of spoken language that may be evaluated in terms of the information given and the grammatical structures used. As each picture was presented to the child, a specific question was asked (e.g. “What has the cat just done?” for the target “Caught the mice”). If a child gave an incomplete response, a general prompt was given to encourage completion of the target utterance. Pictures were presented live by the examiner who also asked the questions; responses were recorded. This task yielded raw scores for grammar and information given, which were awarded according to the published guidelines.

2.3.9 Sentence recall

The Recalling Sentences subtest of the Clinical Evaluation of Language Fundamentals-3 UK (CELF; Semel, Wiig, & Secord, 2000) was administered to evaluate the participant’s ability to repeat sentences of increasing length and complexity without changing syntax, word meanings or word morphology. Each stimulus sentence was presented via headphones and the child’s response was recorded. The repetitions were scored for faithfulness to the stimuli according to published guidelines. The task yielded a raw score for all participants and a standard score and percentile rank for participants over the age of six years.

The Recalling Sentences subtest from the Fourth UK edition of the CELF (Semel, Wiig, & Secord, 2006) was administered in Experiment 2. This edition is standardised for a wider age range of children. The Recalling Sentences subtest yielded a raw score, standard score and percentile rank for all participants over the age of 5 years. The subtest was administered in the same way as for Experiment 1, although some of the stimulus sentences differed in form and content.

2.3.10 Nonword reading (Experiment 2 only)

The Nonword Reading subtest of the Phonological Assessment Battery (PhAB; Fredrickson, et al., 1997) was administered in order to evaluate participants’ reading skills. This subtest required participants to decode letter strings in nonwords of one and two syllables. Some of the nonwords were regular and could be decoded phonologically (e.g. “tib”). Some contained irregular spellings that required the application of common spelling rules (e.g. “leaze”). The task was presented via laptop and headset with recorded instructions. It yielded raw scores for all participants and standard scores for children over the age of six years.

2.4 Experimental tasks

2.4.1 Rationale

Within a psycholinguistic model of speech processing, comparison between performance on different tasks is considered to be as informative as performance on the individual tasks themselves (Vance, et al., 2005). A comparative methodology allows the identification and isolation of skills that are specifically associated with a surface behaviour. The experimental tasks were designed to investigate speech processing at the input versus output stages so that a comparative analysis of performance could be carried out.

It is generally acknowledged that nonword repetition (NWR) taps both speech input and speech output processes. To repeat a nonword, a child must discriminate sounds within the stimulus, before generating and executing an appropriate motor programme. In the developmental literature there are examples of the comparative paradigm being used to evaluate the underlying processes involved in NWR (Gathercole & Baddeley, 1990; Marton & Schwartz, 2003). However a notable feature of such experiments is that they use different stimuli for the input (usually discrimination) task and NWR.

It is arguable that using different stimulus material for nonword discrimination (NWD) and NWR tasks invalidates a direct comparison of performance, perhaps especially when testing children with language impairments. There is evidence that the length of a nonword stimulus affects the NWR of typically developing (TD) children (Vance, et al., 2005) and those with SLI (e.g. Gathercole & Baddeley, 1990). Similarly, children with grammatical SLI (G-SLI) repeat nonwords less accuracy as prosodic complexity increases (Gallon, Harris, & Van der Lely, 2007). The location and stress of a cluster is significant when testing NWR in children with SLI and/or dyslexia (Marshall & van der Lely, 2009). At a perceptual level, the precise acoustic properties of the speech signal are significant to children with SLI, who show worse categorical perception (CP) than controls for synthesised stimuli (Thibodeau & Sussman, 1979) but are unimpaired on natural speech stimuli (Coady, et al., 2005).

It seems, then, that if NWD and NWR stimuli are not carefully matched for e.g. length, complexity or precise acoustic properties, any of these factors could selectively impair performance on one task. A comparison between unmatched tasks could not reliably isolate skills associated with a surface behaviour.

Szenkovits and Ramus (2005) successfully used the same stimuli in a NWD and a NWR task to isolate the underlying processing deficit in dyslexic adults. The same paradigm was used in the experiments presented here. Identical natural speech nonword stimuli were used to evaluate NWD and NWR, and nonword minimal pairs from the same stimulus set were used to evaluate CP (with the addition of a real word minimal pair for Experiment 1; see below). These tasks were chosen in order to isolate skills potentially underlying NWR. CP is implicated in NWD and both are implicated in the input stage of processing for NWR.

2.4.2 Nonword stimuli

The experiments presented here were designed to investigate input and output phonological processing in typically developing (TD) children and children with SLI or SLI with a concomitant speech disorder (SLI+SSD). A set of nonword stimuli was systematically generated to be used in three experimental tasks: categorical perception (CP), nonword discrimination (NWD) and nonword repetition (NWR). The nonwords were carefully designed to allow direct comparisons between the tasks. Specifically, one experimental aim was to answer the question: if a participant cannot repeat a nonword, can (s)he accurately perceive it?

In order to address this question it was essential to elicit NWR error data from all groups, preferably with the TD group showing an age effect (across an age range of 4;0 to 10;0). The following factors were considered when generating the nonword stimuli:

2.4.2.1 Stimulus length

Vance et al. (Vance, et al., 2005) reported that typically developing children aged five to seven years repeat nonwords of three or four syllables significantly less well than they repeat shorter stimuli. They suggested that measures of speech production should use multisyllabic stimuli so that production errors in children over five years can be examined. Children with SLI of the age to be tested here also show consistent deficits repeating nonwords with three or more syllables but show smaller or non-significant difficulties repeating shorter nonwords (Bishop, et al., 1996; Dollaghan & Campbell, 1998).

Nonwords in the base stimulus set (from which all stimuli were generated; see below) consisted of three syllables. It was predicted that nonwords of three syllables would challenge participants in the clinical groups as well as the younger TDs. There would likely be an age effect in the TD group.

2.4.2.2 Complexity

The acquisition of consonant clusters is one of the longest processes in the speech development of TD children (McLeod, et al., 2001). In English, clusters emerge at around two and a half years of age in forms that are not consistent with the ambient language; a full cluster inventory is not established until around the age of nine (Smit, et al., 1990). Word-initial clusters are not morphologically salient and typically emerge later than word final clusters (Watson & Scukanec, 1997). The articulatory output procedures required for consonant clusters are considerably more complex than for singleton consonants. There is a well-defined order of acquisition of consonant clusters, and children typically master those starting with a stop + liquid (e.g. /pl/, /kl/) before those starting with a fricative + liquid (e.g. /sl/) (Smit, et al., 1990).

Most children with speech sound impairments are reported to have difficulty producing appropriate consonant clusters (e.g. Dodd & Iacano, 1989). There is mixed evidence that children with SLI have a selective impairment in the repetition of nonwords containing clusters. Gathercole and Baddeley (1990) found that language impaired children were not affected by the articulatory complexity of NWR stimuli containing clusters. However this finding was not replicated in a larger study by Bishop et al. (1996), who found a small but significant effect of articulatory complexity in children with SLI.

The specific consonant clusters used in the experiments reported here were selected primarily to ensure that the systematic generation of the stimuli resulted in tokens that were phonotactically viable and nonwords in English, and so that a single syllable real word minimal pair could be matched to a single syllable nonword for one CP subtest. The main nonword set and one nonword partial set used the early developing /kl/, /pl/, /pr/ and /br/ clusters. Nonwords containing later developing /sl/ and /fl/ clusters were included in the second nonword partial set. The partial sets were included to provide auditory variation in the NWR and NWD tasks. See Table 2.1 below (p. 62) for a full list of nonword stimuli.

2.4.2.3 Wordlikeness

Wordlikeness is a measure of how much a nonword stimulus resembles a real word (Munson, et al., 2005). Listeners may consider a nonword to be wordlike when, as a whole, it is close in proximity to a real lexical item. A nonword may also be considered wordlike if it contains subparts that are real lexical items. The NWR accuracy of typically developing (TD) children is improved when stimuli are rated as wordlike, suggesting that long-term lexical knowledge influences performance (Dollaghan, et al., 1995). Children with SLI show a larger effect for wordlikeness when compared to TD controls on NWR, with repetition accuracy decreasing as stimuli become less wordlike (Munson, et al., 2005).

Wordlikeness can be measured by the phonotactic probability of constituent subparts or by means of adult listeners’ intuitive ratings (Munson, et al., 2005). Neither measure was explicitly applied to the nonwords generated here. However, care was taken to create stimuli that the experimenter considered were unlike words known by children of the age to be tested.

2.4.3 Predictable error types

The base nonword set was adapted from the NWR errors observed in children with SLI and/or dyslexia by Marshall and van der Lely (2009). These authors generated four basic nonwords each of which consisted of three syllables and contained an onset cluster. Marshall and van der Lely's (2009) basic nonwords were then manipulated by changing the position of the word stress and the cluster (a similar manipulation will be described in further detail below for the current experiments).

van der Lely and Marshall (2009) reported that participants made six types of repetition errors on the clusters:

- C1 deleted e.g. /fləkɛtə/ → /ləkɛtə/;
- C2 deleted e.g. /tʌprəfə/ → /tʌpəfə/;
- C1 substituted e.g. /fəklɛtə/ → /fəglɛtə/;
- C2 substituted e.g. /ləfræpə / → /ləflæpə/;
- Creation of cluster e.g. /ˈfɛklətə/ → /ˈflɛkətə/; cluster creation took place even when the existing cluster was repeated correctly e.g. /ˈfɛklətə/ → /ˈflɛklətə/;
- Other error e.g. /prəfʌtə/ → /bəfʌtə/.

Of relevance to the current report, van der Lely and Marshall (2009) found that their clinical groups (SLI-only, dyslexia-only, SLI+dyslexia) made significantly more errors on word-medial clusters compared to word-initial clusters. The two dyslexic groups had difficulty producing clusters in unstressed syllables. The SLI+Dyslexia group performed worst overall. The cluster accuracy of typically developing (TD) children was not affected by cluster position or word stress, but all groups created clusters significantly more often in initial than in medial positions. The creation of clusters was particularly high in the SLI-only group. These data are suggestive of a selective impairment in a particular type of NWR task in children with linguistic deficits similar to those recruited to Experiment 2 reported here.

In the experiments described below, the same nonword set was to be used in a NWD and a NWR task. A sub-set of the items was to be used for the CP task. It was essential that each experimental task contained sufficient trials for direct comparisons to be carried out. However, the age and attention skills of the participants had to be taken into consideration: the tasks should not contain so many trials that the participants would lose interest. With these factors in mind, the principle characteristics of van der Lely and Marshall's (2009) stress/position manipulation were used to generate a base nonword set from which three NWR error types were predicted.

2.4.3.1 The nonword set

The nonword set was designed to include a base set of stimuli, similar to those used in van der Lely and Marshall's (2009) study. In anticipation of specific NWR production errors, the base nonwords were manipulated so that tokens likely to be produced as errors were also included in the stimulus set. Two partial sets of nonwords were also generated to provide auditory variation in the NWD and NWR tasks. All nonwords were viable according to the rules of English phonotactics.

Four 'base' nonwords were created. Each had three syllables and contained a /kl/ cluster in one of four word positions: word-initial stressed (ˈklepəfə), word-initial unstressed (kləˈpefə), word-medial stressed (fəˈklepə) or word-medial unstressed (ˈfekləpə). In this way, the /kl/ cluster appeared in each combination of word and stress position.

Each base item was then manipulated in three ways to create a set of related nonwords:

- the second consonant of the cluster was deleted (C2 deletion, e.g. /ˈklepəfə/ → /ˈkepəfə/);
- the second consonant of the cluster was moved (C2 movement, e.g. /klepəfə/ → /ˈkepləfə/ or /ˈfəklepə/ → /ˈfləkepə/);
- an additional consonant was inserted to create a second cluster in the word (C2 insertion e.g. /ˈklepəfə/ → /ˈklepləfə/).

These manipulations yielded an additional 12 three syllable nonwords.

A set of single syllable nonwords was added to the stimulus set to check that participants were able to repeat and perceive clusters in the shortest stimuli (Marshall & van der Lely, 2009). The stressed syllable of each three syllable nonword was extracted and this process yielded four word-initial singleton/cluster minimal pairs: `kep - klep`, `pef - plef`, `(kep) - (klep)`, `fek - flek`. As two of the base nonwords contained the same stressed syllable (`kep - klep`), this minimal pair was eliminated from the nonword set (see Table 1.1). The single syllable minimal pair for one stimulus set yielded the nonwords /fek/ and /flek/. Both items are real words (the first in some dialects of the UK), however they were included on the grounds that children of the age and dialect to be tested were unlikely to know them as such.

Finally, auditory variation was introduced by creating two partial sets of nonwords. That is, sets of nonwords that were generated from two additional base nonwords, but which did not undergo all the manipulations described above. These partial sets consisted of a three syllable base nonword with a word initial, stressed cluster /sl/ and a three syllable base nonword with a word-medial unstressed cluster /br/. Both nonwords underwent only the error type manipulation described above. There were a total of 24 three syllable nonwords and a total of 10 one syllable nonwords in the complete set (see Table 1 for a full list of nonwords).

All stimuli were recorded in an anechoic chamber by two adult speakers: a male and a female, both of whom have standard southern British English accents. Stimuli were high pass filtered to exclude low frequency background noise (cut-off = 50Hz), but no further manipulation of the stimuli took place.

Table 2.1: Nonwords used in experimental tasks.

	Base nonword	Manipulation: C2 deletion	Manipulation: C2 movement	Manipulation: C2 insertion	1 syllable derivatives	
Main nonword set						
wi cluster, stressed	ˈklepəfə	ˈkepəfə	ˈkepləfə	ˈklepləfə	kep	klep
wi cluster, unstressed	kləˈpefə	kəˈpefə	kəˈplefə	kləˈplefə	pef	plef
wm cluster, stressed	fəˈklepə	fəˈkepə	fləˈkepə	fləˈklepə	(kep)	(klep)
wm cluster, unstressed	ˈfekləpə	ˈfekəpə	ˈflekəpə	ˈflekləpə	fek	flek
Nonword partial set 1						
wi cluster, stressed	ˈslifətə	ˈsifətə	ˈsiflətə	ˈsliflətə	sif	slif
Nonword partial set 2						
wm cluster, unstressed	ˈpabrənə	ˈpabənə	ˈprabənə	ˈprabrənə	pab	prab

Note: wi = word-initial, wm = word-medial
Total number of three syllable nonwords = 24; total number of 10 one syllable nonwords = 10

2.4.4 Nonword discrimination (NWD) task

A NWD task was carried out using the nonword stimuli described above and SipSli software (custom software developed by Mike Coleman, Division of Psychology and Language Sciences, UCL).

All stimuli that were presented in the NWD task contained a consonant cluster (e.g. /ˈklepəfə/). There were three foil types:

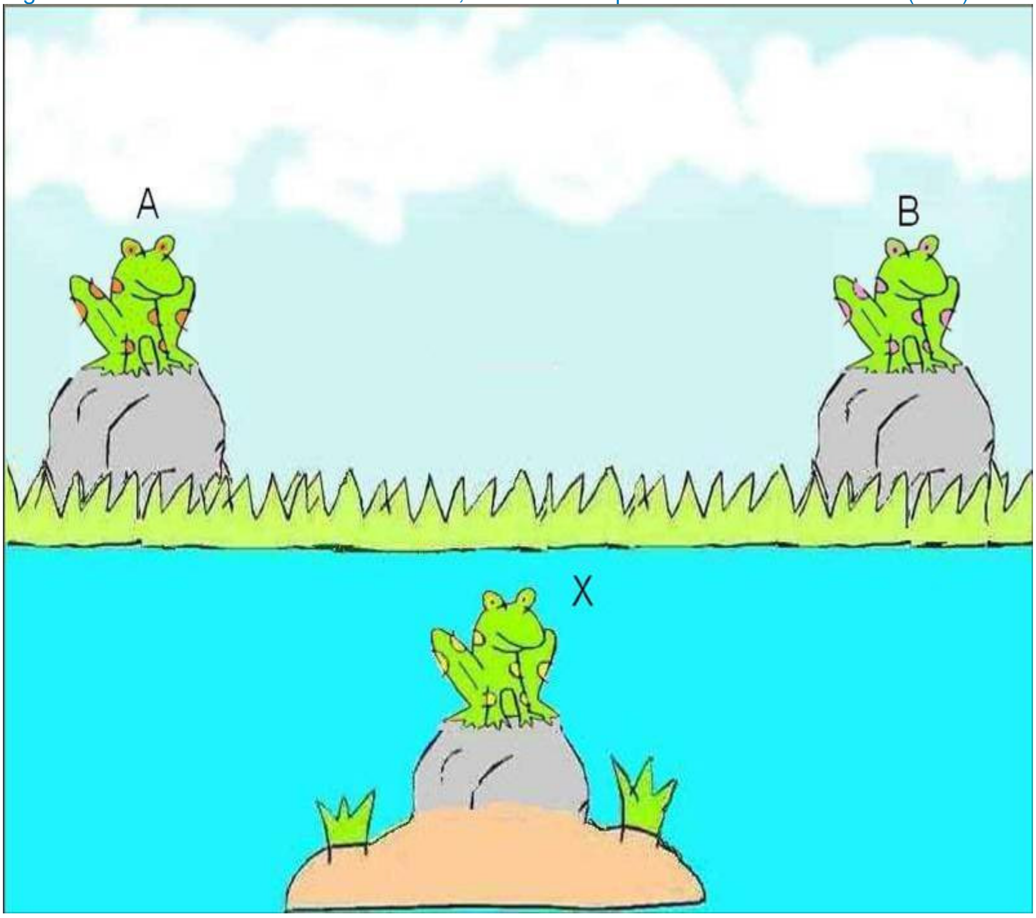
- The second consonant in the cluster (C2) was deleted (e.g. /ˈkepəfə/). Erroneous selection of such a foil was considered to be a C2 deletion error;
- The second consonant in the cluster was moved into a different word position (e.g. /ˈkepləfə/). Erroneous selection of such a foil was considered to be a C2 movement error;
- The stimulus cluster was reproduced, but an additional cluster was created by inserting a second consonant in another word position (e.g. /ˈklepləfə/). Erroneous selection of this foil was considered to be a C2 insertion error.

An AXB procedure was used. This procedure was chosen for two main reasons: (1) in order to avoid the use of arbitrary pictorial or textual labels; (2) to minimise the memory load of the task, as the listener is not required to remember all the stimuli to make a correct response. The child was presented with three frogs on the screen in an inverted triangle formation (see Figure 2.1). (S)he was instructed that each frog would say a ‘silly’ word and that (s)he must decide, then click on, whichever of the top two frogs (A or B) said the same as the frog in the bottom middle (X). Frog A spoke first, followed by frogs X then B. Frogs A and B used the female voice and frog X used the male voice. This design ensured that participants had to phonologically encode the nonwords and could not make a judgement based solely on their acoustic similarities.

A practice block of four items was carried out to familiarise the child with the task demands. Verbal and visual feedback was given during this stage. The practice was followed by 56 test items divided into 4 blocks. Each test block was followed by a ‘reward’ during which the child viewed an on-screen animation of a changeable face and balloons. The child was required to click on a ‘go’ button to proceed from one test item to the next item. This ensured that (s)he was attending to the task before each stimulus was presented. The inter-stimulus interval was set at 450ms.

Each test presentation consisted of either a three- or a one-syllable base nonword as stimulus (X) and one of the manipulated items as a foil. These words pairs represented the manifestation of a previously reported repetition error type (Marshall & van der Lely, 2009). Each word pair was presented twice with the target response being presented as both item A and item B. In Experiment 1, the test items were divided into four blocks of decreasing length (18, 16, 16, 6 items respectively) and presented in a randomised order which was the same for all participants. Experiment 2 controlled for order effects by presenting the test items in a novel random sequence for each participant. This test was divided into three blocks of 15, followed by a final block of 11 items.

Figure 2.1: Screenshot of NWD task, with stimulus presentation order labelled (AXB)



Source: SipSLI (Mike Coleman, UCL)

2.4.5 Nonword repetition (NWR) task

Natural tokens of the female speaker's nonword stimuli (described above) were used again in a NWR task. The participants used a headset with integral microphone and each response was recorded.

The stimuli were presented in a randomised order that was the same for all participants. There was no opportunity to repeat the recorded stimulus, however when unexpected background noise interfered with the stimulus presentation, the nonword was repeated live by the experimenter. This occurred on only a few test items in both experiments.

2.4.6 Categorical perception: Experiment 1

The ability of participants to discriminate singleton/cluster categories was assessed using three categorical perception (CP) tasks in quiet. The first continuum consisted of a single syllable nonword minimal pair of the same structure that had been used in the NWD and NWR tasks (kep-klep). In order to evaluate the effect of stimulus length on CP performance, the second continuum consisted of a three syllable nonword that had also been used in the other experimental tasks, and which contained a singleton/cluster contrast in medial position (/fəˈkepə - fəˈklepə/). Both nonwords in this minimal pair contained the single syllable nonwords used in the first CP subtest (/kep – klep/).

A final CP continuum was included so that the effect of top-down lexical knowledge could be evaluated. This continuum consisted of a single syllable real word minimal pair where the word initial segment consisted of a single consonant at one end of the continuum and a cluster at the other (/kap - klap/). This minimal pair was matched to the single syllable nonword pair, so that the stimulus pairs differed only in the medial vowel.

2.4.6.1 Categorical perception stimuli

A modified system of audio morphing was used to create the kap/klap and kep/klep continua from natural recordings of the female speaker (the kep-klep tokens being those that were used in the other experimental tasks). The morphing system takes two phonetically annotated and pitch-marked utterances and selectively transfers characteristics from a source and a target to create a new output utterance (Huckvale & Yanagisawa, 2007).

Phonetic labelling was carried out manually in the Speech Filing System computing environment (Huckvale, 2004) with each cluster labelled as a single phoneme so that that the corresponding singleton mapped directly onto the cluster when morphing was carried out. Pitch period marking was performed using an automatic tool in SFS (txanal) and hand corrected. The first stage of morphing was to perform pitch synchronous linear predictive coding (LPC) analysis on windows centred on each glottal impulse and of a size equal to two pitch periods. In voiceless regions, the analysis window size was chosen on the basis of a smooth interpolated pitch contour, so as to minimise large changes in window size from frame to frame through the utterance. The LPC coefficients were then converted to a line spectral pair (LSP) representation, which makes the coding of the spectral envelope more amenable to interpolation. The excitation residual was extracted and stored to complement the spectral information. The two utterances were then time-aligned so as to synchronise phonetic events. Alignment was performed using a dynamic programming procedure working from an MFCC spectral representation of the speech, but constrained by the phonetic annotations. This gave an accurate frame-by-frame alignment between source and target, even within individual segments. The morphing system then generated a new output by selecting and interpolating pitch, timing and spectral characteristics from the two input utterances. An output was generated between the source and model (the endpoints of the continuum) at stages specified in terms of percent different from the source/model. There were 51 stages. 25 output items were generated in two percent increments from the source and 26 were generated in the same increments from the model.

Identical procedures were used to create the /kap - klap/ and /kep – klep/ continua. However, while various attempts at manipulating and recoding the endpoints were made, the system of morphing did not provide a satisfactory continuum when applied to the /fəˈkepə - fəˈklepə/ minimal pair. The /fəˈkepə - fəˈklepə/ continuum was eventually created by embedding each morphed item on the /kep – klep/ continuum in a ‘carrier’ which consisted of the first and third syllables of the naturally recorded /fəˈkepə/ token. This system did not take into account the ‘spread’ of formant information from the outer syllables into the middle kep/klep syllable, but the resulting continuum consisted of items that contrasted with the /kap - klap/ and /kep – klep/ continua in terms of length and syllable structure.

2.4.6.2 Categorical perception task

The AXB procedure was the same as that for the nonword discrimination task. This time, participants were told that frogs A and B at the top left and right of the screen would always say the same words (the endpoints of the continuum were given), but that frog X (bottom middle) would say something that would sound similar to, but maybe not the same as, A or B. The child was instructed to click on frog A or frog B, depending on which (s)he thought sounded most similar to frog X. The inter-stimulus interval was set at 450ms.

Stimuli were presented using an interleaved adaptive procedure as described in Ramus et al. (2003). Two randomly interleaved adaptive tracks using a two up/one down procedure (Levitt, 1971) were used to concentrate trials near the 29% and 71% points on the categorisation function. The first two trials were presentations of the endpoints and as such were considered to be training items. Catch trials (continuum endpoints) were randomly interspersed during the task 10% of the time, so that participants would not hear an uninterrupted sequence of ambiguous stimuli. The task ended after three reversals on each track or after a maximum of 30 trials.

2.4.7 Categorical perception: Experiment 2

In response to the findings of Experiment 1 (see Chapter 3), that variation in performance on the real word and three syllable tasks seemed to be affected by artefacts of the stimulus continua, the real word continuum (`kap - klap`) and the 3 syllable nonword continuum were eliminated from Experiment 2. The remaining kep-klep continuum was modified to increase the number of stimuli between the end points, because the categorisation functions obtained were sometimes too steep for their slopes to be estimated. This was achieved by repeating the morphing procedure described above using the same kep-klep endpoints but specifying morphed outputs to be generated at one percent increments. This resulted in a continuum consisting of 102 stages. Stage 1 was the source nonword from which 50 stages were morphed and 102 was the model nonword from which 50 stages were morphed.

The adaptive presentation procedure was identical to that in Experiment 1, with interspersed catch trials and the task ending after three reversals on each track or after a maximum of 30 trials.

2.5 Digit Recall supplementary study: live vs. recorded presentation of stimuli

2.5.1 Introduction

In Experiment 1, the Digit Recall subtest was administered with live presentation of stimuli. During the course of testing, an automated version of this task was scripted in DMDX. The aim was to use a fully automated assessment battery in Experiment 2, thereby eliminating any confounds introduced by subtle variations in presentation of stimuli. In order that typically developing participants from both experiments could be analysed as a single group, a supplementary study was undertaken to compare data collected from live vs. recorded presentation of the Digit Recall stimuli.

2.5.2 Participants

16 children from the cohort recruited from mainstream lower schools in Bedfordshire participated in this supplementary study. They all spoke English as a first and only language and none had a diagnosed or suspected learning difficulty or hearing loss. Information sheets and consent forms had been sent out during the recruitment phase for Experiment 1. The parents/guardians of all 16 children returned signed consent forms.

Two children were excluded from the study because they failed the hearing screen. Data were analysed from 7 boys and 7 girls with a group mean age of 81 months (6 years 9 months; range = 17 months, s.d. = 5 months).

2.5.3 Procedure & Results

The Digit Recall subtest was administered twice to each participant, with an interval of five days between the live stimuli and the recorded stimuli versions. Raw scores were converted to standard scores for the analysis.

There was a significant correlation between the standard scores for the live presentation and the standard scores for the recorded presentation ($p < 0.001$). A paired samples t-test revealed no significant difference between the two presentation versions.

Given that recorded presentation of the Digit Recall task did not affect participants' scores, the automated version of the Digit Recall task was used in the latter stages of Experiment 1 and was included in the assessment battery for Experiment 2.

3. Experiment 1:
speech processing in typical development

3.1 Introduction

Probing the nature of language impairment requires methodological approaches that address the cognitive processes underpinning surface behaviours. Evaluating the performance of typically developing (TD) children on linguistic tasks is one such methodology, and was used here to investigate the speech processing of nonwords.

The relationship between NWR and language and phonological skills has been the focus of considerable attention in the developmental literature. Better NWR accuracy has been associated with the ability to produce longer, more grammatically complex sentences in TD children (Adams & Gathercole, 1996). Correlations are also reported between NWR and phonological awareness (Bowey, 1996), and NWR and digit span tasks (Metsala, 1999).

Gathercole and colleagues have proposed that NWR is a relatively pure measure of phonological memory and that accuracy on NWR is limited by phonological storage capacity (Gathercole & Baddeley, 1990). However, the task undoubtedly relies on intact speech perception and speech production in addition to adequate phonological short-term memory (PSTM). Specifically, acoustic and segmental analysis must be carried out, then a phonological representation must be stored and maintained so that a matching articulatory motor programme can be prepared and executed (Gathercole & Baddeley, 1989). Within the psycholinguistic model of speech processing described by Szenkovits and Ramus (2005), a sub-lexical phonological representation must be generated both at the input and the output stages of NWR. The structure of the phonological representation(s) will undoubtedly be influenced by the quality of the child's perceptual analysis. Indeed, phonemic discrimination is correlated with NWR accuracy in TD children aged four and six years (Masterson, Laxon, Carnegie, Wright, & Horslen, 2005).

The relationship between vocabulary and NWR has been the primary focus of interest in this area. Every new word a child hears is unfamiliar to him/her, and its status as a lexical item will often develop through attempts at repetition. While there is no evidence that NWR is correlated with measures of expressive vocabulary in TD children (Briscoe, Bishop, & Norbury, 2001), a correlation between receptive vocabulary and nonword repetition (NWR) is widely reported. Children with higher receptive vocabulary scores perform better at NWR than children with lower receptive vocabulary scores (Bowey, 1996, 2001) and the ability to learn new words is correlated with NWR (Michas & Henry, 1994). The causal direction of this relationship is complex. In essence, before the age of five years, NWR predicts receptive vocabulary but this relationship is reversed after the age of five years when a child's vocabulary is sufficiently developed that lexical knowledge facilitates NWR (Gathercole, Willis, Emslie, & Baddeley, 1992). Gathercole et al. (1992) suggest that prior to the age of five, children rely primarily on PSTM to perform NWR.

There are two main accounts of the relationship between NWR and word learning (Gathercole, 2006). The first proposes that PSTM capacity affects the ability to repeat a nonword and to retain a new phonological form in the process of learning a new word. In support of this account, NWR is correlated with more conventional measures of short-term storage such as digit span (Gathercole, Frankish, et al., 1999). Stimulus length also affects NWR accuracy, with longer nonwords being repeated less accurately than shorter nonwords (Vance, et al., 2005).

According to the second account of the relationship between NWR and word learning, increases in vocabulary knowledge affect the detail, structure, connection strengths or autonomy of lexical representations in long-term memory. In turn, lexical representations exert a top-down influence to facilitate NWR and vocabulary acquisition (Bowey, 1996, 2001). A number of findings support the theory that NWR is supported by long-term lexical knowledge. Firstly, nonwords that have been rated higher on adult ratings of wordlikeness are repeated more accurately by young TD children than nonwords that are rated low in wordlikeness (Gathercole, 1995). Secondly, nonwords with stressed syllables corresponding to real words are repeated more accurately by TD children (Dollaghan, et al., 1993; Dollaghan, et al., 1995) and real word syllables contained in nonwords are repeated more accurately than nonword syllables in stimuli of four (but not two or three) syllables (Metsala & Chisholm, 2010). Thirdly, TD children repeat nonwords more accurately if the stimuli contain sequences high versus low in phonotactic probability (Edwards, et al., 2004). Finally, syllables from dense versus sparse lexical neighbourhoods are repeated more accurately when constituting part of three or four syllable nonwords (Metsala & Chisholm, 2010).

Irrespective of the theoretical differences between these accounts, both are consistent in their reliance on phonological representations (Munson, 2006). Within the lexicality model, either lexical knowledge influences perceptual analysis and the construction of phonological representations to be stored in PSTM, or nonword stimuli partially overlap existing lexical phonological representations at input, and these representations are used to fill the gaps when incomplete output phonological representations are retrieved from PSTM (Gathercole, Frankish, et al., 1999). Within the phonological store model, variation in storage capacity between individuals may result from differences in perceptual encoding or in the stability of the phonological representations over time.

3.1.1 Aims of the experiment

The experiment presented here investigated the speech processes involved in nonword repetition (NWR). Typically developing (TD) children carried out a NWR task and additional experimental tasks which were designed to tap processing skills implicated in NWR: nonword discrimination (NWD) and three categorical perception (CP) tasks. A full battery of published language assessments was carried out to evaluate relationships between the experimental tasks and language skills.

The NWR and NWD tasks used identical natural speech nonword stimuli. There were three CP continua: a single syllable nonword minimal pair (/kep – klep/), a single syllable real word minimal pair (kap - klap) and a three syllable nonword minimal pair (/fəˈkepə - fəˈklepə/). The real and nonword CP minimal pairs were matched for length and form (they differed only in word-medial vowel). Each of the three syllable nonwords contained one of single syllable nonwords as a constituent syllable. A detailed description of all the stimuli is given in Chapter 2.

The phonological storage account of NWR was explored by using nonword stimuli consisting of both one and three syllables in all tasks. It was predicted that participants would repeat the longer nonwords less accurately than the shorter ones. A similar effect was predicted for NWD, and for the one versus three syllable CP tasks. The length of the nonwords was also affected by the number of phonemes each contained. It was predicted that NWR accuracy would decrease as the number of clusters (hence phonemes) in the nonword increased.

The lexicality account of NWR was explored indirectly by presenting real and nonword CP continua. It was predicted that if top-down lexical knowledge was influencing performance on this task, participants would categorise the real word stimuli (/kap - klap/) more consistently than the single syllable nonword stimuli (/kep – klep/).

The aims of this study were threefold. Firstly, to evaluate and compare the performance of typically developing children on three phonological measures: nonword repetition (NWR), nonword discrimination (NWD) and categorical perception (CP). Secondly, to compare data from the experimental tasks with that from published language measures. The final objective was to evaluate the experimental tasks and the battery of published assessments in preparation for testing children with SLI and SLI+SSD.

Some additional predictions were made about task performance. These are listed below:

- Performance in all experimental tasks would improve significantly with age;
- There would be a positive correlation between performance on the NWR and performance on the measure of receptive vocabulary (BPVS);
- There would be positive correlations between performance on all experimental tasks and performance on measures of phonological awareness (Rhyme, Spoonerism, Naming Speed) and phonological memory (Digit Recall);
- There would be a positive correlation between performance on the NWR task and performance on the NWD task, reflecting shared underlying processes.

3.2 Method

3.2.1 Participants

The participants were recruited from three mainstream lower schools in Bedfordshire and a mainstream primary school in South London. Each lower school was asked to invite children from reception to year four to participate in the study (children at these schools leave for middle school at the end of year four). In order to recruit older children to the study, year five children were invited from the London primary school. To be invited, all children had to meet three basic criteria:

- Speak English as a first and only language;
- No diagnosed or suspected learning difficulty (including, but not limited to, specific language impairments, ASD, ADHD, dyslexia);
- No diagnosed or suspected hearing loss.

Information sheets and consent forms were sent out to the parents/guardians of appropriate children in accordance with UCL Research Ethics Committee approval. 63 children returned consent forms signed by a parent/guardian. Of this number, 43 took part in the study described here and 16 took part in a supplementary study (comparing performance on live vs. recorded presentation of stimuli in the Digit Recall task, described in Chapter 2). The remaining four children were unavailable at the time of testing.

3.2.2 Exclusion criteria

Twelve of the 43 participants were excluded at an early stage of the project, before they had completed the task battery: seven failed the hearing screen, one had a diagnosis of ASD that had not been reported at the invitation stage and one elected not to participate. Three children achieved standardised scores for verbal comprehension that were below 1s.d. from the mean and were eliminated before it was decided that all children meeting the basic criteria and passing the hearing screen would complete the testing. Finally, data from a further participant were excluded from analysis because this participant failed the speech screen.

In summary, data from 15 girls and 15 boys were collected and analysed. These children were aged between 5;0 and 9;11 with a group mean age of 7;3.

3.2.3 Procedure and test battery

The tests were carried out in the schools. Each child was tested individually outside the classroom in a location that was as quiet and distraction-free as the school environments allowed. Where time permitted, the published tests were carried out in the same order with each child, although it was sometimes necessary to prioritise task completion over order of presentation. The experimental tasks were always presented in the same order and with an inter-task interval of at least 24 hours.

The battery of screens and assessments included the following tests (for details of each, please see Chapter 2):

- Hearing screen;
- Computer and mouse familiarisation task;
- Non-verbal cognition: Block Design subtest of the WISC;
- Working memory: Digit Recall subtest of the WMTB-C;
- Phonological awareness: Rhyme, Spoonerism and Naming Speed[3] subtests of the PhAB;
- Speech sound production: diagnostic screen of the DEAP;
- Sentence comprehension: TROG-II;
- Single word comprehension: short form of the BPVS-II;
- Expressive vocabulary: RWFT;
- Expressive language: RAPT;
- Sentence repetition: Recalling Sentences subtest of the CELF-3 UK.

Three experimental tasks were carried out (as described in Chapter 2):

- Nonword discrimination (NWD);
- Nonword repetition (NWR);

---

[3] The four youngest children had difficulty naming digits in an untimed trial, so they only completed the picture naming speed test. A mean Naming Speed score was calculated for all other participants.

- Categorical perception (CP) with /kap-klap/, /kep-klep/ and /fəˈkepə-fəˈklepə/ continua.

The experimental tasks and some of the standardised assessments were administered using a laptop and headphones/headset with integrated microphone. Participants' verbal responses were recorded. For full equipment specifications, please refer to Chapter 2.

3.3 Results

3.3.1 Standardised assessments

Three assessments yielded standard scores (population mean = 100, population standard deviation = 15) for all participants. A summary of these results is presented in Table 3.1.

Table 3.1: Standard scores on published measures

	N	Mean	Std Deviation	Range
TROG	30	102	11	81-127
BPVS	30	111	9	98-131
Digit Recall	30	104	12	83-132

The standard scores for these assessments were converted to z-scores so that they could be compared with the z-scores derived from the other tasks. This conversion was achieved using the following equation: (standard score – population mean)/population standard deviation i.e. (standard score – 100)/15.

The remaining standardised assessments (Recalling Sentences, Rhyme, Spoonerisms, Rapid Picture Naming, WISC) yielded standard scores only for children over the age of six years. In order to have comparable age-normalised scores for the younger children, it was necessary to calculate adjusted z-scores for all participants. Firstly, standardised residuals for linear regressions of the raw scores on age were saved as z-scores. For each assessment, the distribution of these scores was then compared to that of the sub-set of standard scores that was available for older participants. Two-tailed paired-samples t-tests showed that for the Rhyme and Spoonerism tasks, the distributions of age-normalised scores from the whole population differed significantly from the distribution of the sub-set of standardised scores (Rhyme: $t = -2.653$, d.f. = 23, $p < 0.05$; Spoonerisms: $t = -3.980$, d.f. = 23, $p < 0.005$). There was no significant difference between the distributions of raw scores and the sub-set of standardised scores for the WISC, Naming Speed and Recalling Sentences tasks. For each assessment, the mean paired difference obtained from the t-test was subtracted from the z-scores to correct for any distributional difference. Finally, where an assessment was scored according to time taken (a lower score was better), the z-scores were multiplied by -1 so that positive z-scores always indicated better performance than average. One participant had not completed the WISC block test and another had not completed the Naming Speed task. These participants were assigned a z-score of zero for these tasks, zero being the population mean.

3.3.2 Non-standardised published assessments

The remaining published assessments (RWFT, RAPT Grammar and RAPT Information) did not yield standard scores. There was a significant positive correlation between raw score and age for these assessments (see Table 3.2). In order to normalise this age effect, and to allow cross-task comparisons, z-scores were calculated by saving the standardised residuals for linear regressions of the raw scores on age. One participant had not completed the RAPT assessment and was assigned a z-score of zero, being the population mean.

Table 3.2: Correlations of raw scores with age, showing values of r

	RAPT Grammar n = 29	RAPT Information n = 29	RWFT n = 30
Age	0.587**	0.637**	0.722**

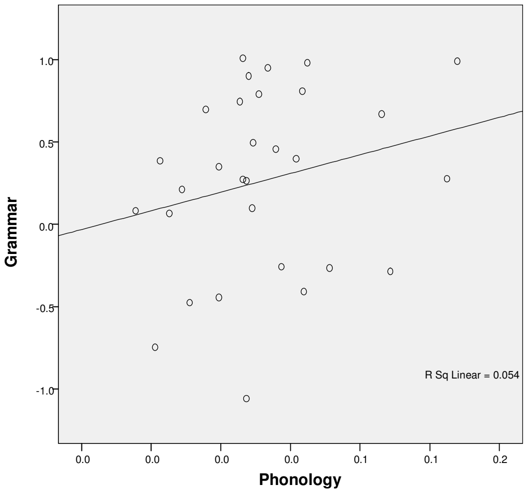
** Correlation is significant at the 0.02 level (1-tailed), Bonferroni corrected.

3.3.3 Composite scores

Three primary measures were extracted from the data for each participant: a non-verbal cognition score (z-score from the WISC block test), a language score (the mean of z-scores from the TROG, BPVS, Recalling Sentences, RWFT and RAPT Grammar), and a phonology score (the mean of z-scores from the Rhyme, Spoonerism, Naming Speed and Digit Recall tasks).

There appeared to be considerable variation in the data (see Figure 3.1), and there was no significant correlation between the grammar and phonology z-scores (r = .231, n = 30, *p* = .219, two-tailed).

Figure 3.1: Scatterplot showing the relationship between grammar and phonology composite z-scores



3.3.4 Experimental tasks

The NWR and NWD tasks yielded a total number of correct responses for each participant. These scores were converted to proportions of the total number of trials.

The output of each categorical perception task was an identification function that plots the proportion of one of the two alternative responses against the points on the stimulus continuum. The identification function for each participant was fitted using logistic regression. The slope of the identification function reflects the consistency with which the listener is categorising items on the continuum, with a steeper slope indicating higher consistency. Where a child had a fixed category boundary and consistently identified stimuli either side of this boundary, an infinite gradient was extracted. In such cases, responses at the category boundary were altered by a constant of 0.5 so that the slope of the function was not infinite. This manipulation was necessary for 10% of participants in the /kap - klap/ condition, 27.6% of participants in the kep-klep condition, and 13% of participants in the /fəˈkepə-fəˈklepə/condition. A gradient could not be calculated for one participant in the kep-klep condition. This participant gave the same response to all stimuli, and was assigned a standard score of zero, this being the population mean.

3.3.5 Age effects

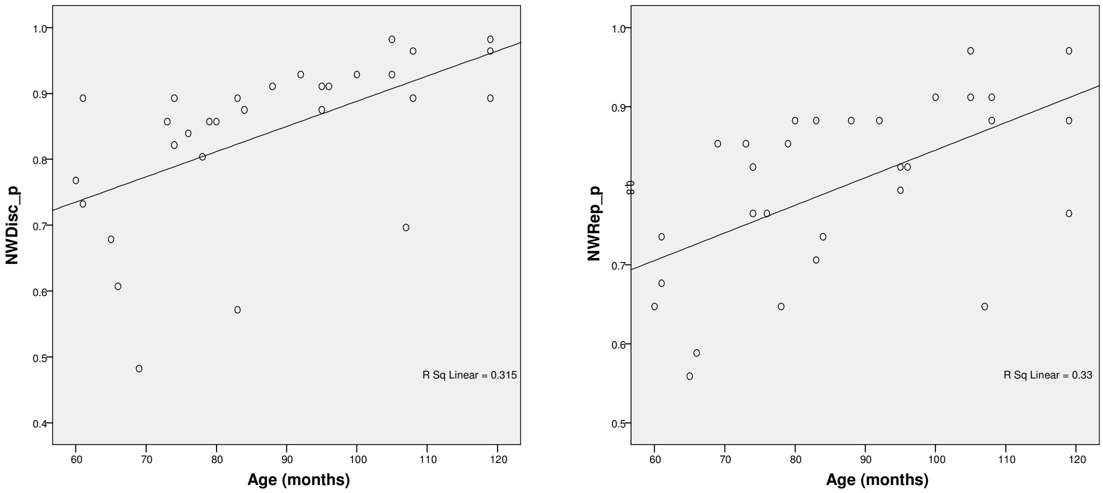
The scatterplots in Figures 3.2 and Figures 3.3 depict the relationship between age and performance in the experimental tasks. As expected, performance on all tasks improved with age, although this correlation was just below significance for the /fəˈkepə – fəˈklepə/ CP continuum (see Table 3.3).

Table 3.3: Correlations between age, NWD, NWR and CP, showing values of r (n = 30)

	NW Disc	NW Rep	kap-klap grad	kep-klep grad	fəˈkepə-fəˈklepə grad
Age	0.561*	0.574*	0.512*	0.432*	0.400

*p < .01, 1-tailed, Bonferroni corrected

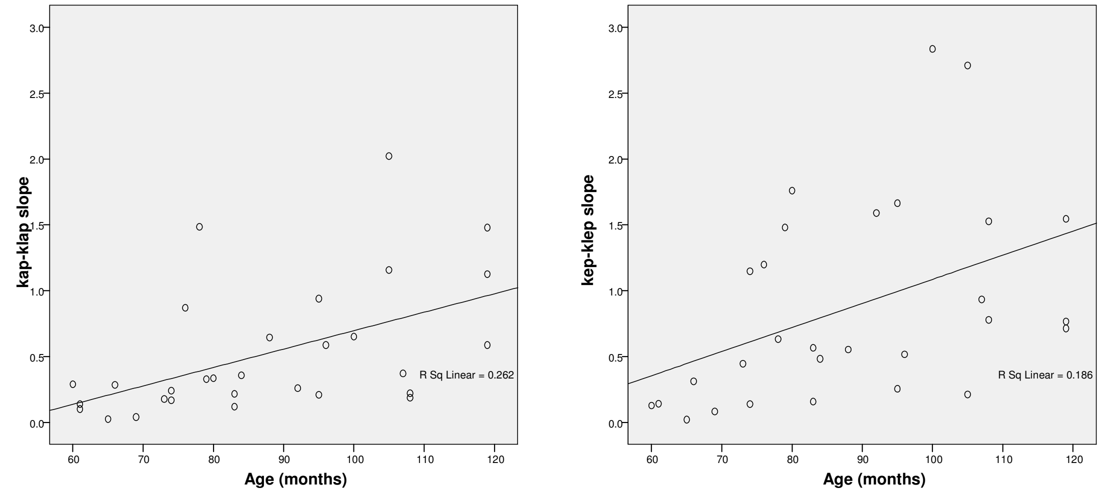
Figures 3.2a,b: The relationship between age, NWD and NWR



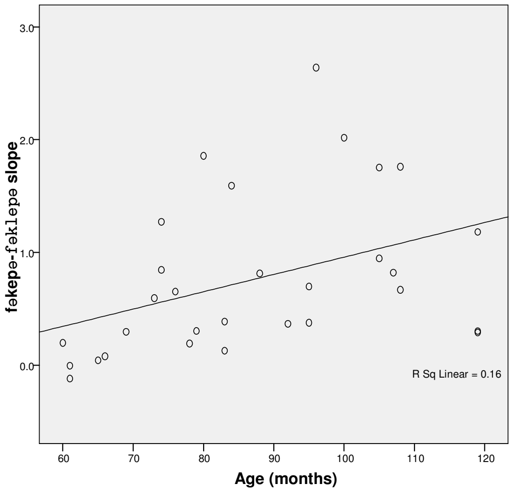
Figures 3.3a, b, c: The relationship between age and CP

a. 1 syllable real word

b. 1 syllable nonword



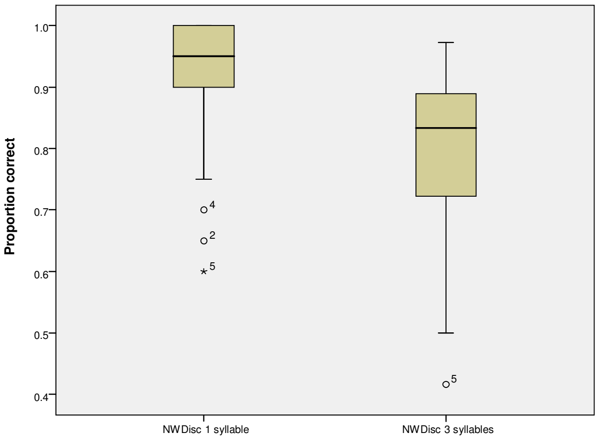
c. 3 syllable nonword



3.3.6 NWD task

Participants were better at discriminating nonwords of one syllable than nonwords of three syllables (see Figure 3.4). This performance difference was highly significant (paired samples, *t* = 8.990, d.f. = 29, *p* < 0.001, one-tailed).

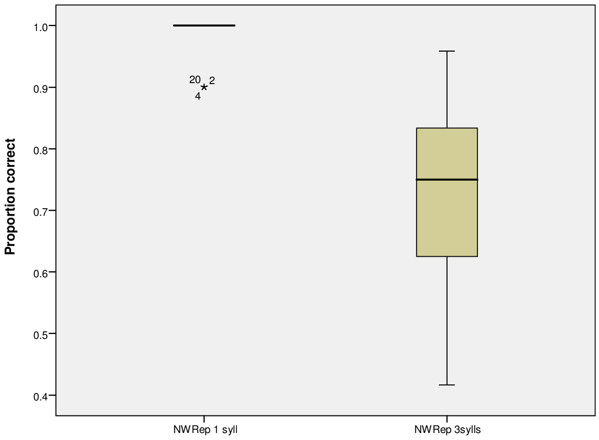
Figure 3.4: The effect of syllable number on NWD performance.



3.3.7 NWR task

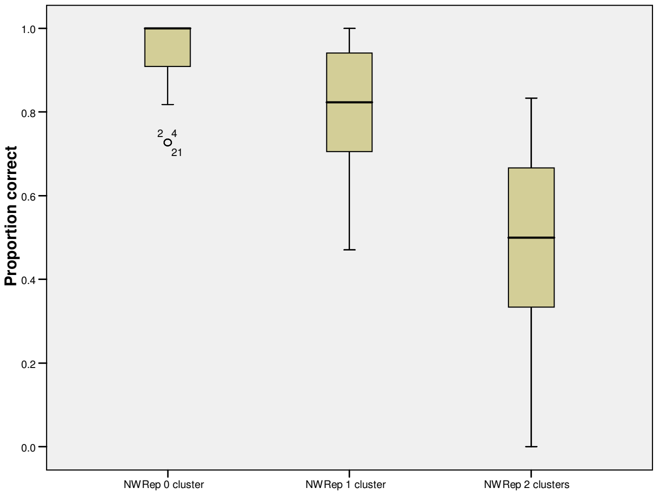
Participants were better able to repeat nonwords containing one syllable than nonwords containing three syllables (see Figure 3.5). Most participants performed at ceiling on the single syllable stimuli. A paired samples t-test confirmed that the difference in syllable structure was highly significant (*t* = 9.340, d.f. = 29, p < 0.001).

Figure 3.5: The effect of syllable number on NWR performance



Most participants performed at ceiling on one and three syllable nonwords containing no cluster (see Figure 3.6). As the number of clusters in all the nonwords increased, repetition became worse. A one-way repeated measures ANOVA showed that the effect of number of clusters was highly significant ($F_{(2,58)} = 74.855$, $p < 0.001$). Post hoc testing with Bonferroni correction revealed that performance on each cluster condition (zero, one, two) was significantly different to the others ($p < 0.05$).

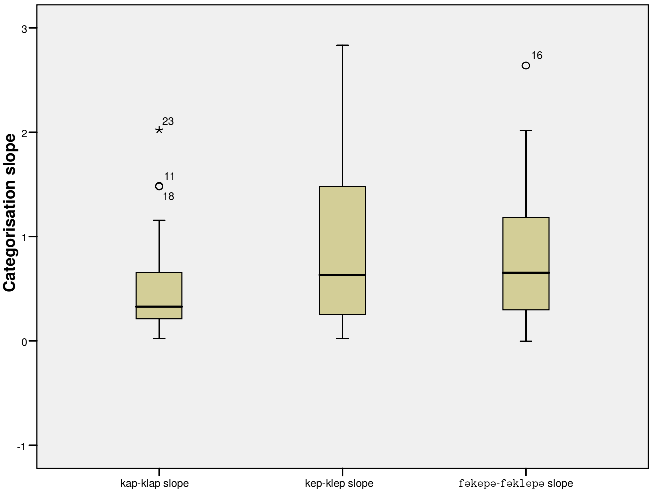
Figure 3.6: The effect of cluster number on NWR (1 and 3 syllable nonwords)



3.3.8 Categorical perception

The boxplots in Figure 3.7 depict performance on the categorical perception tasks. It appears that participants categorised `/kap - klap/` less consistently than the nonword pairs (kep-klep, `fəˈkepə - fəˈklepə`), however a repeated measures ANOVA failed to show a significant effect of continuum type ($F_{(2,56)} = 3.024$, $p = 0.057$).). Post hoc testing with Bonferroni correction revealed that there was a significant difference between the consistency of categorisation of the `kap - klap` vs. kep-klep continua ($p < 0.05$).

Figure 3.7: Performance on the three CP tasks



The relationships between the categorical perception tasks are depicted in Figure 3.8 and Figure 3.9. The scatterplots suggest there may be positive correlations between the real vs. nonword and single syllable vs. three syllable word pairs. Partial correlations using Bonferroni correction and factoring out the effects of age, showed that these correlations were not significant.

Figure 3.8: Real vs. nonword in CP task.

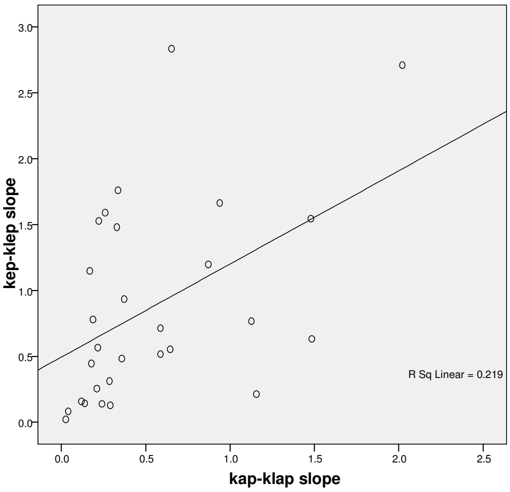
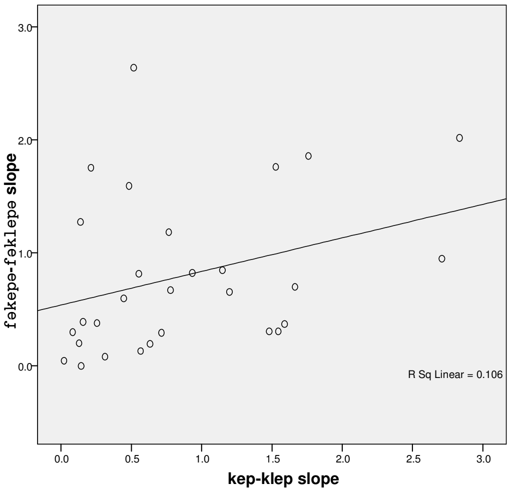


Figure 3.9: 1 vs. 3 syll nonwords in CP task.



3.3.9 Relationships between the experimental tasks and published language measures.

Measures for all experimental tasks were normalised for age in the manner described above. The participant for whom no gradient was extracted from the kep-klep categorical perception task was assigned a z-score of zero, this being the population mean. A composite categorical perception z-score was calculated for each participant by standardising the mean value of the three slopes.

Pearson's correlations were carried out with Bonferroni correction to investigate the relationships between the z-scores for the experimental tasks and those obtained from the published assessments. There was no correlation between the experimental tasks and any of the published measures of grammar (see Table 3.4).

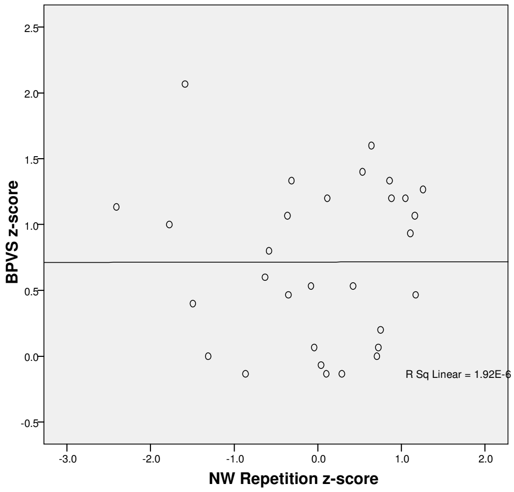
Table 3.4: Correlations between experimental tasks and measures of grammar showing values of r (n=30)

	NWD	NWR	Composite CP
TROG	0.267	0.045	0.169
BPVS	-0.048	0.001	-0.227
Recalling Sentences	0.189	0.198	-0.018
RWFT	-0.124	0.142	-0.248
RAPT Grammar	-0.022	0.185	-0.103

No correlations significant at the 0.003 level, two-tailed, Bonferroni corrected.

Of particular interest was the relationship between NWR and the test of receptive vocabulary (BPVS). The scatterplot in Figure 3.10 clearly shows that these two tasks are unrelated.

Figure 3.10: The relationship between NWR and receptive vocabulary (BPVS) z-scores



Pearson's correlations were also carried out to compare the relationships between the experimental tasks and the published measures of phonological awareness (Naming Speed, Spoonerism task, Rhyme task) and phonological memory (Digit Recall). After Bonferroni correction, no correlation was significant (see Table 3.5).

Table 3.5: Correlations between experimental tasks and measures of phonological awareness and phonological memory, showing values of r for n = 30

	NWDisc	NWRep	Composite CP
Naming Speed	0.037	-0.094	-0.115
Spoonerism	0.061	-0.006	-0.037
Rhyme	0.421	0.327	0.099
Digit Recall	0.083	0.210	0.322

No correlations significant at the 0.004 level, one-tailed, Bonferroni corrected.

Finally, Pearson's correlations were carried out to investigate the relationships between the experimental tasks and the composite performance measures (non-verbal cognition, phonology and grammar). Using Bonferroni correction there were no significant correlations.

Table 3.6: Relationships between experimental tasks and non-verbal and language measures, showing values of r for n = 30

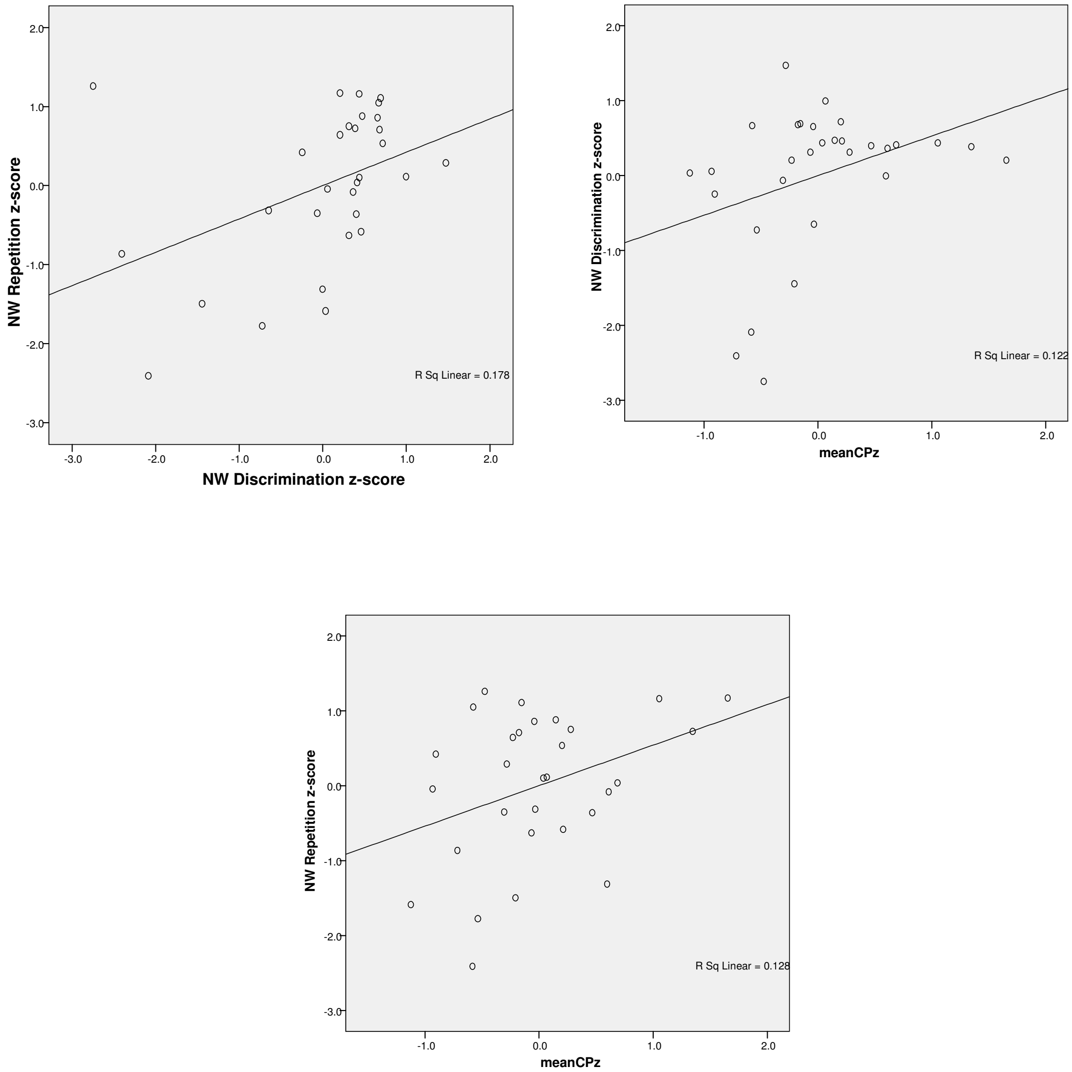
	NWD	NWR	Composite CP
WISC	-0.004	-0.158	0.189
Phonology	0.253	0.174	0.092
Grammar	0.077	0.196	-0.133

No correlations significant at the 0.006 level, one-tailed, Bonferroni corrected.

3.3.10 Relationships between experimental task

The scatterplots in Figure 3.11 depict the relationships between the experimental tasks.

Figure 3.11: Relationships between experimental tasks



Pearson’s correlations were carried out to investigate these relationships. Using Bonferroni correction, no correlations were significant (see Table 3.7).

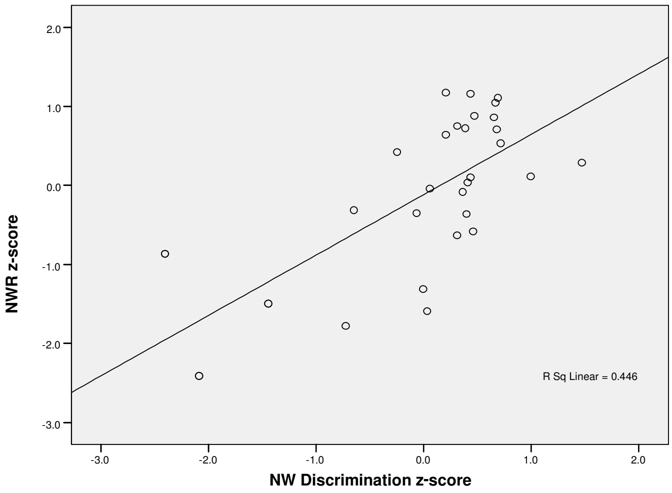
Table 3.7: Correlations between experimental tasks showing values of r for n = 30

	NWD	NWR	Composite CP
NWD	1	0.422	0.350
NWR	0.422	1	0.358
Composite CP	0.350	0.358	1

No correlation is significant at the 0.008 level (one-tailed, Bonferroni corrected).

A closer look at the scatterplot of NWR plotted against NWD (see Figure 3.11, top) revealed that the participant who had performed best at NWR had also performed worst at NWD. This participant's data appeared to be skewing the results. The data from this participant were excluded and the positive correlation between NWR and NWD was found to be highly significant (r = 0.668, n = 29, *p* < 0.001).

Figure 3.12: The relationship between NWR and NWD (after the exclusion of 1 participant)



3.4 Discussion

The aim of this study was to investigate the performance of typically developing children on a series of published language measures and speech processing tasks.

Performance on the experimental tasks improved with age. The tasks clearly tapped phonological skills that develop between five and ten years of age. In support of the phonological storage account of nonword repetition (NWR), NWR accuracy was best for the single syllable nonwords. This finding extends that of Vance et al. (2005) for an older age range of children (5;0 to 9;11 versus Vance et al.'s three to seven years).

Nonword discrimination (NWD) and NWR were highly correlated, supporting the claim that NWD taps a stage of phonological processing that is also involved in NWR. NWD accuracy was better for one versus three syllable nonwords, suggesting phonological short-term memory (PSTM) capacity was critical to performance. However, participants were reminded not to repeat the NWD stimuli, but some were observed silently rehearsing the occasional nonword. Even when repetition was suppressed, it is possible that covert rehearsal was being carried out. This would indicate that output processing systems may have been involved in performing the NWD task and that PSTM capacity may not have been entirely responsible for poorer discrimination of the longest stimuli.

The categorical perception (CP) data were somewhat anomalous. There was no correlation between the mean CP score and the other experimental tasks, a finding that cannot be readily explained. It seems unlikely that CP measured skills that were not also involved in NWD and NWR. Stimulus length was not significant to CP, which suggests that PSTM capacity did not affect performance. Participants performed equally well on the single syllable vs. three syllable nonword continua (/kep – klep/ or /kap - klap/ versus /fəˈkepə - fəˈklepə/). It should be noted that the morphed segments of both nonword continua were acoustically identical, with each of the three syllable stimuli consisting of an embedded version of /kep/ or /klep/. It seems that if the extraction of salient features from the carrier syllables does result in an increased processing load, the increase is not relevant to a task of this nature.

There was a difference between CP of the single syllable real words and the nonwords of the same length. Surprisingly, the real words were categorised less consistently than the nonwords. Why were participants poorest at categorising the real word minimal pair (/kap - klap/)? It may be that top-down influence of existing lexical knowledge did not influence categorisation of the real word stimuli, so that only sub-lexical processes were involved in task performance. Alternatively, participants may have been able to create adequate or better phonological representations for the acoustic properties of the nonword items, thereby negating any lexical advantage that the real words may have had. Either way, it seems likely that the pattern of performance was due to an artefact of the /kap - klap/ stimulus continuum.

The NWR cluster data seem to support a phonological storage account of NWR. NWR accuracy decreased as the number of clusters (hence number of phonemes) increased. Moreover, while the nature of the phonological representation(s) to be created, stored and retrieved is relevant to NWR accuracy in children with SLI and dyslexia, the location and stress of a cluster do not affect typically developing (TD) children (Marshall & van der Lely, 2009). Again, phonological storage capacity rather than the processing of complex stimuli is implicated with respect to cluster number in NWR. However when the experimental tasks are compared with published measures of language and phonology, the picture changes.

None of the experimental tasks was correlated with the published assessments, even the measures of phonological awareness and PSTM. The nonword stimuli used here were devised systematically and targeted very specific phonological structures. Unlike the stimuli of Marshall and van der Lely (2009) some of the NWD and NWR stimuli presented here contained two clusters. It is plausible that adding a second consonant cluster significantly increased the processing load and that this factor, rather than PSTM, accounted for the drop in performance on the three syllable NWD and NWR stimuli. It should be noted that the correlation between Rhyme and NWD fell slightly short of significance, so perhaps these tasks do tap a similar sub-set of phonological skills.

Of particular interest was the failure to replicate the correlation between NWR and receptive vocabulary. This suggests that the nonword stimuli used here did not activate existing whole or partial lexical representations. The results confirm that although the same nonwords were used in all experimental tasks, no lexicalisation of the nonwords took place over the course of testing. There was at least a period of 24 hours between each experimental task, and this seemed to prevent any word-learning taking place.

The structure of the stimuli also guarded against top-down lexical influences on NWR performance. Real word syllables, especially those from dense lexical neighbourhoods, that are contained in nonwords are repeated more accurately than nonword syllables in stimuli of four syllables (Metsala & Chisholm, 2010). Although the nonwords used here were phonotactically viable in English, they contained no real words. The wordlikeness of the stimuli was not formally evaluated, but the lack of correlation between NWR and receptive vocabulary suggests that participants did not consider the nonwords to resemble real words.

Given that NWR was so clearly unrelated to receptive vocabulary and as NWD was implicated in NWR, it may be unsurprising that NWD and NWR were unrelated to the measure of PSTM. The stimuli in the digit recall task were familiar lexical items for the participants and it is reasonable to assume that the application of long-term knowledge could not be eliminated from their performance. It is possible that top-down lexical influences account for some of the shared variability in studies that have found a correlation between PSTM and NWR. There was no evidence that top-down processing affected performance on the NWR task used here, which may explain the lack of correlation with PSTM.

In summary, the results indicate that the NWD, NWR and CP tasks measure phonological skills that develop in childhood. The tasks did not appear to index receptive vocabulary or PSTM. It is proposed that these tasks isolate phonological representations at an unspecified sub-lexical stage of speech processing.

4. Experiment 2:
speech processing in SLI and SLI+SSD

4.1 Introduction

This experiment was carried out to investigate the phonological processing skills of children with SLI and SLI+SSD.

Experiment 1 investigated the performance of typically developing (TD) children on three experimental tasks: nonword repetition (NWR), nonword discrimination (NWD) and categorical perception (CP). The results showed that NWD and NWR were highly correlated, implicating the same phonological processes in performance accuracy. None of the tasks was related to published measures of language and phonology. Of particular interest was the failure to find a correlation between NWR and receptive vocabulary or between any task and a measure of phonological short term memory (PSTM). It seemed that participants did not rely on lexical knowledge or PSTM when performing the experimental tasks. Instead it seemed that the tasks tapped phonological representations at an unspecified stage of sub-lexical speech processing.

These findings motivated the current experiment in three main ways:

1. The stimuli used in the experimental tasks seemed to tap phonological processing at a sub-lexical level that was relatively independent of PSTM capacity. Given that children with SLI have impaired lexical knowledge (e.g. McGregor, Newman, Reilly, & Capone, 2002)and impaired PSTM (e.g. Archibald & Gathercole, 2006b), it was predicted that CP, NWD and NWR would elicit data from the clinical groups that was not significantly confounded by either deficit. In this way, the locus of any phonological impairment could be identified.
2. The nature of any transfer of information between input and output stages of speech processing could be explored.
3. If the NWR and NWD tasks are similarly correlated for children with speech and language impairments, a NWD task using stimuli of the kind used in Experiment 1 may be an alternative to NWR as a candidate marker for SLI. NWD does not require a verbal response so would not be subject to the same scoring issues as verb morphology, sentence recall or NWR (i.e. distinguishing speech production errors from linguistic errors). If NWD could

replicate the diagnostic power of NWR, it could be used effectively to identify language deficits in children with SSD.

Although the population of children with concomitant SLI and SSD is small (see Chapter 1), appropriate identification of these children is important. If an underlying language deficit is not identified, a speech sound disorder may be wrongly characterised and treatment may be ineffective. Furthermore, children who present with both disorders rather than SSD alone may be at increased risk for other developmental disorders such as attention deficit disorder (McGrath, et al., 2008) or reading and spelling difficulties (e.g. Lewis, et al., 2000; Peterson, et al., 2009).

Assessment of expressive language skills in children with moderate to severe SSD presents a clinical challenge, especially with respect to differentiating speech sound versus morphological errors. This is particularly significant when one considers that two of the leading candidates for an SLI diagnostic marker are verb morphology (Rice & Wexler, 1996) and sentence recall (Conti-Ramsden, et al., 2001). Both measures require precise articulation of morphological suffixes that children with SSD are likely to find difficult. The third potential marker for SLI, NWR, is usually scored such that any phoneme substitution, deletion or insertion is scored as an error. A child with SSD is likely to score poorly on any or all of these measures, regardless of whether (s)he has an underlying language impairment.

In order to address the potential of NWD as a diagnostic tool, it is first necessary to consider the range of deficits that are associated with this and NWR in children with SLI or SLI+SSD. A phonological deficit has been proposed as a core underlying cause of both SLI and SSD, but its precise manifestation will be different depending on the presence and severity of other deficits (Pennington & Bishop, 2009). Direct comparison between SLI and SLI+SSD allows us to ask whether the pattern of underlying impairment is the same in both disorders and addresses the cognitive processes that underpin surface behaviours. Is the speech impairment in SLI+SSD simply additional to the language impairment, or do these children have a more severe or qualitatively different cognitive deficit when compared to children with SLI only?

NWR has been proposed as a relatively pure measure of a PSTM deficit in SLI (Bishop, et al., 1996). However, as the results from Experiment 1 showed, NWR is a complex task which also engages cognitive processes such as speech discrimination and production. In theory, impairment in any of these processes may lead to poor NWR. Some researchers also believe that SLI is characterised by an underlying auditory processing deficit which would also affect NWR, but this debate will not be addressed here (for a review, see Rosen, 2003).

Comparing performance on NWD with that on a task such as CP potentially yields information about the speech processing skills implicated in NWD. Phonological categorisation is a fundamental input process whereby a listener categorises speech sounds according to the phonological system of the language. A deficit at this stage of processing could result in the formation of incorrect phonological representations for nonwords in a discrimination or repetition task, and may account for a deficit on either task. There is evidence that the categorical perception of children with SLI is impaired for synthesised speech stimuli (e.g. Thibodeau & Sussman, 1979), but not natural speech stimuli (Coady, et al., 2005). The stimuli used in the experiment reported here are natural, so the expectation was that the SLI participants would perform comparably to TD peers.

CP in children with SLI+SSD has been investigated in several studies. Sussman (1993) found that children with a speech and language impairment performed similarly to younger, typically developing controls on an identification task (pointing to the relevant letter in response to stimuli on a /ba/-/da/ continuum). The clinical participants were not impaired on a discrimination task using the same stimuli. While the former task most closely resembles the paradigm used in the current experiment, Sussman's stimuli were synthesised which may account for the deficit she observed in the SLI+SSD participants (see above). Assuming stimulus type was not significant, the SLI+SSD group in the current experiment was predicted to perform worse than TD participants on the categorical perception task.

There is mixed evidence that poor NWR in children with SLI is due to poor speech discrimination. Gathercole and Baddeley (1990) reported that children with SLI were able to discriminate between single syllable nonwords differing on one phonetic feature. Montgomery

(1995) found that children with SLI were impaired on discrimination of nonwords that differed by one phoneme, but this effect was only observed in four syllable stimuli, implying that nonword length affected performance. Marton and Schwartz (2003) failed to replicate this finding, showing instead that language impaired children performed comparably to TD controls on NWD. The authors suggested that methodological differences may explain the conflicting results. In particular, Marton and Schwartz noted that their minimal pair nonword stimuli differed only in stress pattern, so that participants may have used multiple phonetic cues to enhance discrimination.

NWR requires the planning and execution of a sequence of articulatory gestures to produce a spoken output which corresponds to a retrieved phonological representation (Gathercole, Service, Hitch, Adams, & Martin, 1999). This process presents a considerable challenge to a child’s processing system, especially as articulation accuracy is of the utmost importance. In most NWR tasks, published or informal, any phoneme substitution, deletion or insertion is scored as an error. This contrasts to measures of PSTM such as digit recall, where phoneme errors are acceptable as long as the child has produced an unambiguous version of the target digit. Experimental evidence has shown that NWR scores correlate with articulation rates in typically developing four-year-old children (Gathercole, Service, et al., 1999). Language-impaired children do not differ from controls in the onset or rate of their articulation (Gathercole & Baddeley, 1990), despite performing worse on a NWR task. However, children with SSD do have slower articulation rates than TD controls (e.g. Raine, et al., 1991). The effects of slower articulation rates and phoneme errors could mitigate against successful NWR in SSD.

NWR combines the different processing demands detailed above. It is difficult to state with any certainty which process or processes are impaired unless each is measured independently. Most studies that have compared NWR with its underlying input processes in children have looked at a single process (e.g. CP or NWD) and compared findings to previous studies that have used entirely different NWR stimuli, and a different cohort of participants. Alternatively, input processes have been investigated alongside NWR using different speech materials (often including real words) to evaluate each processing pathway (Gathercole & Baddeley, 1990; Marton & Schwartz, 2003).

The primary aim of this experiment was to investigate possible causal relationships between categorical perception (CP), nonword discrimination (NWD) and NWR in children with SLI and SLI+SSD, with a view to identifying the locus of any processing deficit. The experiment used identical natural speech nonwords for the NWD and NWR tasks and a continuum derived from two of the nonword stimuli for CP to eliminate any effects of stimulus type. This is a paradigm that was successfully used by Szenkovits and Ramus to pinpoint impaired input phonological processing as the cause of poor performance on NWD and NWR (Szenkovits & Ramus, 2005). The secondary aim of the experiment was to evaluate these tasks as possible markers of language impairment. To that end, the relationships between the experimental tasks and language, phonological and reading composites were investigated. The diagnostic power of the experimental tasks was compared to that of sentence recall.

Finally, top-down processes have been implicated in NWR. There is a body of research that finds receptive vocabulary predicts NWR accuracy in typically developing (TD) children. This suggests that TD children use lexical knowledge to aid performance on NWR (for a review, see Gathercole, 2006). In children with SLI, the relationship is somewhat different, with NWR instead predicting receptive vocabulary (Coady & Evans, 2008). Experiment 1 (reported in Chapter 3) failed to replicate the relationship between receptive vocabulary and NWR in TD participants, which suggested that the lexicon was not involved in top-down processing of the stimuli. It is anticipated, then, that NWR will not predict receptive vocabulary in the SLI or SLI+SSD groups.

4.1.1 Aims of the experiment

This experiment was carried out to investigate the speech processing of children with SLI and SLI+SSD versus typically developing (TD) children. Specifically, NWR accuracy was compared to accuracy on tasks implicated in NWR, namely NWD and CP, with the aim of identifying the locus of any processing deficit. A full battery of published linguistic assessments was also carried out in order to characterise the groups and evaluate the reliability of the experimental tasks in identifying language impairment.

Specifically, the following issues were addressed:

- An age effect on experimental task performance was observed in the Experiment 1 TD participants. It was predicted that this effect would be replicated in the TD group in this experiment;
- NWD was correlated with NWR in the first experiment, and potentially taps the same skills as NWR. If NWR is considered to be a marker for SLI, could NWD be used as an equivalent marker that eliminates the confounds of speech difficulties? Speech output deficits will affect NWR accuracy but such errors may not represent an underlying linguistic deficit. On the other hand, impaired speech production may mask an underlying or concomitant language deficit if NWR errors are attributed to a 'pure' speech deficit.
- CP and NWD are both implicated in NWR. Children with SLI and those with SLI+SSD have difficulty with NWR. Does the language deficit in SLI+SSD mitigate further against NWR accuracy or can repetition errors be explained solely in terms of the speech impairment? In other words, do children with SLI and those with SLI+SSD perform similarly on input processing tasks (CP and NWD)?
- Processing abstract stimuli (nonwords or real words with low concreteness and imageability for children) may be more difficult for children with SLI (Coady, et al., 2005) in CP tasks. This may be due to fragile underlying phonological representations or to the cumulative effects of synthetic speech stimuli and high memory demands. The CP and NWD tasks presented here were designed to minimise memory load using an AXB paradigm and they used natural speech stimuli. In this way, the phonological representations of the clinical groups could be targeted for evaluation.

4.2 Method

4.2.1 Participants

27 children with SLI or SLI with concomitant SSD were recruited from five specialist language provisions located in the south east of the UK. Four of the language provisions were attached to

mainstream schools and one was a specialist school for children with communication difficulties. Schools were given a list of inclusion criteria before identifying suitable participants. The criteria were as follows:

- Aged between 5;0 and 10;0;
- Speak English as a first and only language;
- No known hearing loss;
- No known neurological disorder;
- No known orofacial abnormality;
- No additional learning difficulty e.g. ASD, dyslexia;
- Non-verbal cognition within normal limits;
- Receptive language skills at 5 year level or above (this was lowered to a TROG age equivalent of 4 years after testing commenced);
- Has statement of special educational needs that identifies spoken language difficulties and/or speech difficulties or has been identified as requiring support through School Action or School Action Plus for spoken language and/or speech difficulties.

Participants were screened for hearing difficulties, below average non-verbal cognition (scaled score of 1s.d. below the mean on the WISC or WPPSI) and receptive language difficulties in case they had been inappropriately referred to the project. Thirteen of the children recruited from language provisions were eliminated after the first testing session: four failed the hearing screen, seven did not achieve a standard score above one standard deviation below the mean on the WISC/WPPSI subtest, one participant elected not to participate and another was not available to complete the assessment battery.

Two clinical subgroups were formed for analysis on the basis of the results of the DEAP screen. One group included participants with a language impairment and age appropriate speech skills (SLI). The other group included participants with a language impairment and a concomitant speech difficulty (SLI+SSD).

26 typically developing (TD) participants were recruited from the mainstream classes of the schools with specialist language provisions, from an additional mainstream lower school and

from two nurseries in the south east of the UK. To be invited to participate, all children had to meet three basic criteria:

- Aged 4;0 – 10;0;
- Speak English as a first and only language;
- No diagnosed or suspected learning difficulty (including, but not limited to, specific speech/language impairments, ASD, ADHD, dyslexia);
- No diagnosed or suspected hearing loss.

TD participants were also screened for hearing difficulties, below average non-verbal cognition and receptive language difficulties using the criteria applied to the clinical groups. Five of the TD participants were excluded before they had completed the task battery: three failed the hearing screen, one did not achieve an age equivalent score of 4;0 for receptive language (TROG), and another elected not to complete the test battery. 21 TD children took part in this experiment. Their data were combined with data from the TD participants from Experiment 1 to form a TD control group consisting of 51 participants aged 4;0 to 9;11 (group mean age = 6;11, s.d. = 1;6).

Information sheets and consent forms were sent out to the parents/guardians of all children in accordance with University College London Research Ethics Committee approval. Children who returned signed consent forms were given simplified information about what to expect from testing. Each participant was given the opportunity to withdraw from the research at any time.

4.2.2 Procedure and test battery

The tests were carried out in the schools. Each child was tested individually outside the classroom in a location that was as quiet and distraction-free as the school environments allowed. Where time permitted, the published/informal tests were carried out in the same order with each child, although it was sometimes necessary to prioritise task completion over order of presentation. The experimental tasks were always presented in the same order and with an inter-task interval of at least 24 hours.

The battery of published and informal assessments consisted of the following (for details of each test, please see Chapter 2):

- Hearing screen;
- Computer and mouse familiarisation task;
- Informal reading screen;
- Reading: nonword reading subtest of the PhAB;
- Non-verbal cognition: block design subtest of the WISC or WPPSI;
- Working memory: digit recall and word recall subtests of the WMTB-C;
- Phonological awareness: Rhyme and Alliteration with pictures subtests of the PhAB;
- Speech sound production: diagnostic screen of the DEAP;
- Sentence comprehension: TROG-2;
- Single word comprehension: short form of the BPVS-II;
- Expressive vocabulary: RWFT;
- Expressive language: RAPT (grammar scores only);
- Sentence repetition: Recalling Sentences subtest of the CELF-4UK.

This battery of assessments was not the same as that administered for Experiment 1. Firstly, the Spoonerisms and Naming Speed sub-tests were eliminated from the earlier battery. Many participants in Experiment 1 did not engage with the Spoonerisms task and it was often difficult to motivate them to complete it. It was anticipated that there would be additional difficulties motivating the youngest participants and those with language and/or speech difficulties in the current experiment. It was also anticipated that valid scoring of the Naming Speed subtest would be particularly difficult in the case of participants with SSD.

Three tasks were introduced into the battery: the Alliteration with pictures and Nonword Reading subtests of the PhAB and an informal reading screen which assessed letter knowledge and reading of real word and nonword digraphs. None of the participants had a diagnosis of specific literacy difficulties but the reading measures were introduced to further characterise the groups. There is evidence that children with a language impairment in the absence of a reading impairment are unimpaired on phonological processing tasks including nonword repetition (Bishop, et al., 2009; Catts, et al., 2005). The Alliteration task was included to evaluate the phonological awareness of language impaired children whose speech was insufficiently intelligible for reliable scoring on the Rhyme task.

Three experimental tasks were carried out (as described in Chapter 2):

- Nonword discrimination (NWD);
- Nonword repetition (NWR);
- Categorical perception (CP) with kep-klep continuum. That the TD participants in Experiment 1 did not differ in their performance on the one versus three syllable nonword continua or the real word versus nonword continua was likely due to artefacts of the /kap - klap/ and /fəˈkepə - fəˈklepə/ continua. These continua were eliminated from the battery for Experiment 2.

The experimental tasks and some of the standardised assessments were administered using a laptop and headphones/headset with integrated microphone. Participants' verbal responses were recorded. For full equipment specifications, please refer to Chapter 2.

4.2.3 Statistical methods

PASW Statistics 18 software was used for all analyses. Normal distribution and homogeneity of variance were considered for all variables before further analyses were undertaken. Boxplots, histograms and normal Q-Q plots were visually examined to assess distribution and Levene's test or Mauchly's Test of Sphericity was used to confirm homogeneity of variance as appropriate. Unless otherwise stated, the assumptions of normal distribution and homogeneity of variance were met and parametric analyses were carried out. Performance on all tasks was predicted to improve with age, so one-tailed significance is reported for such correlations. Where parametric post-hoc pair-wise comparisons were carried out, the Bonferroni-corrected statistic is reported. The statistics of t-tests are reported for equal variances not assumed where appropriate.

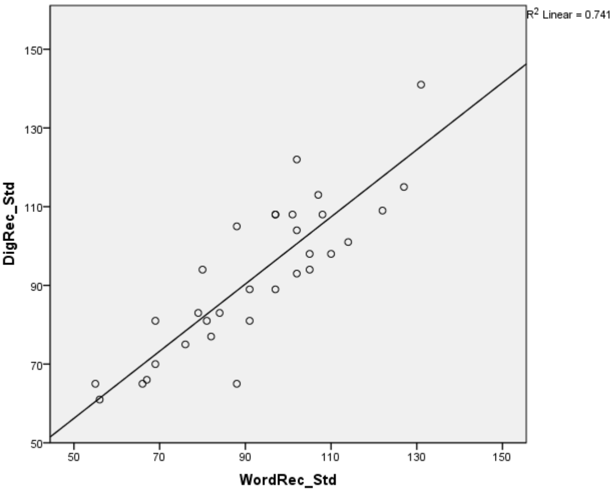
There were a few missing points in the data, most being where the youngest TD participant(s) were not able to complete the most challenging tasks. Where a participant had not completed a task, no score was assigned. This applied to the following tasks in the numbers specified: WISC (1), Rhyme (2), Recalling Sentences (1), Reading Screen (1), Nonword Reading (2), CP (1).

4.3 Results I: Performance battery

Data from the Word Recall subtest of the WMTB and the Alliteration subtest of the PhAB were eliminated for the following reasons:

- A paired sample t-test revealed no significant difference between the distributions of standard scores for the recall subtests (Digit Recall and Word Recall) within the typically developing (TD) group. There was also a high correlation between the two tasks (see Figure 4.1). It was decided that standard scores from either recall subtest alone adequately represented phonological short-term memory (PSTM) skills. The Digit Recall data were selected as this task had been completed by participants in both experiments.
- Most participants performed at ceiling on the Alliteration subtest of the PhAB. The range of raw scores in the control group was insufficient to calculate reliable z-scores for participants falling outside the published standardisation range (i.e. those younger than 6;0). The Alliteration task had been introduced as a measure of phonological awareness that did not require a verbal response, in case Rhyme data was difficult to score reliably in the SLI+SSD group. In the event, this alternative phonological measure was not required as the speech of all children in the SLI+SSD group was sufficiently intelligible. Moreover, the Rhyme subtest was completed by participants in both experiments.

Figure 4.1: Scatterplot showing the relationship between Digit Recall and Word Recall.



4.3.1 Calculation of z-scores (performance battery)

With the exception of Recalling Sentences, all tasks that were administered in both experiments were identical in content.

The performance of the typically developing (TD) groups from the two experiments was compared in order to determine the extent to which they were similar, as would be expected. A univariate general linear model was applied to the raw scores from each task (published and experimental) with age as the covariate, group (TD1 vs. TD2) as the fixed factor and raw score as the dependent variable. There was no significant interaction between group and age in each analysis, so the interaction was removed from the models and the analyses were re-run. No significant difference between raw scores for the TD groups was found on any task except Recalling Sentences. The raw, scaled or standard scores from both experiments were combined for all tasks except Recalling Sentences, so that N=51 for the combined TD group.

Standard scores were calculated for the TROG, BPVS and Digit Recall tests for all participants (with a theoretical population mean = 100 and population standard deviation = 15). These were converted to z-scores using the equation: (standard score – population mean)/(population standard deviation) i.e. (standard score – 100)/15.

The Rhyme and Nonword Reading tasks yielded standard scores only for participants over the age of six years. In order to have comparable age-normalised scores for the younger participants, it was necessary to calculate adjusted z-scores based on the raw scores of the TD participants. Firstly, standardised residuals for linear regressions of the raw scores on age (using the TD participants only) were saved as z-scores. Next, two-tailed paired samples t-tests were carried out for both tasks to compare the distributions of the TD z-scores with the sub-sets of TD standard scores that were available for older participants. Finally, the mean paired difference was subtracted from each z-score to correct for any distributional difference.

Younger participants in Experiment 2 had carried out the Block Design sub-test from the WPPSI in place of that from the WISC[4]. Consequently, all Experiment 2 participants were assigned published scaled scores, which was not the case for the youngest participants in Experiment 1. Z-scores were calculated and adjusted for all participants based on the raw scores of the combined TD group in the way described above for the Rhyme and Nonword Reading tasks.

The trend in performance as a function of age was significant and linear for the informal reading screen, RWFT and RAPT ($p < 0.01$). Standardised residuals for linear regression of the raw score on age (based on the TD participants only) were saved as unadjusted z-scores.

The Recalling Sentences subtest from version 3 of the CELF was administered in Experiment 1 and the same subtest from version 4 of the CELF was administered in Experiment 2. The versions differed in their standardisation (version 4 being standardised from five years rather than six) and in the form and content of some sentence stimuli. There was a significant difference in the distribution of Recalling Sentences raw scores for TD participants in Experiment 1 and those for TD participants in Experiment 2. A univariate general linear model was applied to the raw scores with age as the covariate, group (TD1 vs. TD2) as the fixed factor and raw score as the dependent variable. There was no significant interaction between group and age, so the interaction was removed from the model and the analysis was re-run. As there was a significant main effect of group (TD1 vs. TD2; $p < 0.001$) which accounted for 68% of the variance, Recalling Sentences data from the two experiments could not be combined before the calculation of z-scores. Z-scores were therefore calculated for both experiments separately before being combined into a single variable. For Experiment 2, a two-tailed, paired samples t-test showed that there was no significant difference between the scaled scores assigned from published tables and the z-scores calculated by linear regression based on the performance of the TD group. Therefore the computed z-score was used for the youngest male TD (who was outside the age range of the published norms) but the z-scores derived from the published scaled scores were used for the other children. For Experiment 1, the difference between the

---

[4] The label ‘WISC’ will be used to refer to either version of the Block Design sub-test.

distributions for assigned scaled scores vs. calculated z-scores almost reached significance. As seven out of 29 participants were too young to be assigned a scaled score, the calculated z-scores (adjusted by the mean paired difference) were used for all participants in this experiment.

4.3.2 Characterising the groups: descriptive statistics

Summary data for the standardised assessments is presented below (see Table 4.1). One-sample t-tests were carried out to compare the typically developing (TD) group's performance on each measure with the population mean (10 for a scaled score, 100 for a standard score). For the WISC, TROG and Rhyme tasks, the TD group was a representative sample of the population. There was no difference between the group mean and that expected from the population in general ($p > 0.05$). However, the TD participants performed better than average on Recalling Sentences, BPVS, Digit Recall and Nonword Reading ($p < 0.05$).

Table 4.1: Summary data for standardised assessments.

	TD				SLI (N=6)			SLI+SSD (N=8)		
	Valid N	Mean	s.d.	Range	Mean	s.d.	Range	Mean	s.d.	Range
Age	51	83.45	18.40	71.00	94.83	18.56	48.00	94.13	16.60	45.00
WISC^1	45	10.47	3.10	16.00	8.80	1.60	4.00	9.13	1.36	4.00
TROG^2	51	101.96	12.18	51.00	72.33	14.00	35.00	77.38	13.04	35.00
BPVS^2	51	110.71	9.99	44.00	86.17	13.09	30.00	97.38	11.81	39.00
Rhyme^2	43	102.81	9.82	62.00	81.50	13.58	35.00	72.75	8.29	24.00
Digit Recall^2	51	103.75	12.75	60.00	76.50	9.63	24.00	74.00	11.84	37.00
NW Reading^2	19	105.89	10.05	43.08	99.03	16.94	46.00	86.02	11.75	36.85
Recalling Sentences^1	44	10.89	2.68	12.00	3.17	2.04	5.00	3.13	1.25	4.00

Note: ^{1} Scaled scores: population mean = 10 population s.d. = 3
^{2} Standard scores: population mean = 100, population s.d. = 15
Valid scores for TD group dependent on experiment and age of participant.

For every assessment, normal distribution within each group and homogeneity of variance between groups were considered. Levene's test indicated homogeneity of variance across groups for each assessment except the WISC ($F_{(2,61)} = 4.953$, $p = 0.010$). Group differences in performance on all tasks except the WISC were evaluated using parametric statistics (one-way, between-subjects ANOVAs).

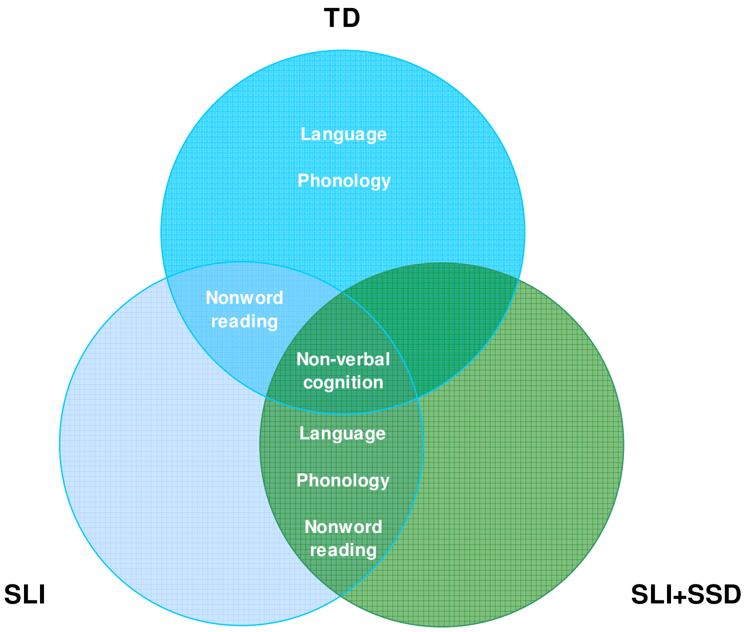
Group differences on WISC were investigated using the Kruskal-Wallis test. There was a significant effect of group ($\chi^2 = 6.938$, df = 2, $p = 0.031$). Mann-Whitney U-tests were used for post-hoc, pair-wise comparisons. The difference between the TD and SLI groups reached significance (U = 75.000, $N_1 = 50$, $N_2 = 6$, $p = 0.47$, two-tailed), but there was no significant difference between the performance of the SLI+SSD group and that of either the SLI or the TD groups. Given the disparity in the group sizes and how close the TD vs. SLI pair-wise comparison came to significance, non-verbal cognition was considered to be similar across groups and will not be considered further.

Table 4.2: Descriptive statistics for WISC z-scores: median and range

Group			Effect of group (Kruskall-Wallis, df=2)		Post-hoc pair-wise comparison (Mann-Whitney U-tests, two-tailed)					
TD (n = 50)	SLI (n = 6)	SLI+SSD (n = 8)			TD vs. SLI		TD vs. SLI+SSD		SLI vs. SLI+SSD	
.10 4.12	-.73 1.31	-.52 1.13	χ2 = 6.938	p = 0.031	U = 75.000	p = 0.047	U = 114.000	p = 0.053	U = 19.000	p = 0.573

Descriptive statistics for the variables that met the requirements for parametric analysis are shown in Table 4.3. The groups did not differ by age ($F_{(2,62)} = 1.996$, $p = 0.145$). For ease of reference, age is reported here in years and months but elsewhere it will be reported in months. Post-hoc pair-wise comparisons of group performance on each assessment revealed that there was no significant difference between the SLI and SLI+SSD groups on any published task, although these groups did differ on the informal reading screen ($p = 0.039$). The TD group performed similarly to the SLI group on nonword reading. In summary, then, the groups did not differ in terms of age or non-verbal cognition but the TD group was characterised by superior language and phonology skills, sharing a similar level of nonword reading with the SLI group. The SLI and SLI+SSD groups performed similarly on all published measures (see Figure 4.2 below).

Figure 4.2: Diagram illustrating the characteristics defining the experimental groups (published assessments).



Note: To be included in the TD or SLI groups, participants were required to pass the DEAP Diagnostic Screen. Participants with language difficulties who failed the DEAP screen were assigned to the SLI+SSD group.

Table 4.3: Descriptive statistics: mean, standard deviation and minimum/maximum values for z-scores

	Group			Effect of group (ANOVA)		Post-hoc pair-wise comparison (Bonferroni)		
	TD ($n = 51$)	SLI ($n = 6$)	SLI+SSD ($n = 8$)	F and p values		TD vs. SLI	TD vs. SLI+SSD	SLI vs. SLI+SSD
Age (yr;m)	6;11	7;10	7;10	$F_{(2,62)} = 1.996$	$p = 0.145$	$p = 0.459$	$p = 0.386$	$p = 1.000$
	1;6	1;6	1;4					
	4;0 - 9;11	5;7 - 9;7	6;2 - 9;11					
TROG	0.13	-1.84	-1.51	$F_{(2,62)} = 25.70$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p = 1.000$
	0.81	0.93	0.87					
	-1.6 - 1.8	-3 - -0.67	-2.67 - -0.33					
BPVS	0.71	-0.92	-0.18	$F_{(2,62)} = 18.37$	$p < 0.001$	$p < 0.001$	$p = 0.004$	$p = 0.126$
	0.67	0.87	0.79					
	-0.8 - 2.13	-2 - 0	-1.2 - 1.4					
RAPT Grammar	0.00	-1.63	-1.99	$F_{(2,62)} = 15.03$	$p < 0.001$	$p = 0.004$	$p < 0.001$	$p = 1.000$
	0.98	1.60	1.54					
	-3.11 - 1.53	-4.83 - -0.58	-4.85 - 0.28					
RWFT	0.00	-3.08	-2.23	$F_{(2,62)} = 33.58$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p = 0.434$
	0.99	1.52	1.20					
	-2.55 - 1.79	-4.85 - -1.13	-3.9 - -0.82					
Rhyme [b]	0.17	-1.23	-1.82	$F_{(2,60)} = 26.26$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p = 0.556$
	0.83	0.91	0.55					
	-2.07 - 2.22	-2.07 - 0.27	-2.07 - -0.47					
Digit Recall	0.25	-1.57	-1.73	$F_{(2,62)} = 29.33$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p = 1.000$
	0.85	0.64	0.79					
	-1.27 - 2.73	-2.33 - -0.73	-2.6 - -0.13					
Recalling Sentences [a]	-0.04	-2.28	-2.29	$F_{(2,61)} = 37.53$	$p < 0.001$	$p < 0.001$	$p < 0.001$	$p = 1.000$
	0.92	0.68	0.42					
	-2.6 - 2.33	-3 - -1.33	-3 - -1.67					
	TD ($n = 21$)	SLI ($n = 6$)	SLI+SSD ($n = 8$)	Effect of group (ANOVA)		TD vs. SLI	TD vs. SLI+SSD	SLI vs. SLI+SSD
Reading Screen [a]	0.00	-0.82	-2.33	$F_{(2,31)} = 15.73$	$p < 0.001$	$p = 0.324$	$p < 0.001$	$p = 0.039$
	0.97	1.07	1.03					
	-2.17 - 1.46	-2.25 - 0.53	-3.89 - -0.85					
NW Reading [b]	0.39	-0.06	-0.93	$F_{(2,30)} = 7.92$	$p = 0.002$	$p = 0.680$	$p = 0.001$	$p = 0.153$
	0.67	1.13	0.78					
	-0.81 - 2.07	-1 - 2.07	-2.59 - -0.13					

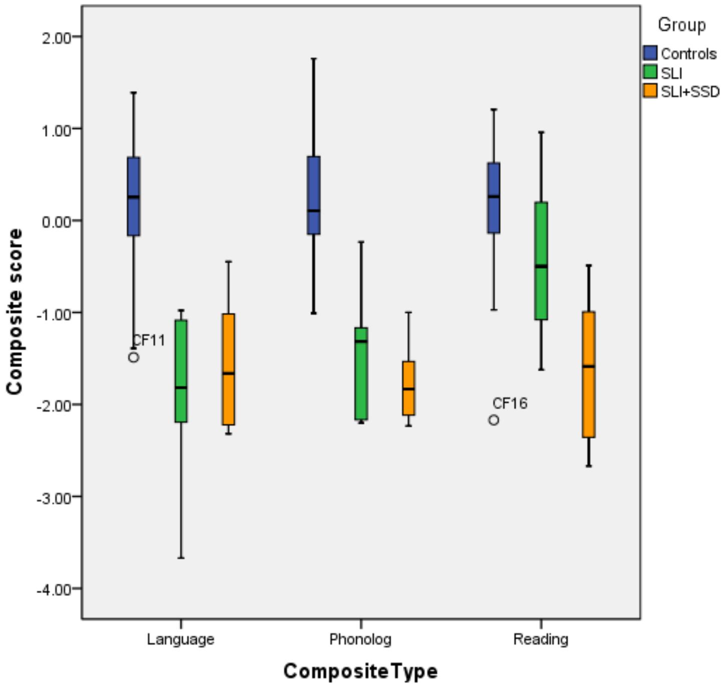
Notes: [a] One missing data point: Recalling Sentences (TD); Reading Screen (SLI). [b] Two missing data points: Rhyme (TD); Nonword Reading (TD).

4.3.3 Characterising the groups: composite measures

Three primary measures were extracted from the data to index performance in key skills: language, phonology and reading. The language composite score was the mean z-score achieved on the TROG, BPVS, RAPT Grammar, RWFT and Recalling Sentences tasks. The phonology composite was the mean z-score achieved on the Rhyme and Digit Recall tasks. Finally, the reading composite was the mean z-score achieved on the informal Reading Screen and Nonword Reading. Where a data point was missing, the participant's composite was calculated as the mean of the tasks (s)he had completed within the skill area. Reading data was only available for Experiment 2 participants.

Consistent with the analysis of individual performance measures, one-way between-subjects ANOVAs revealed a significant effect of group for each composite score. For composite language, $F_{(2,62)} = 49.439$, $p < 0.001$; for composite phonology, $F_{(2,62)} = 45.219$, $p < 0.001$; for composite reading, $F_{(2,32)} = 13.219$, $p < 0.001$. Post-hoc pair-wise comparisons of the groups confirmed the pattern of performance illustrated above in Figure 4.2 for language and phonology composites. However, there was a significant difference between the SLI and SLI+SSD groups on the composite reading score ($p = 0.031$). Boxplots of the composite scores confirmed that the TD group performed better than both SLI groups on language and phonology, and better than the SLI+SSD group on reading (see Figure 4.3). The SLI and SLI+SSD groups had similar language and phonology skills, but the SLI group's reading was better than that of the SLI+SSD group.

Figure 4.3: Boxplots showing group differences on composite measures.



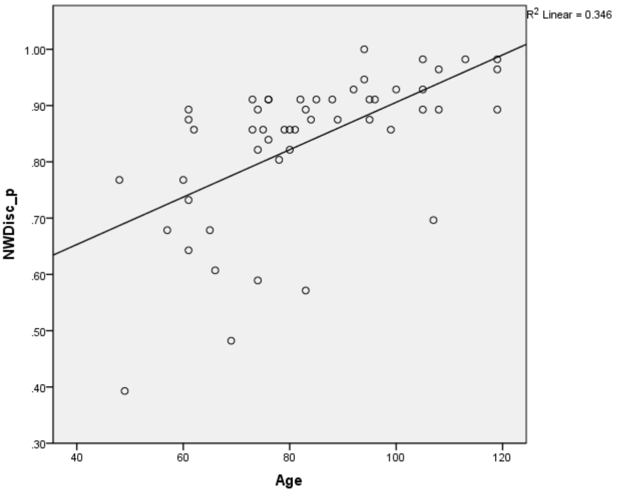
4.4 Results II: Experimental tasks

4.4.1 NWD

This task was scored out of 56 trials. An item was scored as correct if the participant identified the nonword that matched the stimulus in an AXB format. NWD stimuli were presented in a fixed random order in Experiment 1. In Experiment 2 the stimuli were randomised for each participant to control for presentation order effects. A comparison was made between NWD data from the typically developing groups (TD1 vs. TD2) and the presentation order modification proved non-significant so data from the two experiments were combined.

Overall performance on NWD improved linearly with age for the TD participants (r = 0.588, n = 51, p < 0.001, one-tailed; see Figure 4.4).

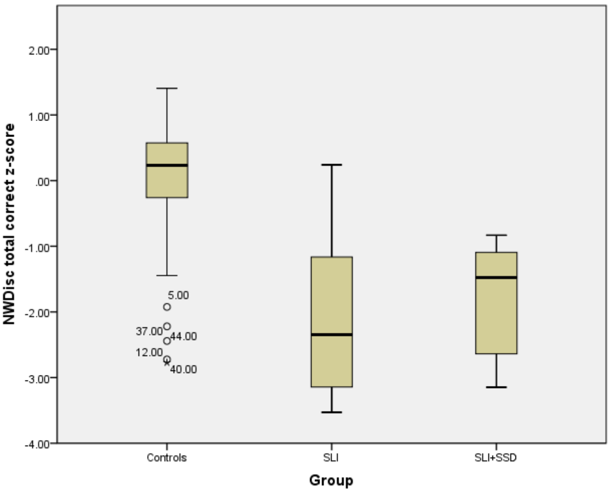
Figure 4.4: Scatterplot showing the effect of age on overall NWD performance



Note: TD participants only (N=51).

In order to factor out the effect of age, standardised residuals for linear regression of the raw score on age (using the TD participants as a selection variable) were saved as unadjusted z-scores. There were some outliers in the TD group, but a one-sample Kolmogorov-Smirnov (K-S) test confirmed that the distribution was normal (Z = 1.349, *p* = 0.53) and parametric analysis was carried out. A oneway ANOVA on the age corrected z-scores revealed a significant main effect of group ($F_{(2,62)}$ = 19.167, *p* < 0.001). Post-hoc pair-wise comparisons (Bonferroni) confirmed that there was a significant difference between the performance of the TD group and both clinical groups (*p* < 0.001). There was no difference in the overall performance of the SLI and SLI+SSD groups (*p* = 1.000). Figure 4.5 illustrates the relative performance of each group, with TDs performing best.

Figure 4.5: Boxplots showing group differences in overall performance on NWD (z-scores)

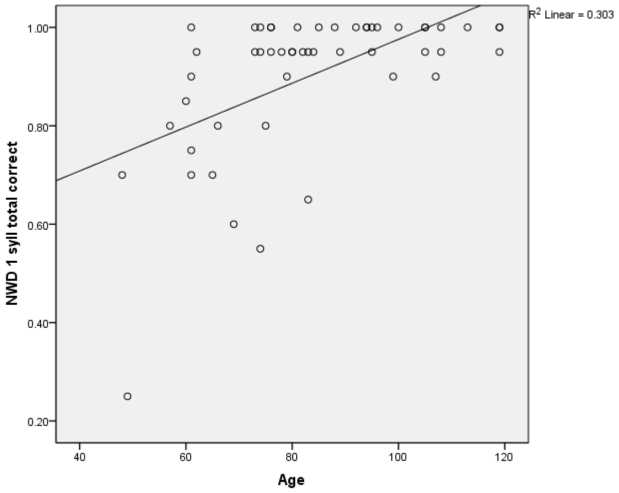


Note: Controls = TD participants

4.4.1.1 Number of syllables

The NWD stimuli consisted of either one or three syllables. The scatterplot in Figure 4.6 shows the relationship between age and performance on single syllable stimuli for the TD group. Many TD participants performed at, or close to, ceiling on this stimulus type and the distribution of scores was not normal (Z = 2.233, *p* <0.001). However a non-parametric test of correlation (Spearman's rho) was carried out and there was a positive correlation between age and NWD performance on single syllable words for the TD group (rho = 0.551, N = 51, *p* <0.001, one-tailed).

Figure 4.6: Scatterplot showing the effect of age on NWD of single syllable words

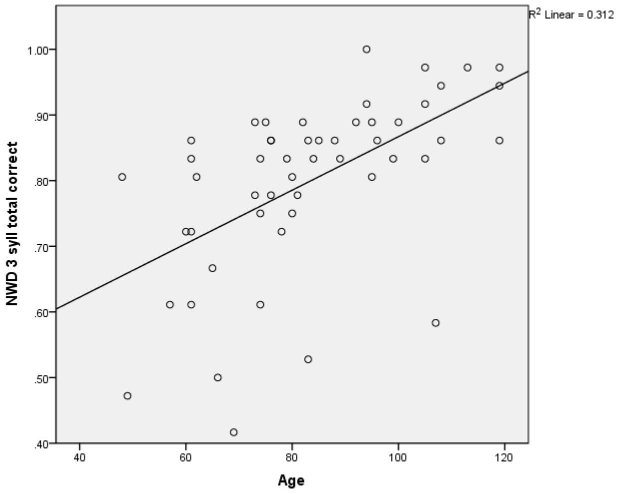


Note: Plot shows proportion correct for TD group only (N=51).

Figure 4.7 shows the effect of age on discrimination of three syllable nonwords in the TD group. A K-S test confirmed that the distribution was normal. A Pearson's R test was carried out and revealed a significant positive correlation between age and performance (r = 0.559, n = 51, *p* < 0.001, one-tailed).

All participants' raw scores for the one and three syllable word types were converted to z-scores through linear regression of the raw scores on age using the TD group as a selection variable. This factored out the effect of age.

Figure 4.7: Scatterplot showing the relationship between age and NWD of three syllable stimuli.



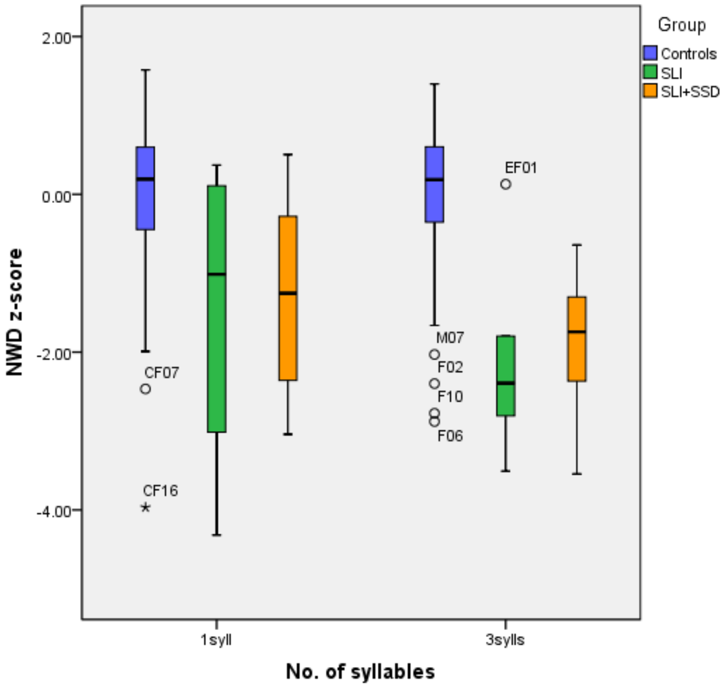
Note: Plot shows proportion correct for TD group only (N=51).

Figure 4.8 shows how the groups performed on nonwords consisting of one vs. three syllables. A two-way mixed ANOVA revealed main effect for number of syllables ($F_{(1,62)} = 5.317$, $p = 0.024$) and for group ($F_{(2,62)} = 17.317$, $p < 0.001$). The interaction between number of syllables and group was non-significant ($p = 0.114$). Post-hoc independent samples t-tests showed that in single syllable words, the only significant difference was between the TD and SLI+SSD groups (t = 3.296, d.f. = 57, $p = 0.002$), with TDs performing better than the SLI+SSD participants. For nonwords containing three syllables, the TD group performed better than the SLI group (t = 4.853, d.f. = 55, $p < 0.001$) and the SLI+SSD group (t = 5.041, d.f. = 57, $p < 0.001$). There was no difference between the clinical groups on longer nonwords.

Post-hoc paired samples t-tests were also carried out on the clinical groups (as the TD data were constrained to have the same distributions). There was no significant difference in performance within the clinical groups on one vs. three syllable nonwords, indicating that the pattern of errors was the same across groups.

Further analysis of the NWD task was restricted to nonwords of three syllables.

Figure 4.8: Boxplots showing NWD performance on 1 and 3 syllable stimuli (z-scores).



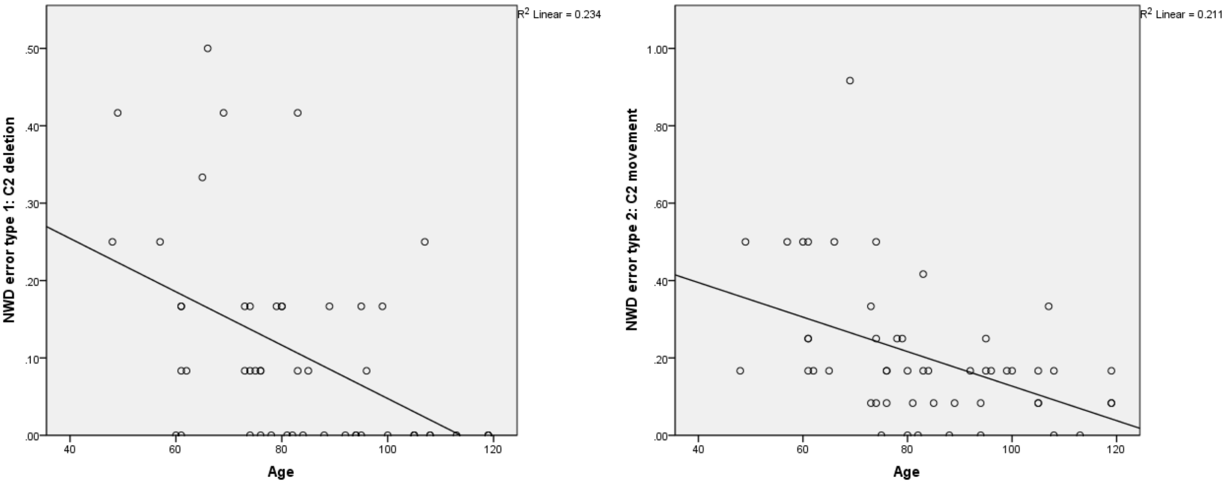
4.4.1.2 Error analysis

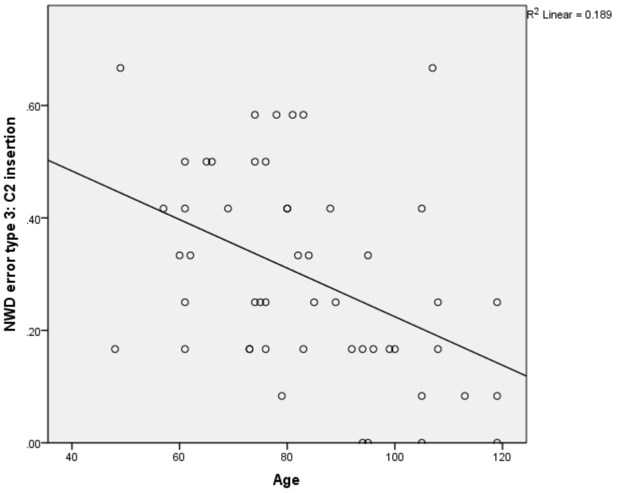
All stimuli that were presented in the NWD task contained a consonant cluster (e.g. /ˈklepəfə/). There were three foil types:

- The second consonant in the cluster (C2) was deleted (e.g. /ˈkepəfə/). Erroneous selection of such a foil was considered to be a C2 deletion error;
- The second consonant in the cluster was moved into a different word position (e.g. /ˈkepləfə/). Erroneous selection of such a foil was considered to be a C2 movement error;
- The stimulus cluster was reproduced, but an additional cluster was created by inserting a second consonant in another word position (e.g. /ˈklepləfə/). Erroneous selection of this foil was considered to be a C2 insertion error.

The scatterplots in Figure 4.9 show the proportion of errors by age for each error type. There were a total of 56 trials in the NWD task so the number of trials for each foil type was relatively small (C2 deletion = 22, C2 movement = 12, C2 insertion = 22). It is clear from the figures below, that for each type there were a considerable number of TD participants who made no errors.

Figure 4.9: Scatterplots showing the effect of age on NWD errors (by type).





Note: TD group only. Proportion of errors is shown, so a lower score indicates fewer errors and better performance.

A series of K-S tests on the error data confirmed that the C2 insertion error distribution was normally distributed (Z = 1.086, *p* = 0.189), but the C2 deletion and C2 movement distributions deviated significantly from normality (for C2 deletion, Z = 1.703, *p* = 0.006; for C2 movement, Z = 1.873, *p* = 0.002). However, one-tailed correlation tests (n = 51) revealed a significant age effect for each error type: for C2 deletion errors, rho = -0.509, *p* < 0.001; for C2 movement errors, rho = -0.486, *p* < 0.001; for C2 insertion errors, r = -0.435, *p* < 0.001. Note that the correlations were negative to reflect decreasing errors (improving performance) with increasing age.

To take account of the nonlinear distributions, quadratic regression of the TD raw scores on age was carried out. The resulting standardised residuals were compared to those from linear regression of the raw scores. There was no significant difference between the quadratic and the linear calculation, so standardised residuals from the linear regression of raw scores were saved as age-normalised z-scores and parametric analyses were carried out.

A two-way mixed ANOVA showed no significant effect of error type and no interaction between error type and group. Consistent with overall performance, there was a significant main effect of group ($F_{(62,1)}$ = 22.020, *p* < 0.001) (see Figure 4.10).

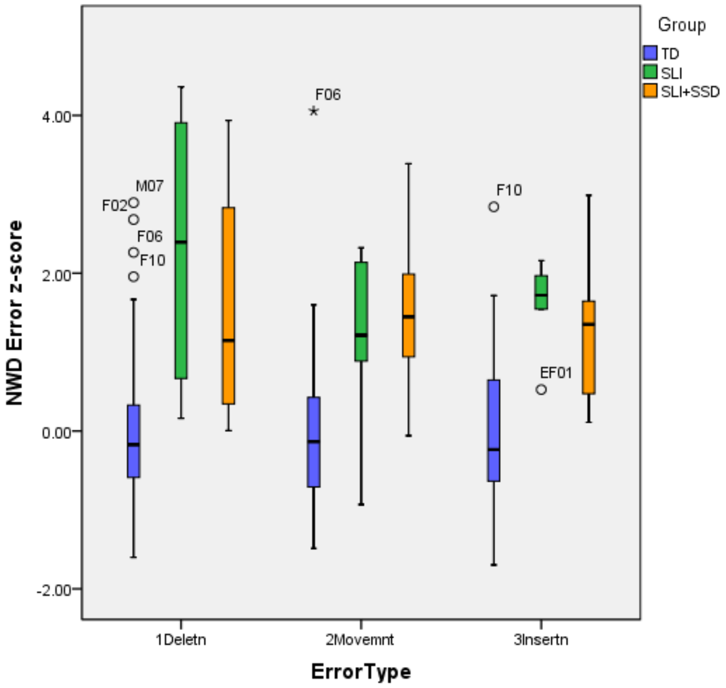
Post-hoc independent samples t-tests (two-tailed) were carried out to investigate the nature of the group effect. The TD group performed significantly better than both clinical groups on each error type (i.e. the controls produced fewer errors):

- TD vs. SLI, deletion errors: t = -3.008, d.f. = 5.341, *p* = 0.027;
- TD vs. SLI, movement errors: t = -2.625, d.f. = 55, *p* = 0.011;
- TD vs. SLI, insertion errors: t = -3.881, d.f. = 55, *p* < 0.001.
- TD vs. SLI+SSD, deletion errors: t = -3.908, d.f. = 57, *p* < 0.001;
- TD vs. SLI+SSD, movement errors: t = -3.990, d.f. = 57, *p* < 0.001;
- TD vs. SLI+SSD, insertion errors: t = -3.361, d.f. = 57, *p* = 0.001.

There was no significant difference between the clinical groups on any error type.

Within group comparisons were made for each error type in the clinical groups (the distribution of errors being constrained to be the same within the TD group). A series of pairwise comparisons showed that neither clinical group made significantly more errors of any type. The pattern of errors was therefore the same across all groups.

Figure 4.10: Boxplots showing the effects of group and NWD error type.



Note: The y-axis represents number of errors, so that a higher value denotes poorer performance.

4.4.1.3 Summary

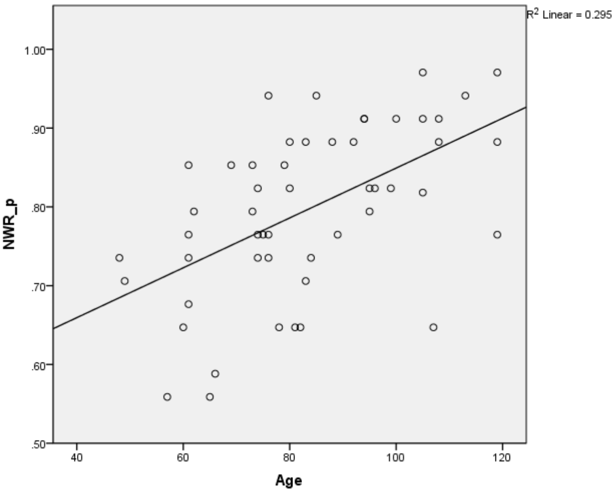
- Total number correct: the TD group performed better than both clinical groups. There was no difference between the SLI and SLI+SSD groups.
- One syllable NWD stimuli: the TD group performed similarly to the SLI group and better than the SLI+SSD group. There was no difference between the clinical groups.
- Three syllable NWD stimuli: the TD group performed better than both clinical groups. There was no difference between the SLI and SLI+SSD groups.
- There were no differences within the clinical groups when one and three syllable nonwords were compared. All groups performed in a similar pattern across one and three syllable stimuli.
- Errors: the TD group made fewer errors of each type compared to both clinical groups. There was no difference between the SLI and SLI+SSD groups in the numbers of each error they produced.
    - Neither clinical group favoured any one error type. All groups made a similar pattern of errors.

4.4.2 NWR

A broad phonetic transcription of the 34 nonwords was made at the time of testing. Transcriptions were later checked against recordings of each response. A segmental error of any nature was scored as an incorrect response. In order to check the reliability of the transcriptions, response data from four participants were randomly selected and transcribed by a second phonetically trained researcher. Intertranscriber reliability was 85%.

Overall performance on NWR improved with age for the typically developing (TD) participants. There was a significant positive correlation between age and proportion correct in the TD group ($r = 0.543$, $n = 51$, $p < 0.001$, one-tailed; see Figure 4.11).

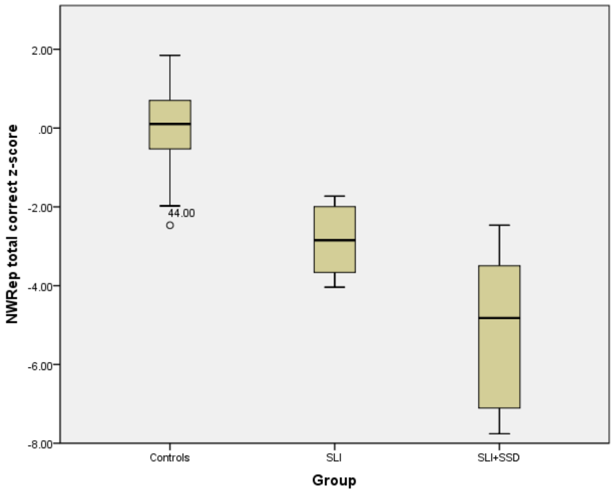
Figure 4.11: Scatterplot showing the effect of age on overall NWR performance.



Note: Plot shows proportion correct for the TD group only (N=51).

In order to factor out the effect of age, standardised residuals for linear regression of the raw score on age (using the TD participants as a selection variable) were saved as unadjusted z-scores. A Kruskall-Wallis test was carried out to investigate the effect of group on NWR performance as the two variables failed Levene's test for homogeneity of variance ($p = 0.003$). There was a main effect of group ($\chi^2 = 32.005$, df = 2, $p < 0.001$). Post-hoc pair-wise comparisons were made with two-tailed Mann-Whitney U-tests. Each group performed significantly differently from the others: for TD vs. SLI, U = 3.000, $N_1 = 51$, $N_2 = 6$, $p < 0.001$; for TD vs. SLI+SSD, U = 1.000, $N_1 = 51$, $N_2 = 8$, $p < 0.001$; for SLI vs. SLI+SSD, U = 7.000, $N_1 = 6$, $N_2 = 8$, $p = 0.029$. Figure 4.12 shows that the TD group performs better than both clinical groups and that participants with SLI perform better than those with SLI+SSD.

Figure 4.12: Boxplots showing group differences in overall performance on NWR (z-scores).



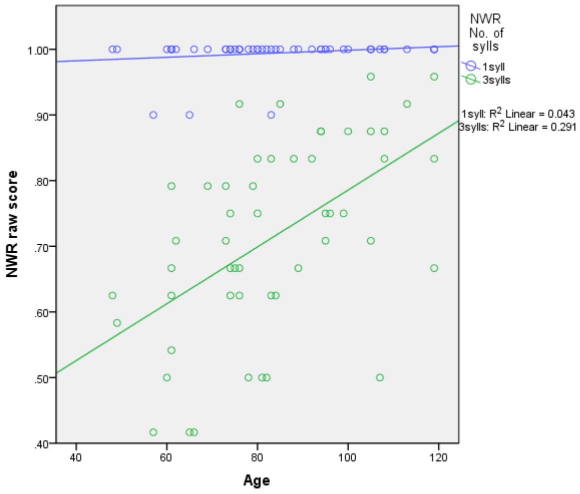
Note: Controls = TD participants.

4.4.2.1 Number of syllables

As in NWD, the nonwords in this task consisted of either one or three syllables. Figure 4.13 shows that almost without exception, the TD participants performed at ceiling on the single syllable nonwords (see Figure 4.13). There was no evidence of an age effect in the TD group, so z-scores were calculated using the equation (raw score – TD group mean) / (TD group s.d.). This calculation did not alter the distribution or variance of the TD group data for single syllable stimuli, but it maintained consistency of scale for all variables. It should be noted that this method yielded a very wide range of z-scores for the clinical groups, probably because the ceiling effect resulted in a small estimate of variance in the TD group.

The three syllable nonword data was normally distributed in the TD group ($Z = 0.743$, $p = 0.638$), and there was a significant positive correlation between performance and age ($r = 0.539$, $p < 0.001$, one-tailed; see Figure 4.13). Z-scores for the three syllable nonwords were calculated by saving standardised residuals for linear regression of the raw scores on age (using the TD participants as a selection variable).

Figure 4.13: Scatterplot showing the relationship between age and NWR performance on 1and 3 syllable stimuli.



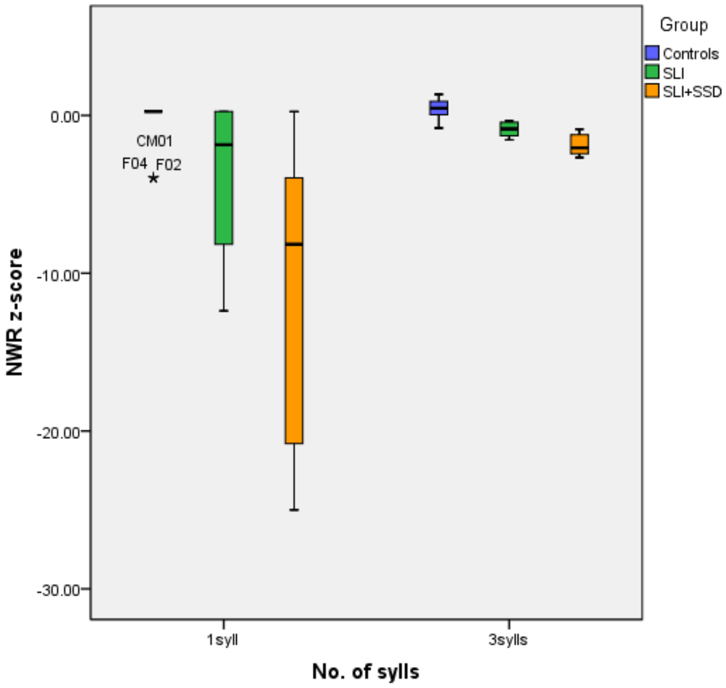
Note: Plot shows proportion correct for TD group only (N=51).

Despite the ceiling effect that was evident in the control group, a Kruskall-Wallis one-way between subjects test showed a significant effect of group for one syllable stimuli (for 1 syllable, $\chi^2 = 34.304$, df = 2, $p < 0.001$). Post-hoc pairwise comparisons were carried out with two-tailed Mann-Whitney U-Tests. The only significant difference was between the TD and SLI+SSD groups (U = 31.500, $N_1 = 51$, $N_2 = 8$, $p < 0.001$; see Figure 4.14).

A one-way ANOVA revealed a significant main effect of group for the three syllable stimuli ($F_{(2,62)} = 56.789$, $p < 0.001$). Post-hoc pairwise comparisons showed that each group performed significantly differently to the others (for TD vs. SLI, $p < 0.001$; for TD vs. SLI+SSD, $p < 0.001$; for SLI vs. SLI+SSD, $p = 0.011$; see Figure 4.14).

Within group differences were considered for the clinical groups. The number of syllables did not significantly affect performance in the SLI group. A Wilcoxon Signed Ranks Test revealed a significant difference between performance on single syllable vs. three syllable nonwords in the TD group (z = 2.831, N-Ties = 50, $p = 0.005$, two-tailed). There was also a significant difference between performance on one vs. three syllable nonwords in the SLI+SSD group (t = -2.834, d.f. = 7, $p = 0.025$).

Figure 4.14: Boxplots showing the effect of syllable number on NWR.



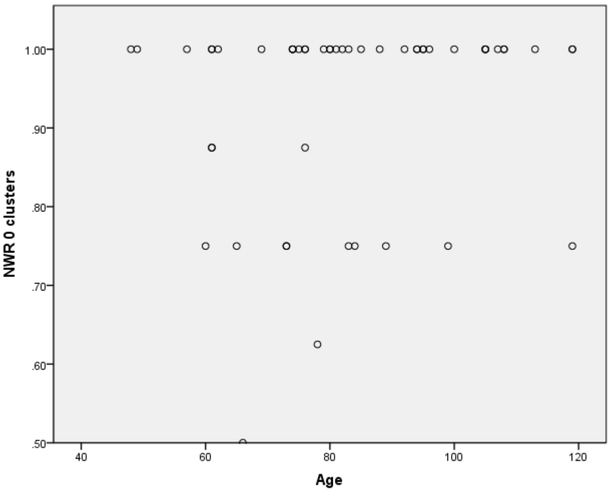
Note: The TD group ceiling effect on 1 syllable nonwords results in a wide range of z-scores for the clinical groups.

Further analysis of the NWR data was restricted to nonwords of three syllables.

4.4.2.2 Cluster number

All TD participants performed at or close to ceiling on three syllable nonwords containing no cluster, such as /ˈkepəfə/ (see Figure 4.15). There was no evidence of an age effect on stimuli of this type so raw scores were converted to z-scores using the equation (raw score – TD group mean) / (TD group standard deviation).

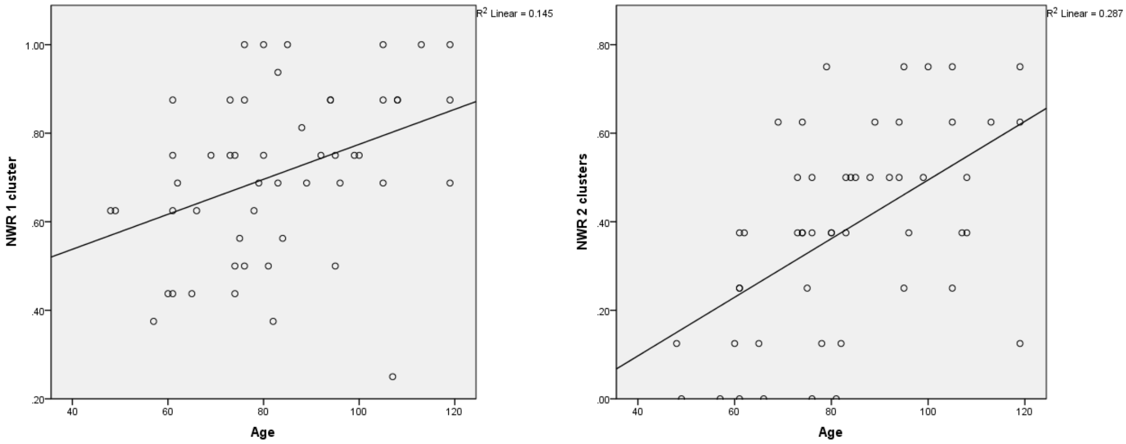
Figure 4.15: Scatterplot showing raw score against age for NWR stimuli containing no clusters.



Note: Plot shows proportion correct for TD group only (N=51).

There was a significant positive correlation between age and nonwords containing one cluster (e.g. /ˈklepəfə/; r = 0.381, n = 51, $p = 0.006$) and two clusters (e.g. /ˈkleplәfə/; r = 0.536, n = 51, $p < 0.001$) (see Figure 4.16). Z-scores were calculated on the basis of linear regression of the raw TD scores for both stimulus types.

Figure 4.16: Scatterplots showing raw score against age for NWR stimuli containing 1 or 2 clusters.



Note: Plot shows proportion correct for TD group only (N=51).

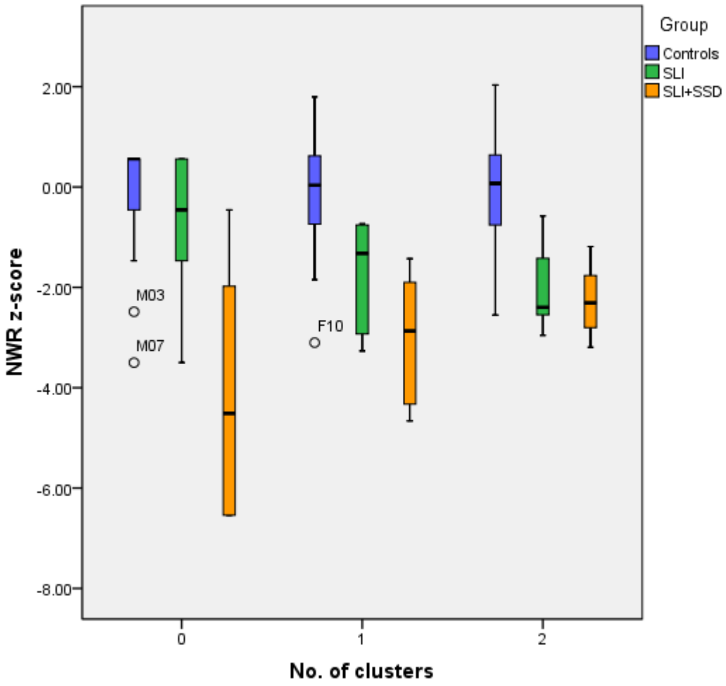
Despite the ceiling effect that was evident in the TD group, a Kruskall-Wallis one-way between subjects test showed a significant effect of group for zero clusters ($\chi^2 = 23.040$, df = 2, $p < 0.001$). Post-hoc pairwise comparisons were made with Mann-Whitney U-Tests. There was no difference in performance between the TD and SLI groups (U = 97.000, $N_1 = 51$, $N_2 = 6$, $p =$

0.152). The SLI+SSD group differed from both the TD group (U = 21.000, $N_1 = 51$, $N_2 = 8$, $p < 0.001$) and the SLI group (U = 6.000, $N_1 = 6$, $N_2 = 8$, $p = 0.020$).

One-way ANOVAs showed a main effect of group for nonwords containing one ($F_{(2,62)} = 34.030$, $p < 0.001$) and two clusters ($F_{(2,62)} = 28.725$, $p < 0.001$). Post-hoc pairwise comparisons revealed that the TD group performed better than the SLI ($p = 0.001$) and SLI+SSD ($p < 0.001$) groups on nonwords containing one cluster. There was weak evidence that the SLI group performed better than the SLI+SSD group on nonwords of this type ($p = 0.066$). The TD group performed better than both clinical groups on nonwords with two clusters ($p < 0.001$) but there was no significant difference between the clinical groups themselves.

Within group comparisons were made between each cluster type (zero-one, zero-two, one-two) for the clinical groups (the TD group being constrained to show no differences across number of clusters). None of the tests reached significance, although there was weak evidence that the SLI+SSD group performed better on nonwords containing one cluster vs. two (t = -2.146, d.f. = 7, $p = 0.069$).

Figure 4.17: Boxplots showing group performance on NWR stimuli containing 0, 1 or 2 clusters.



4.4.2.3 Error analysis

Three types of error in cluster production were coded to correspond with the stimulus/foil items in the NWD task:

- C2 deletion, whereby the second consonant in a cluster was deleted (e.g. /ˈklepəfə/ → /ˈkepəfə/);
- C2 movement, whereby the second consonant in a cluster was moved into a different word position (e.g. /ˈklepəfə/ → /ˈkepləfə/);
- C2 insertion, whereby the stimulus cluster was reproduced, but an additional cluster was created by inserting a second consonant in another word position (e.g. /ˈkepləfə/ → /ˈklepləfə/).

The effect of age on number of errors was investigated for each error type in the TD group (see Figure 4.18).

K-S tests were carried out to assess the distributions. Unsurprisingly, C2 deletion and C2 insertion errors deviated significantly from a normal distribution (for type 1, Z = 1.608, $p = 0.011$; for type 2, Z = 1.342, $p = 0.055$; for type 3, Z = 2.084, $p < 0.001$). There was, however, a significant effect of age for C2 deletion and C2 movement errors (n = 51): for C2 deletion, rho = -0.359, $p = 0.005$; for C2 movement, r = -0.386, $p = 0.003$.

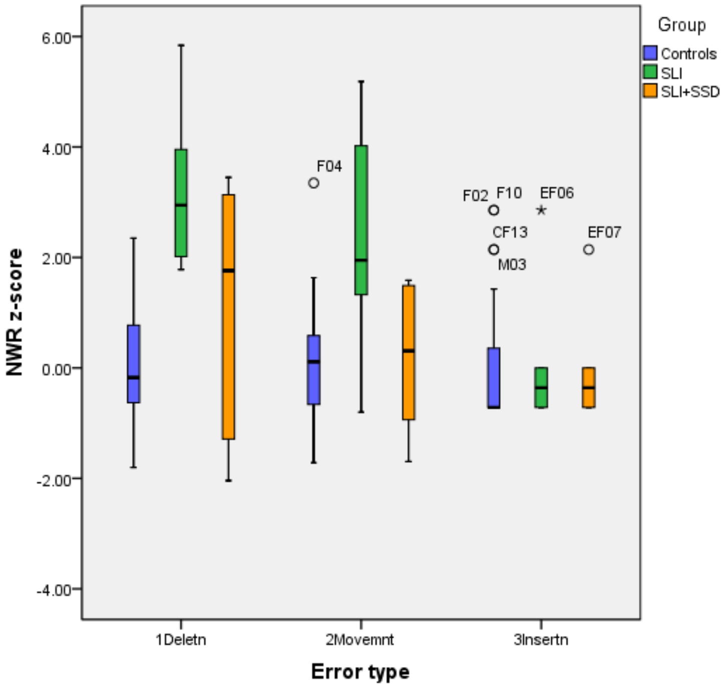
Z-scores were calculated for C2 deletion and C2 movement errors on the basis of linear regression of raw scores on age, using the TD group as a selection variable. Z-scores for C2 insertion errors were calculated with the equation (raw score – TD group mean) / (TD group standard deviation), which yielded a wide range of scores in the clinical groups.

As the data for each type of error violated the assumptions of normality and/or equality of variance, non-parametric group comparisons were made using Kruskall-Wallis test (df = 2). Post-hoc pair-wise comparisons were carried out with Mann-Whitney tests (two-tailed, see

Figure 4.18). There was a significant effect of group for C2 deletion errors ($\chi^2 = 14.917$, $p = 0.001$). The TD controls made significantly fewer of this error type than the SLI participants ($U = 7.000$, $N_1 = 51$, $N_2 = 6$, $p < 0.001$). There was also a significant effect of group for C2 movement errors ($\chi^2 = 7.984$, $p = 0.018$), again with only the TD group performing better than the SLI group ($U = 45.000$, $N_1 = 51$, $N_2 = 6$, $p = 0.003$). There was no effect of group for C2 insertion errors.

Within subject effects were examined in the clinical groups only (the TD data for each error type being constrained to be the same by the way the z-scores are calculated). Post-hoc paired t-tests revealed that the SLI+SSD group did not produce more errors of any type (i.e. they produced a similar pattern of errors to the TD group against which the z-scores were calculated). In the SLI group, however, there was a significant difference between C2 deletion and C2 insertion errors ($t = 3.197$, df = 5, $p = 0.024$), indicating that participants in this group produced a different pattern of errors to the TDs.

Figure 4.18: Boxplots showing the effects of group and NWR error type.



Note: The y-axis represents number of errors, so that a higher value denotes poorer performance.

4.4.2.4 Summary

- Total number correct: the TD group performed better than both clinical groups. The SLI group performed better than the SLI+SSD group.
- One syllable nonword stimuli: the TD group performed similarly to the SLI group and better than the SLI+SSD group. There was no significant difference between the clinical groups.
- Three syllable nonword stimuli: the TD group performed better than both clinical groups. The SLI group performed better than the SLI+SSD group.
- The TD and SLI+SSD groups performed better on nonwords containing one syllable as compared to those containing three syllables. The SLI group performed similarly on nonwords of either length.
- Zero clusters (three syllables): the TD group performed similarly to the SLI group and better than the SLI+SSD group. The SLI group performed better than the SLI+SSD group.
- One cluster (three syllables): the TD group performed better than both clinical groups. There was weak evidence that the SLI group performed better than the SLI+SSD group.
- Two clusters (three syllables): the TD group performed better than both clinical groups. There was no difference between the SLI and SLI+SSD groups.
- There was weak evidence that the SLI+SSD group repeated nonwords containing one cluster better than nonwords containing two clusters. There were no other within group differences in the SLI or SLI+SSD groups. The pattern of performance was similar across the TD and SLI groups.
- Errors: the TD group produced fewer C2 deletion and C2 movement errors than the SLI group. The SLI+SSD group produced a similar number of these errors to the TD and SLI groups. All groups produced a similar number of C2 insertion errors.
  - The SLI+SSD group produced a similar pattern of errors to the TD group, favouring no one error type. The SLI group produced more C2 deletion than C2 insertion errors, which indicated a different pattern of performance to the TD group. I.e. SLI+SSDs were quantitatively different to TDs while SLIs were quantitatively and qualitatively different to TDs.

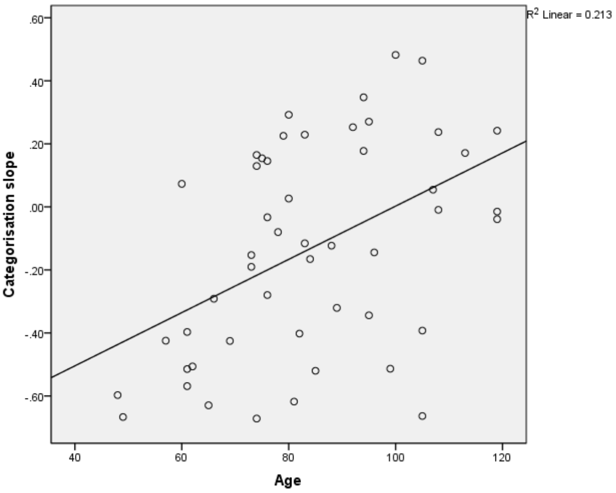
4.4.3 CP

The output of the CP task was an identification function that plotted the proportion of one of the two possible responses against the points on the stimulus continuum. The identification function for each participant was fitted using logistic regression and excluded the interspersed endpoint presentations. The slope of the identification function reflected the consistency with which the listener categorised items on the continuum, with a steeper slope indicating higher consistency. The CP task in Experiment 1 used a 50 point continuum. Experiment 2 used the same endpoints, but with a continuum of 100 points. Values from Experiment 1 were multiplied by two so that both sets of data could be combined.

Where a child had a fixed category boundary and consistently identified stimuli either side of this boundary, the gradient was not estimable, being essentially infinite. In such cases, responses at the category boundary were altered by a constant of 0.5 so that the slope of the function was not infinite. The distribution of the gradients was skewed, so a constant of 0.2 was added to each (to eliminate negative values) before a log transform was carried out. An identification function could not be extracted from the responses of one participant in the Experiment 1 typically developing (TD) cohort (thus for the TD group, n = 50).

As with the other experimental tasks, consistency of categorisation improved significantly with age for the TD participants (r = 0.462, n = 50, *p* = 0.001; see Figure 4.19).

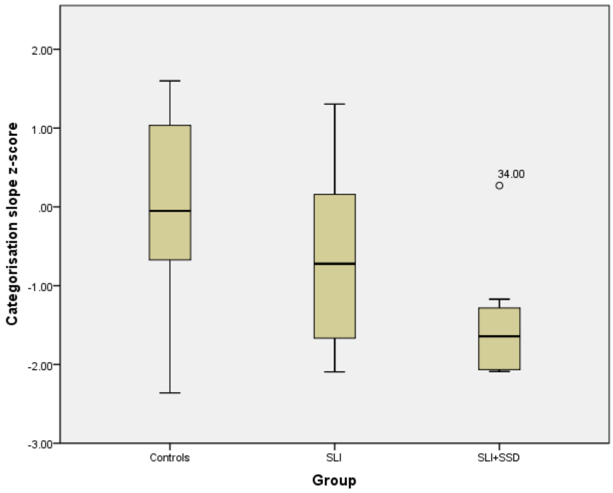
Figure 4.19: Scatterplot showing the relationship between age and categorisation consistency.



Note: Gradient for TD group (N = 50).

In order to factor out the effect of age, standardised residuals for linear regression of the raw score on age (using the TD participants as a selection variable) were saved as unadjusted z-scores. Figure 4.20 shows the performance of the three groups on the CP task. Despite an outlier in the SLI+SSD group, a one-sample K-S test confirmed that the distribution was normal (Z = 0.767, $p = 0.599$) so parametric analyses were carried out. A one-way ANOVA of the age corrected z-scores revealed a significant main effect of group ($F_{(2,61)} = 8.079$, $p = 0.001$). Post-hoc testing revealed that controls performed significantly better than the SLI+SSD group (t = 4.009, d.f. = 56, $p < 0.001$). There was no difference between the performance of the controls and the SLI participants, or between the SLI and SLI+SSD groups. It was noted that there was an outlier, EM14, in the SLI+SSD group. When data from this participant was eliminated from the analyses, there was weak evidence that the SLI group performed better than the SLI+SSD group on the CP task (t = 2.058, d.f. = 5.739, $p = 0.087$).

Figure 4.20: Boxplots showing group differences in categorisation consistency



Note: Controls = TD participants

4.4.3.1 Summary

- The TD group were more consistent than the SLI+SSD group at categorising the stimuli. There was no difference in categorisation consistency between the TD and SLI groups;
- There was some evidence that the SLI group was more consistent at categorising the stimuli than the SLI+SSD group.

4.5 Results III: Relationships between experimental tasks

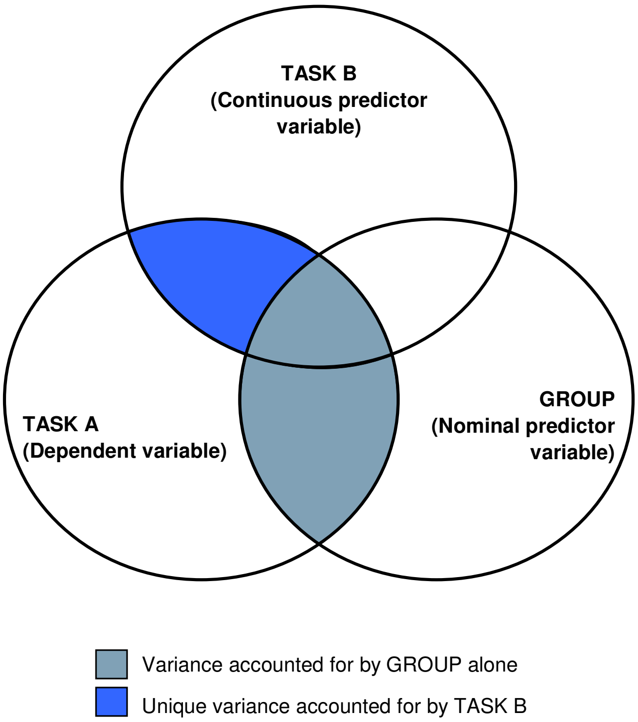
The regression technique described later in this section and in the sections that follow was designed to investigate the relationships between measures while also accounting for group differences. In essence, the technique assessed whether or not a model obtained from a scatterplot with more than one distinct regression line, was necessary. A detailed description of this technique can be found in Rosen (2003), with a statistical account in Cook and Weisberg (1999).

For each relationship a series of regressions was carried out to test whether a model with a single regression line best accounted for the data, or whether separate regression lines were necessary. If the nature of the regression was indeed different in one or more groups, combining them for a single regression could have been misleading. The effect of group was therefore taken into account, firstly by evaluating whether the nature of the regression was different in the clinical groups. If it was not, the SLI and SLI+SSD groups were combined and the variability in the clinical groups was compared to that in the typically developing (TD) group. If the nature of the regression was different in the SLI and SLI+SSD groups, they were included separately in a model with three groups (TD, SLI, SLI+SSD). In effect, distinct regression lines were fit to the data for each group or combination of groups that required it.

A non-significant interaction between group and the continuous predictor variable was always removed from a model as this indicated that there was the same relationship between the variables within each group and each regression line had the same slope. A non-significant interaction may ‘soak up’ the main effects so if it is removed from the model, the main effect sizes increase.

The model of the main effects (after the interaction had been removed) is illustrated in Figure 4.21. A main effect of group in the model showed that group membership accounted for a significant proportion of the variance in performance of Task A. If the effect of group was the only main effect, the relationship between Task B and Task A was dependent on which group a participant belonged to. In such models, knowledge about performance on Task B was irrelevant in predicting performance on Task A, once a participant's group membership was known. If group and Task B both predicted performance on a task, the group effect was eliminated from the model to leave only the unique variance accounted for by Task B.

Figure 4.21: Diagram illustrating the variance accounted for by group and task predictors.



In summary, a general linear model was used, including a nominal predictor variable (group) and a continuous predictor variable (Task B). This measured the extent to which group and a task (Task B) predicted performance on another measure (Task A). It should be noted that some of the variables in the analyses did not meet the assumptions of normality or equality of variance. Linear regression is reasonably robust against such violations. Moreover, the current clinical groups were so small that there were few statistical alternatives for analyses of this nature.

This section continues with a comparison of performance on the experimental tasks: nonword repetition (NWR), nonword discrimination (NWD) and categorical perception (CP).

4.5.1 Overall task performance

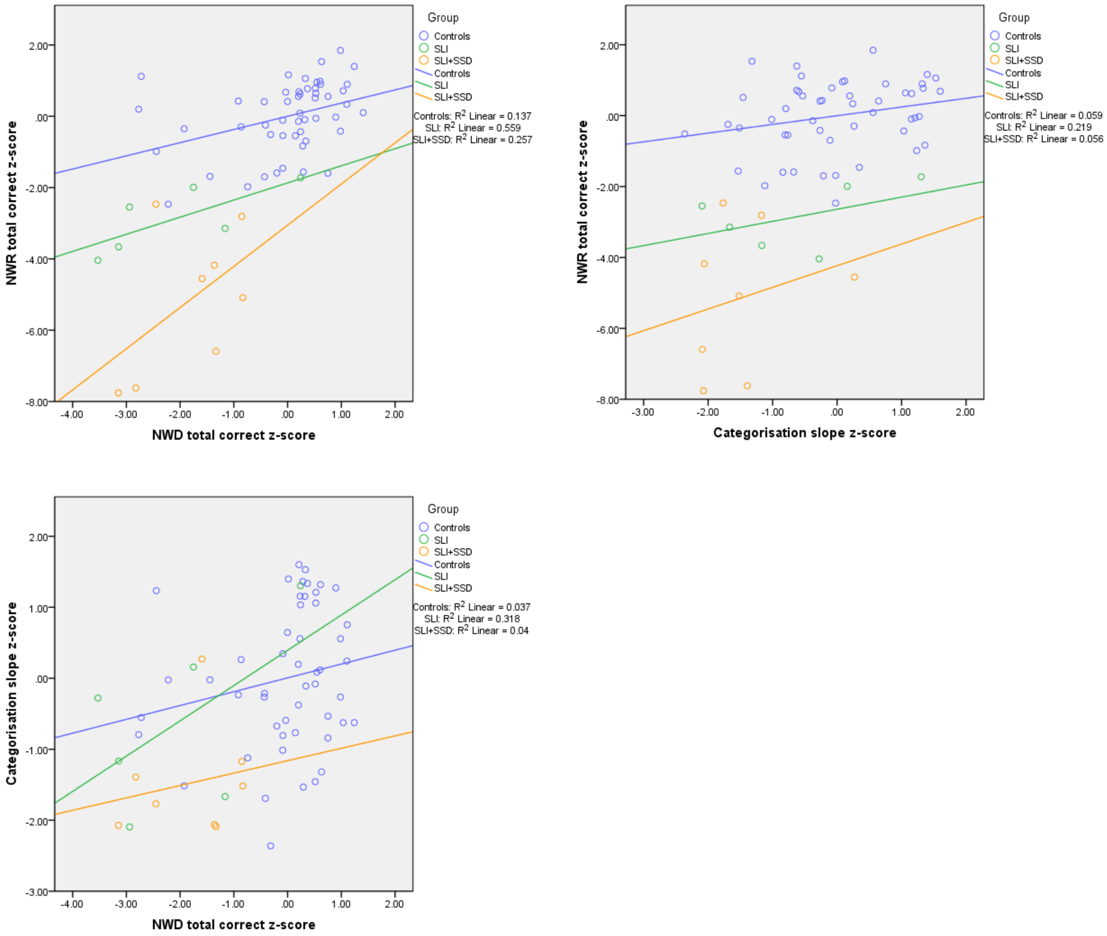
The TD group was constrained to perform similarly on each task. Repeated measures ANOVAs restricted to each clinical group were carried out to investigate task effects within the SLI and SLI+SSD groups. There was a significant main effect of task in the SLI+SSD group ($F_{(2,14)} = 23.916$, $p < 0.001$). Post-hoc pairwise comparisons showed that this group performed significantly better on the NWD task than on the NWR task (t = 5.361, d.f. = 7, $p = 0.001$). They performed significantly better on the CP task than the NWR task (t = -5.172, d.f. = 7, $p = 0.001$). There was no difference between their performance on the NWD and CP tasks (t = 0.860, d.f. = 7, $p = 0.418$). Removing EM14 from the analyses made no difference to the significance of any comparison.

There was also a significant main effect of task in the SLI group ($F_{(2,10)} = 11.762$, $p < 0.002$), but the pattern of performance was somewhat different to that observed in the SLI+SSD group. The SLIs performed the same on the NWD and NWR tasks (t = 2.041, d.f. = 5, $p = 0.097$), but better on CP than each of the other tasks (CP vs. NWD: t = -2.749, d.f. = 5, $p = 0.040$; CP vs. NWR: t = -4.688, d.f. = 5, $p = 0.005$).

4.5.2 Predicting performance

Figure 4.22 shows the relationships between the experimental tasks. There are three distinct regression lines for each plot. It seems that for the task comparisons depicted, the nature of the regression is different in each group. This assumption was checked by carrying out a series of linear regressions with models including different group variables.

Figure 4.22: Scatterplots showing the relationships between experimental groups.



Note: Regression lines are plotted for each participant group.
The range of the y-axis scale in the bottom plot is different to those at the top.

Within the psycholinguistic framework, CP can be regarded as a lower level skill than NWD which, in turn, is a lower level skill than NWR. Analyses were undertaken to evaluate whether NWR could be predicted from performance at the levels of NWD and CP, and whether NWD could be predicted from performance at the level of CP.

4.5.2.1 Predicting NWR

An attempt was made in the clinical groups alone to predict NWR from NWD. As noted above, there was a significant difference between the SLI and SLI+SSD groups in NWR (*p* = 0.012). A linear regression model was used to assess the value of NWD and group (TD, SLI, SLI+SSD) in predicting NWR. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.257$) so that term was eliminated and the model re-fit. Both group ($p < 0.001$) and NWD ($p = 0.001$) were significant predictors of NWR, the complete model accounting for 76% of the variance. Eliminating NWD from the model showed group membership to account for 72% of the variance implying that including NWD accounted for a further 4%.

As the calculation was dominated by the numbers in the TD group, the analysis returned to the linear regression model of the clinical groups alone. The interaction between group and NWD was non-significant ($p = 0.421$) so it was eliminated from the model and the model re-fit. Group was a significant predictor of NWR ($p = 0.012$) and, given the small number of participants, there was weak evidence that NWD predicted NWR in the clinical groups ($p = 0.085$). This model accounted for 51% of the variance. Eliminating NWD showed group membership to account for 35% of the variance, implying that including NWD accounted for a further 16% in the clinical groups.

There was no difference between the SLI and SLI+SSD groups when CP was used as a predictor for NWR ($p = 0.081$). These groups were combined into a single, ‘clinical’ group for the analysis. A linear regression model was used to assess the value of CP and group (TD vs. clinical) in predicting NWR. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.119$), so that term was eliminated and the model re-fit. Both group ($p < 0.001$) and CP ($p = 0.017$) were significant predictors of NWR, the complete model accounting for 69% of the variance. Eliminating CP from the model showed group membership to account for 65% of the variance, implying that including CP accounted for a further 4%. When the linear regression model of the clinical groups alone was reconsidered, neither group ($p = 0.081$) nor CP ($p = 0.381$) were significant predictors of NWR.

4.5.2.2 Predicting NWD

When CP was considered as a predictor for NWD, there was no difference between the SLI and SLI+SSD groups ($p = 0.317$). These groups were again combined into a single, ‘clinical’ group.

A linear regression model was used to assess the value of CP and group (TD vs. clinical) in predicting NWD. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.562$), so that term was eliminated and the model re-fit. Group was a significant predictor of NWD ($p < 0.001$) and accounted for 38% of the variance. CP did not predict NWD ($p = 0.069$). When the linear regression model of the clinical groups alone was reconsidered, neither group ($p = 0.317$) nor CP ($p = 0.136$) were significant predictors of NWD.

Next, regressions were carried out to investigate whether output processes were involved in the NWD task. Firstly, an attempt was made to predict NWD from NWR. In the clinical groups alone, there was no difference between SLI and SLI+SSD ($p = 0.168$) so these groups were combined. A linear regression model was used to assess the value of NWR and group (TD vs. clinical) in predicting NWD. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.341$) so that term was eliminated and the model re-fit. There was no main effect of group membership ($p = 0.129$) but NWR did predict NWD ($p = 0.006$). When group was eliminated from the model, NWR accounted for 38% of the variance.

4.5.2.3 Predicting CP

Finally, an attempt was made to see if output processes were involved in the CP task, by using NWR as a predictor. In the clinical groups alone, there was no difference between SLI and SLI+SSD ($p = 0.494$) so these groups were combined. A linear regression model was used to assess the value of NWR and group (TD vs. clinical) in predicting CP. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.979$) so that term was eliminated and the model re-fit. There was no main effect of group membership ($p = 0.807$) but NWR did predict CP ($p = 0.017$). When group was eliminated from the model, NWR accounted for 25% of the variance.

4.5.3 Predicting NWR errors

Error data for the NWR and NWD tasks was used to evaluate whether a specific NWD error would predict a similar error on the NWR task.

4.5.3.1 Predicting NWR C2 deletion errors

An attempt was made in the clinical groups alone to predict NWR C2 deletion errors from C2 deletion errors on the NWD task. There was no difference between the SLI and SLI+SSD groups ($p = 0.059$) so these groups were combined to form a single, ‘clinical’ group. A linear regression model was used to assess the value of NWD C2 deletion errors and group (TD vs. clinical) in predicting NWR errors of this type. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.836$) so that term was eliminated and the model re-fit. Group was a significant predictor of NWR C2 deletion errors ($p < 0.001$), accounting for 28% of the variance, but there was no main effect of the NWD error type ($p = 0.562$). When the analysis was originally carried out on the clinical groups alone, the difference between the SLI and SLI+SSD groups came close to significance. The linear regression was also carried out with three groups included in the model (TD, SLI, SLI+SSD). This model showed a significant interaction between group and NWD error ($p = 0.002$) which undermined any main effects that existed.

4.5.3.2 Predicting NWR C2 movement errors

An attempt was made in the clinical groups alone to predict NWR C2 movement errors from C2 movement errors on the NWD task. As the difference between the SLI and SLI+SSD groups came close to significance ($p = 0.059$) and the groups were so small, linear regressions were carried out on two models: one with the clinical groups combined and one with the group distinction maintained. A complete model including both main effects and their interaction where the clinical groups were distinct showed no interaction between group (TD, SLI, SLI+SSD) and NWD C2 movement errors ($p = 0.964$) so that term was eliminated from the model. Group was a significant predictor of NWR errors of this type, and accounted for 25% of the variance. NWD errors of the same type did not predict NWR errors (*p* = 0.933).

When a similar linear regression was carried out with the clinical groups combined, NWD errors remained non-significant (*p* = 0.735), and group membership accounted for 12% of the variance (*p* = 0.010).

4.5.3.3 Predicting NWR C2 insertion errors

Finally, an attempt was made in the clinical groups alone to predict NWR C2 insertion errors from NWD C2 insertion errors. There was no group difference between SLI and SLI+SSD (*p* = 0.776), so these groups were combined in the analysis to form a single 'clinical' group. A linear regression model was used to assess the value of NWD C2 insertion errors and group (TD vs. clinical) in predicting this type of error on the NWR task. A model including both main effects and their interaction showed the interaction term to be non-significant (*p* = 0.381) so that term was eliminated and the model re-fit. There was no main effect of group (*p* = 0.164), but NWD C2 insertion errors did predict production of the same error in the NWR task (*p* = 0.009), accounting for 10% of the variance.

Within group comparisons were carried out for the clinical groups' NWD and NWR data. Only C2 deletion and C2 movement errors were considered. Z-scores for NWR insertion errors were calculated differently to those for the NWD errors of the same type, so a direct comparison could not be made between the two tasks. The SLI group made comparable numbers of C2 deletion and C2 movement errors on each task. There was weak evidence that the SLI+SSD group produced more C2 movement errors on NWD than on NWR (t = -2.129, d.f. = 7, *p* = 0.071).

4.5.4 Summary

See also Figure 4.23.

Predicting NWR:

- In a model using the three separate groups (TD, SLI, SLI+SSD), group membership accounted for 72%. Adding NWD to the model accounted for a unique 4% of the variance
- In the clinical groups alone, NWD accounted for 16% of the variance and group membership accounted for 35%.
- In a model using two groups (TD vs. clinical), CP accounted for 4% of the variance but group membership accounted for 65%. Neither CP nor group predicted NWR in a model using the clinical groups alone.

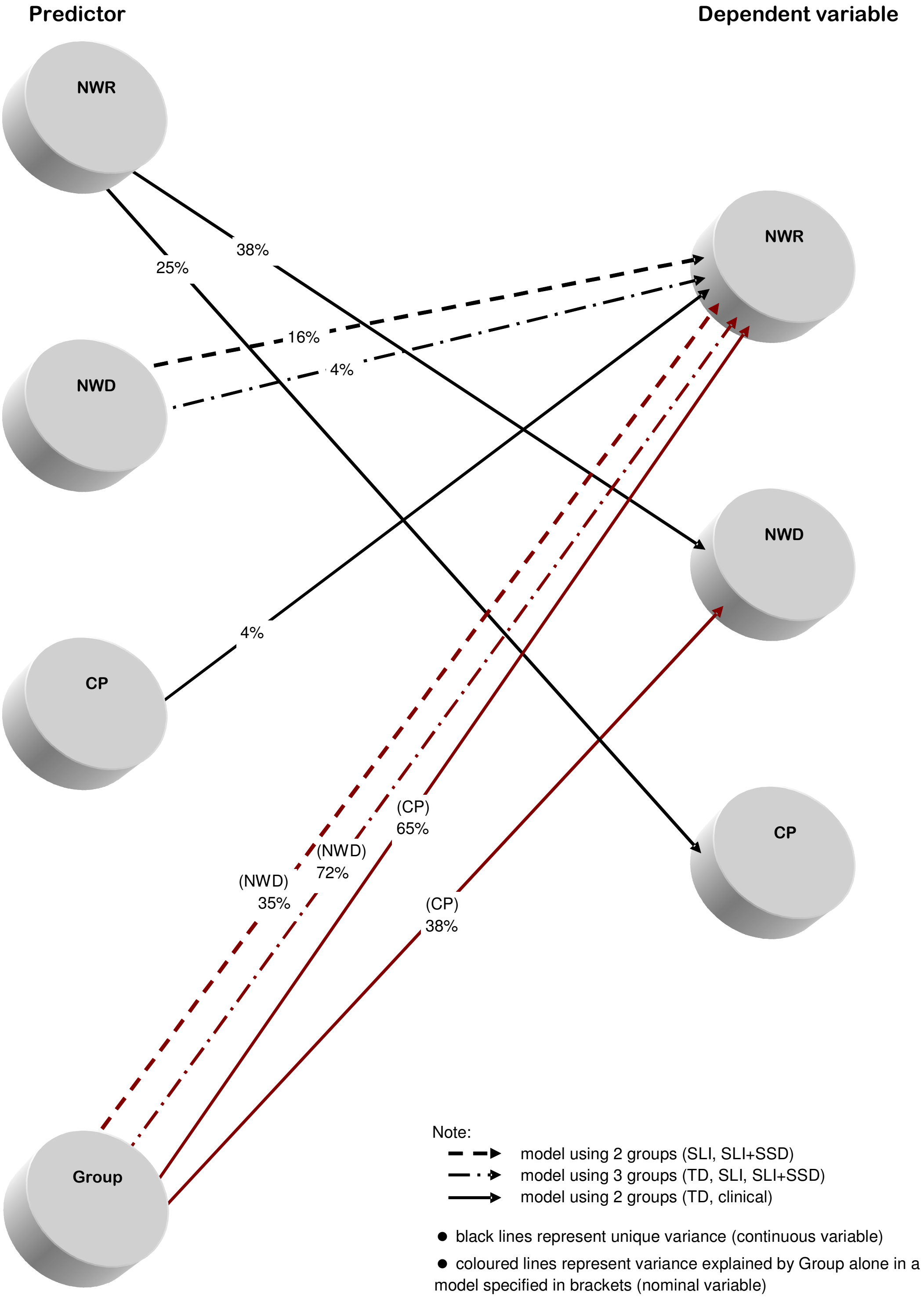
Predicting NWD:

- In a model using two groups (TD vs. clinical), group membership accounted for 38% of the variance. CP did not predict NWD. Neither CP nor group predicted NWD in a model using the clinical groups alone.
- NWR predicted NWD, accounting for 38% of the variance in a model with two groups (TD vs. clinical). Group membership did not predict NWD.

Predicting CP:

- NWR predicted CP in a model with two groups (TD vs. clinical). NWR accounted for 25% of the variance.

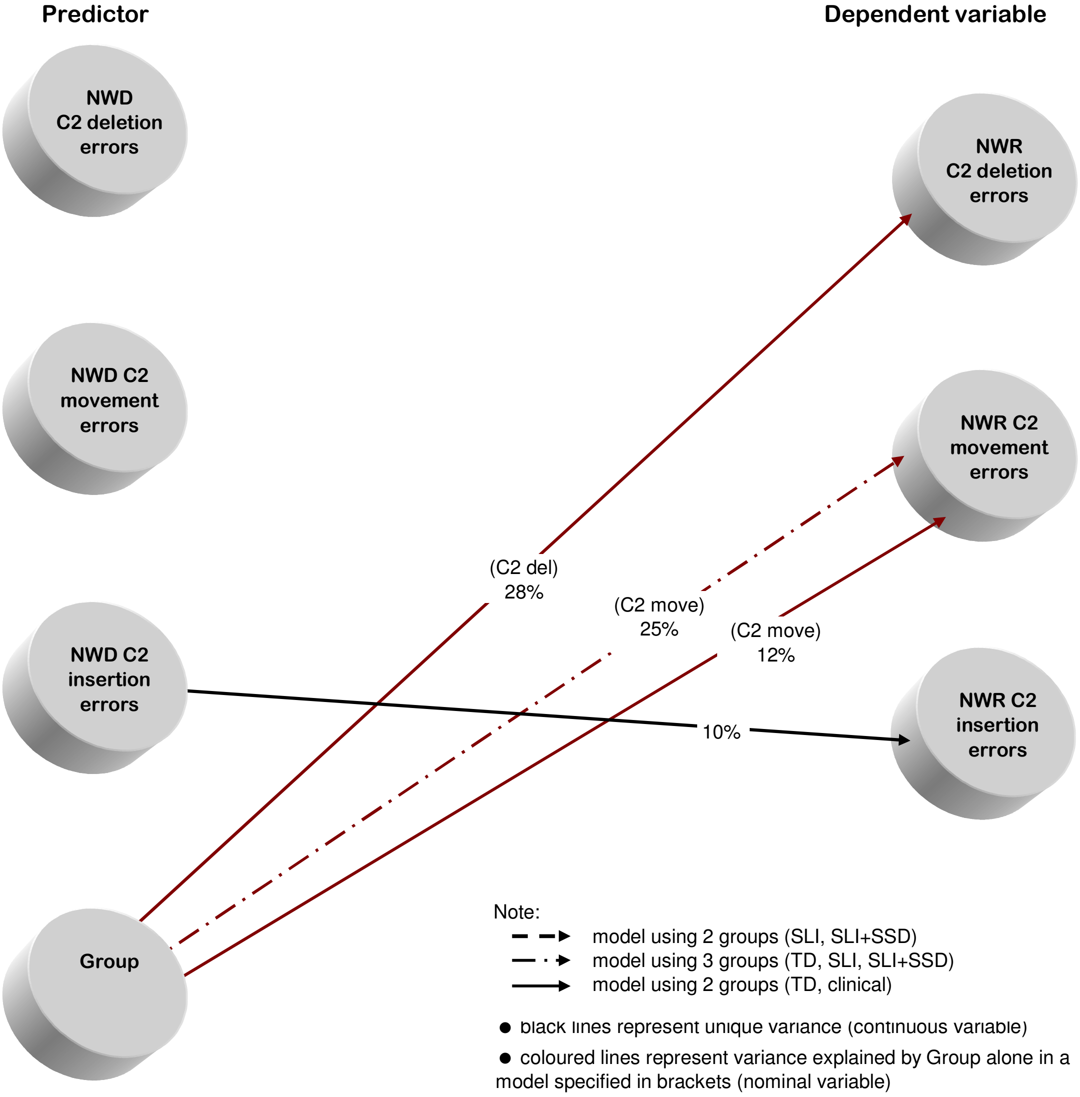
Figure 4.23: Diagram illustrating predictors of NWR, NWD and CP.



Predicting NWR errors (see Figure 4.24):

- NWR C2 deletion: in a model using two groups (TD vs. clinical), group accounted for 28% of the variance. NWD errors of this type did not predict NWR errors of the same type.
- NWR C2 movement errors: in a model using three groups (TD, SLI, SLI+SSD), group accounted for 25% of the variance. NWD errors of this type did not predict NWR errors of the same type.
- NWR C2 insertion errors: in a model using two groups (TD vs. clinical), NWD errors of this type accounted for 10% of the variance. Group did not predict C2 insertion errors.

Figure 4.24: Diagram illustrating predictors of NWR errors.



Note:
model using 2 groups (SLI, SLI+SSD)
model using 3 groups (TD, SLI, SLI+SSD)
model using 2 groups (TD, clinical)

- black lines represent unique variance (continuous variable)
- coloured lines represent variance explained by Group alone in a model specified in brackets (nominal variable)

4.6 Results IV: Relationships between experimental tasks and linguistic skills

4.6.1 Predicting receptive vocabulary

4.6.1.1 NWR as a predictor (clinical groups only)

An attempt was made in the clinical groups alone to predict receptive vocabulary (BPVS z-score) from NWR. There was no group difference between SLI and SLI+SSD ($p = 0.100$), so these groups were combined in the analysis to form a single 'clinical' group. A linear regression model was used to assess the value of NWR and group (TD vs. clinical) in predicting receptive vocabulary. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.230$) so that term was eliminated and the model re-fit. Group was a significant predictor of receptive vocabulary ($p = 0.007$) and accounted for 34% of the variance. NWR did not predict receptive vocabulary ($p = 0.538$). When the linear regression model of the clinical groups alone was reconsidered, neither group ($p = 0.100$) nor NWR ($p = 0.451$) predicted receptive vocabulary. Finally, the TD data from Experiment 1 was compared to that from Experiment 2 so that the predictive effect of NWR in TD participants only could be evaluated. Neither experimental group ($p = 0.856$) nor NWR ($p = 0.165$) predicted receptive vocabulary in TDs.

4.6.1.2 NWD as a predictor

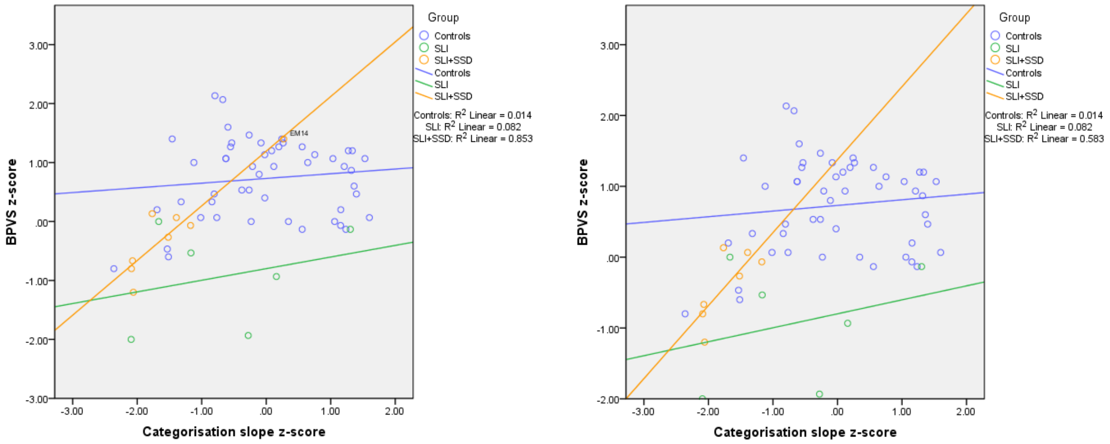
There was no difference between the SLI and SLI+SSD groups when NWD was used as a predictor for receptive vocabulary ($p = 0.144$). These groups were again combined into a single group. A linear regression model was used to assess the value of NWD and group (TD vs. clinical) in predicting receptive vocabulary. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.173$), so that term was eliminated and the model re-fit. Group was a significant predictor of receptive vocabulary ($p < 0.001$) and accounted for 35% of the variance. NWD did not predict receptive vocabulary ($p =$

0.234). When the linear regression model of the clinical groups alone was reconsidered, neither group (*p* = 0.144) nor NWD (*p* = 0.184) predicted receptive vocabulary.

4.6.1.3 CP as a predictor

When CP was considered as a predictor for receptive vocabulary in the clinical groups alone, there was a significant difference between the SLI and SLI+SSD groups (*p* = 0.022). Closer inspection of a scatterplot of CP plotted against receptive vocabulary (BPVS z-score; see Figure 4.25) showed that there was a high performing outlier in the SLI+SSD group (participant EM14). When the data from this participant was removed from the group comparison, there was no longer a significant difference between the SLI and SLI+SSD groups (*p* = 0.100), so they were combined to form a single 'clinical' group, with data from EM14 included in the analysis. A linear regression model was used to assess the value of CP and group (TD vs. clinical) in predicting receptive vocabulary. A model including both main effects and their interaction showed the interaction term to be non-significant (*p* = 0.484), so that term was eliminated and the model re-fit. Group was a significant predictor of receptive vocabulary (*p* < 0.001) and accounted for 36% of the variance. CP did not predict receptive vocabulary (*p* = 0.204).

Figure 4.25: Scatterplot showing the relationship between CP and receptive language (BPVS).



Note: Participant EM14 is included in scatterplot at left, but not at right. Note that the R2 value changes considerably in the SLI+SSD group with data from EM14 eliminated, even though the relationship between the two variables is little affected in that group

When the linear regression model of the clinical groups alone was reconsidered (without data from EM14), neither group ($p = 0.100$) nor CP ($p = 0.258$) predicted receptive vocabulary.

4.6.2 Predicting NWR from receptive vocabulary

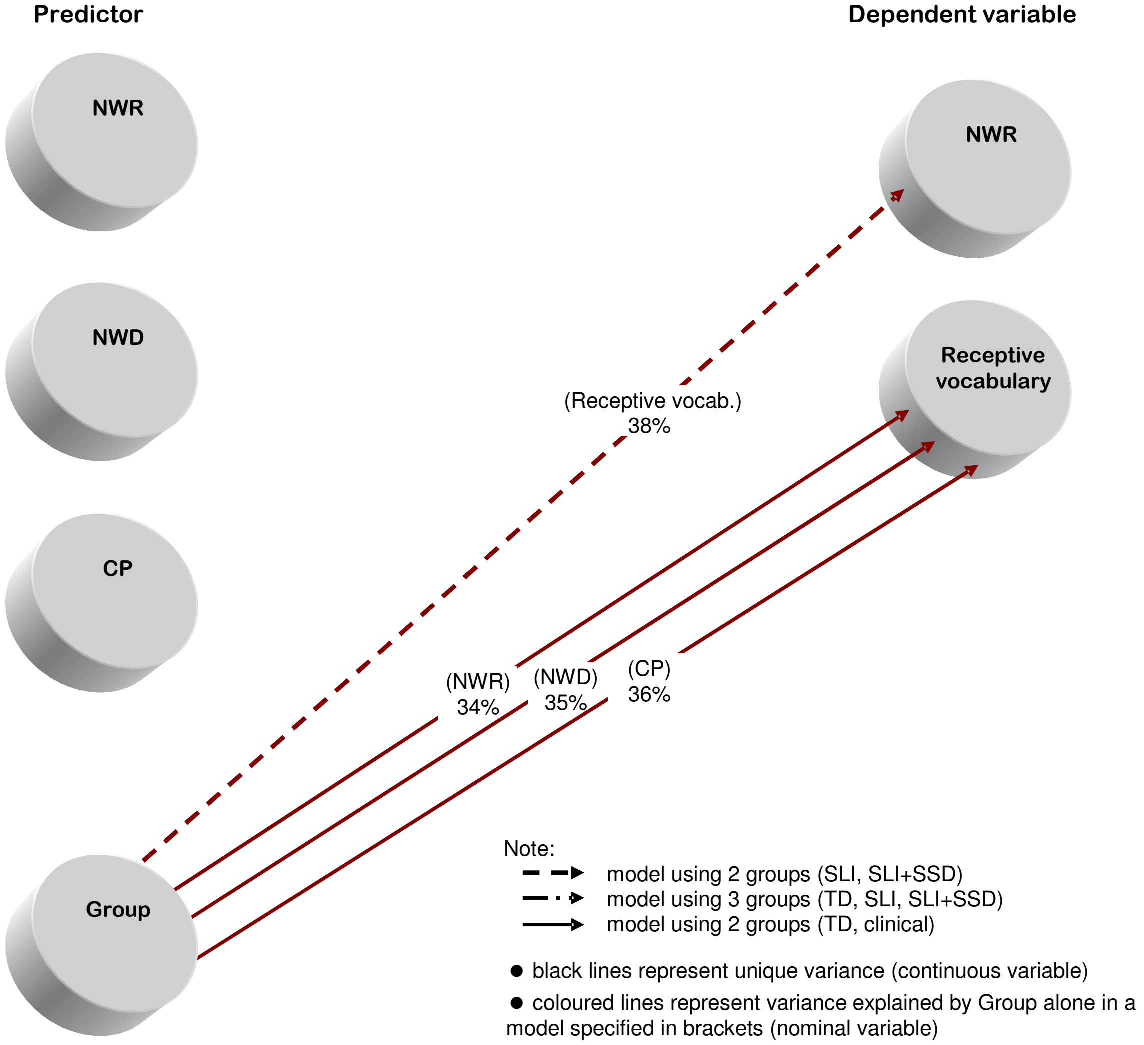
Some researchers suggest that the relationship between receptive vocabulary and NWR is unidirectional in children with SLI, with receptive vocabulary predicting repetition accuracy and NWR not predicting receptive vocabulary (as it is reported to do in typically developing children). A further analysis was undertaken to investigate the value of receptive vocabulary in predicting NWR within the clinical groups alone. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.817$), so that term was eliminated from the model and the model re-fit. Group predicted NWR ($p = 0.025$), accounting for 38% of the variance, but there was no main effect of receptive vocabulary ($p = 0.451$).

4.6.3 Summary

See Figure 4.26.

- In a model using two groups (TD vs. clinical), NWR did not predict receptive vocabulary. Group membership accounted for 34% of the variance. Neither group nor NWR predicted receptive vocabulary in the clinical groups alone.
- In a model using two groups (TD vs. clinical), NWD did not predict receptive vocabulary. Group membership accounted for 35% of the variance. Neither group nor NWD predicted receptive vocabulary in the clinical groups alone.
- In a model using two groups (TD vs. clinical), with data from one outlier eliminated, CP did not predict receptive vocabulary. Group membership accounted for 36% of the variance. Neither group nor CP predicted receptive vocabulary in the clinical groups alone.
- Receptive vocabulary did not predict NWR in the clinical groups alone. Group membership accounted for 38% of the variance.

Figure 4.26: Diagram illustrating relationships between variables and receptive vocabulary.



4.6.4 Phonological short term memory (PSTM)

In order to evaluate whether PSTM affected performance on the experimental tasks and sentence recall[5], regression analyses were carried out with Digit Recall as a predictor.

4.6.4.1 Predicting NWR

An attempt was made in the clinical groups alone to predict NWR from Digit Recall. There was a significant difference between the SLI and SLI+SSD groups ($p = 0.037$). A linear regression

---

[5] ‘Sentence recall’ equates to the Recalling Sentences subtest of the CELF.

model was used to assess the value of Digit Recall and group (TD, SLI, SLI+SSD) in predicting NWR. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.876$) so that term was eliminated and the model re-fit. Group was a significant predictor of NWR ($p < 0.001$) and there was weak evidence that Digit Recall predicted NWR ($p = 0.076$). This complete model accounted for 73% of the variance. Eliminating Digit Recall from the model showed group membership to account for 72% of the variance, implying that including Digit Recall accounted for only a further 1%. When the linear regression model of the clinical groups alone was reconsidered, group predicted NWR and accounted for 37% of the variance. Digit recall did not predict NWR in the clinical groups alone.

4.6.4.2 Predicting NWD

There was no difference between the SLI and SLI+SSD groups when Digit Recall was used as a predictor for NWD in the clinical groups alone ($p = 0.629$). These groups were combined into a single, ‘clinical’ group. A linear regression model was used to assess the value of Digit Recall and group (TD vs. clinical) in predicting NWD. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.727$), so that term was eliminated and the model re-fit. Group was a significant predictor of NWD ($p = 0.001$) and accounted for 40% of the variance. Digit Recall did not predict NWD ($p = 0.130$). Neither group ($p = 0.629$) nor Digit Recall ($p = 0.412$) predicted NWD in the clinical groups alone.

4.6.4.3 Predicting CP

When Digit Recall was considered as a predictor for CP in a model using only the clinical groups, there was no difference between the SLI and SLI+SSD groups ($p = 0.159$). These groups were again combined into a single, ‘clinical’ group. A linear regression model was used to assess the value of Digit Recall and group (TD vs. clinical) in predicting CP. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.400$), so that term was eliminated and the model re-fit. Digit Recall was a significant predictor of CP ($p < 0.001$) and accounted for 34% of the variance. Group did not predict CP ($p = 0.861$). The model using the clinical groups alone was reconsidered. Group did not predict CP ($p = 0.159$) but Digit Recall just reached significance ($p = 0.050$), accounting for 30% of the variance.

4.6.4.4 Predicting sentence recall

Finally, Digit Recall was considered as a predictor for sentence recall. In a model using only the clinical groups, there was no difference between the SLI and SLI+SSD groups ($p = 0.781$). These groups were again combined into a single, ‘clinical’ group. A linear regression model was used to assess the value of Digit Recall and group (TD vs. clinical) in predicting sentence recall. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.978$), so that term was eliminated and the model re-fit. Both group ($p < 0.001$) and Digit Recall ($p < 0.001$) were significant predictors of sentence recall and accounted for 64% of the variance. Eliminating Digit Recall from the model showed group membership to account for 55% of the variance implying that Digit Recall accounted for another 9%. The model using the clinical groups alone was reconsidered. Digit Recall ($p = 0.016$) but not group ($p = 0.781$) predicted sentence recall, accounting for 42% of the variance.

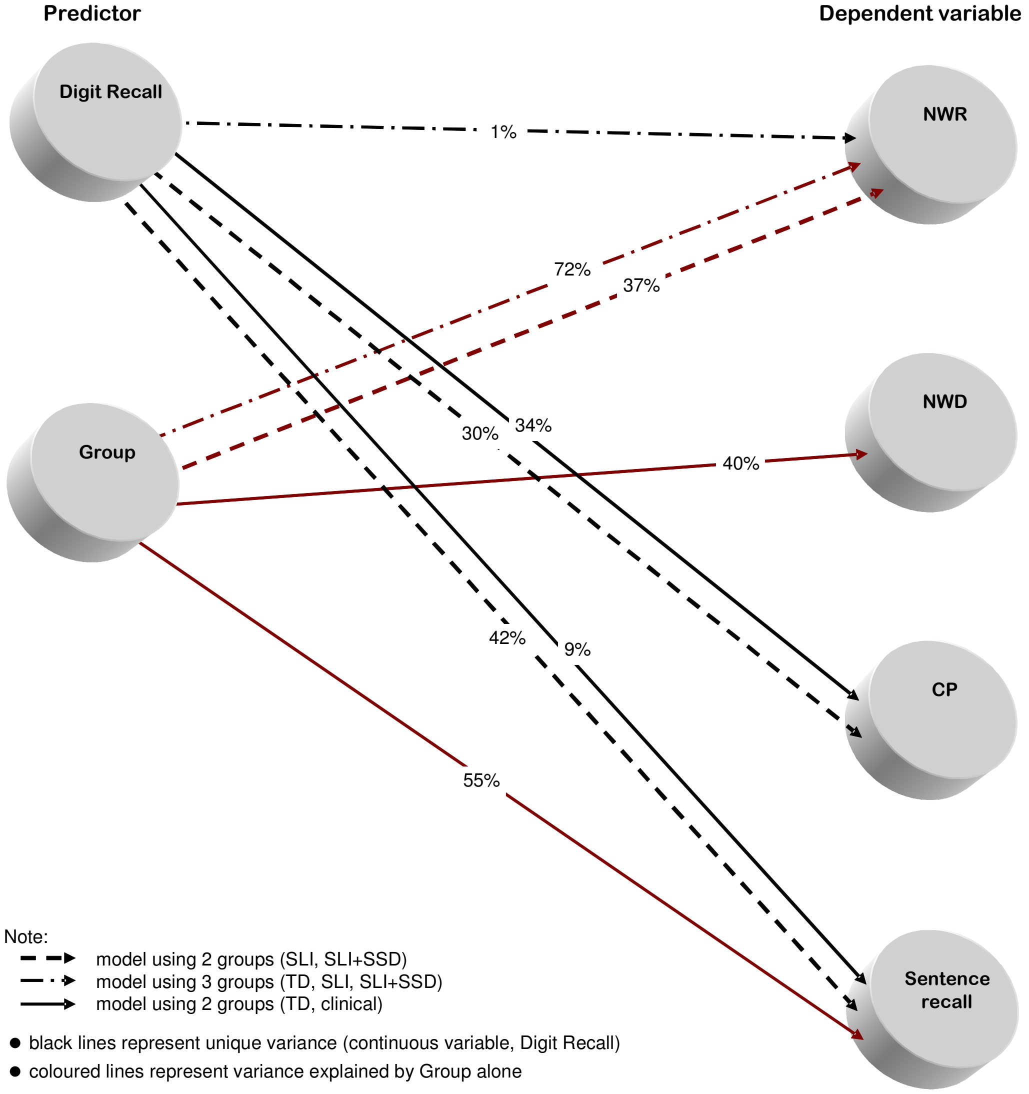
4.6.4.5 Summary

See Figure 4.27.

- In a model using three groups (TD, SLI, SLI+SSD), Digit Recall predicted NWR, accounting for 1% of the variance. Group membership accounted for 72% of the variance. Similarly, group but not Digit Recall predicted NWR in the clinical groups alone and accounted for 37% if the variance.
- In a model using two groups (TD vs. clinical), group predicted NWD but Digit Recall did not. Group accounted for 40% of the variance. Neither group nor Digit Recall predicted NWD in the clinical groups alone.

- In a model using two groups (TD vs. clinical), Digit Recall predicted CP, accounting for 34% of the variance. A similar result was found in the clinical groups alone where Digit Recall accounted for 30% of the variance. Group membership did not predict CP in either model.
- Digit Recall predicted sentence recall using two groups (TD vs. clinical), accounting for 9% of the variance. In a model comparing the clinical groups alone, Digit Recall accounted for 42% of the variance.

Figure 4.27: Diagram illustrating tasks predicted by digit recall and group.



4.7 Diagnostic markers for language and literacy difficulties?

Sentence recall has been proposed as a marker for SLI, so this measure was excluded from the composite language measure, and a mean language score computed from the remaining scores. A one-way ANOVA confirmed that the group differences described above did not change (see section 4.3.2). There was a significant main effect of group ($F_{(2,62)} = 43.352$, $p < 0.001$), with the TDs performing significantly better than both clinical groups ($p < 0.001$) and both clinical groups performing similarly ($p = 0.839$). Linear regressions were then carried out to assess the value of sentence recall and the experimental tasks in predicting language and literacy skills.

4.7.1 Predicting language skills

4.7.1.1 NWR as a predictor

An attempt was made in the clinical groups alone to predict language skills (revised composite language score) from NWR. There was no difference between the SLI and SLI+SSD groups ($p = 0.460$), so these groups were combined in the analysis to form a single 'clinical' group. A linear regression model was used to assess the value of NWR and group (TD vs. clinical) in predicting language skills. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.075$) so that term was eliminated and the model re-fit. Group was a significant predictor of language skills ($p < 0.001$) and accounted for 58% of the variance. NWR did not predict language skills ($p = 0.268$).

As this model was dominated by the performance of the TD controls, the model that included the clinical groups alone was reconsidered. A model including the main effects of group and NWR and their interaction showed the interaction term to be non-significant ($p = 0.714$) so that term was eliminated and the model re-fit. Neither group membership ($p = 0.460$) nor NWR ($p = 0.858$) predicted language skills in the clinical groups.

4.7.1.2 NWD as a predictor

There was no difference between the SLI and SLI+SSD groups when NWD was used as a predictor for language skills ($p = 0.485$). These groups were again combined into a single group. A linear regression model was used to assess the value of NWD and group (TD vs. clinical) in predicting language skills. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.385$), so that term was eliminated and the model re-fit. Group was a significant predictor of language skills ($p < 0.001$) and accounted for 60% of the variance. NWD did not predict language skills ($p = 0.086$). Reconsideration of the model including the clinical groups alone showed neither group membership ($p = 0.485$) nor NWD ($p = 0.263$) to be significant predictors of language skills in participants with SLI or SLI+SSD.

4.7.1.3 CP as a predictor

When CP was considered as a predictor for language skills, there was weak evidence of a difference between the SLI and SLI+SSD groups ($p = 0.082$). The clinical groups were not combined for this analysis. A linear regression model was used to assess the value of CP and group (TD, SLI, SLI+SSD) in predicting language skills. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.181$) so that term was eliminated and the model re-fit. Both group ($p < 0.001$) and CP ($p < 0.022$) were significant predictors of language skills, the complete model accounting for 62% of the variance. Eliminating CP from the model showed group membership to account for 58% of the variance, implying that including CP accounted for a further 4%.

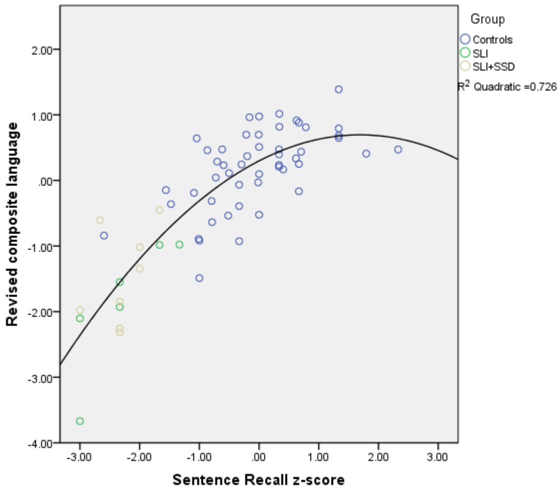
Given the large number of participants in the TD group, a model with just the clinical groups was again considered. As mentioned above, the interaction between group and CP was non-significant, so that term was eliminated from the model. The main effect of group was marginal ($p = 0.082$) but CP was a significant predictor or language skills ($p = 0.034$), the complete model accounting for 38% of the variance. Eliminating CP from the model showed group membership to account for 6% of the variance, implying that including CP accounted for a further 32% in the SLI and SLI+SSD groups.

4.7.1.4 Sentence recall as a predictor

An attempt was made in the clinical groups alone to predict language skills (revised composite language) from sentence recall. There was no difference between the SLI and SLI+SSD groups ($p = 0.274$), so these groups were combined in the analysis to form a single, 'clinical' group. A linear regression model was used to assess the value of sentence recall and group (TD vs. clinical) in predicting language skills. There was no main effect of group ($p = 0.393$), but sentence recall was a significant predictor of language skills ($p < 0.001$) and there was a significant interaction between sentence recall and group ($p = 0.026$). This complete model accounted for 74% of the variance.

However, inspection of the scatterplot suggested the relationship between sentence recall and language skills was not quite linear (see Figure 4.25). The plot seemed to flatten for high scores on sentence recall. Accounting for this with a quadratic term in the regression led to a model that accounted for 73% of the variance. In other words, a very good estimate of revised composite language can be obtained from sentence recall, without regard for the diagnostic group.

Figure 4.28: Scatterplot showing the relationship between language and sentence recall.



Note: A single quadratic regression curve is fitted to the plot as there is no main effect of group membership.

4.7.2 Predicting phonological skills

4.7.2.1 NWR as a predictor

An attempt was made in the clinical groups alone to predict phonological skills from NWR. There was no group difference between the SLI and SLI+SSD groups (*p* = 0.604), so these groups were combined in the analysis to form a single ‘clinical’ group. A linear regression model was used to assess the value of NWR and group (TD vs. clinical) in predicting phonological skills. A model including both main effects and their interaction showed the interaction term to be non-significant (*p* = 0.425) so that term was eliminated and the model re-fit. Both group (*p* < 0.001) and NWR were significant predictors of phonological skills (*p* < 0.015) and accounted for 62% of the variance. Eliminating NWR from the model showed group membership to account for 59% of the variance, implying that including NWR accounted for another 3%.

The model that only included the clinical groups was reconsidered. A model including the main effects of group and NWR and their interaction showed the interaction term to be non-significant (*p* = 0.137) so that term was eliminated and the model re-fit. Neither group membership (*p* = 0.604) nor NWR (*p* = 0.480) predicted phonological skills in the clinical groups.

4.7.2.2 NWD as a predictor

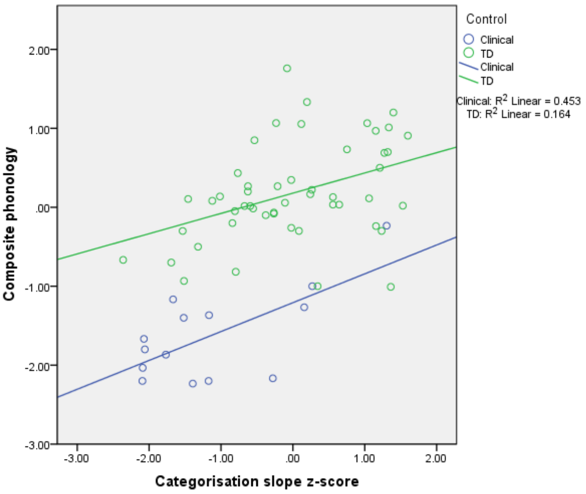
There was no difference between the SLI and SLI+SSD groups when NWD was used as a predictor for phonological skills in the clinical groups alone ($p = 0.091$). These groups were combined into a single, 'clinical' group. A linear regression model was used to assess the value of NWD and group (TD vs. clinical) in predicting phonological skills. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.732$), so that term was eliminated and the model re-fit. Both group ($p < 0.001$) and NWD ($p = 0.001$) were significant predictors of phonological skills, the complete model accounting for 65% of the variance. Eliminating NWD from the model showed group membership to account for 59% of the variance, implying that NWD accounted for another 6%.

The model that only included the clinical groups was reconsidered. A model including the main effects of group and NWD and their interaction showed the interaction term to be non-significant ($p = 0.099$) so that term was eliminated and the model re-fit. Group membership did not predict phonological skills ($p = 0.091$), but NWD was a significant predictor ($p = 0.016$), the complete model accounting for 49% of the variance. Eliminating group from the model showed NWD to account for 33% of the variance in the clinical groups.

4.7.2.3 CP as a predictor

When CP was considered as a predictor for composite phonology in a model using only the clinical groups as a fixed factor, there was no difference between the SLI and SLI+SSD groups ($p = 0.785$). These groups were again combined into a single, 'clinical' group. A linear regression model was used to assess the value of CP and group (TD vs. clinical) in predicting phonological skills. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.510$), so that term was eliminated and the model re-fit. Both group ($p < 0.001$) and CP ($p < 0.001$) were significant predictors of phonological skills, the complete model accounting for 68% of the variance. Eliminating CP from the model showed group membership to account for 60% of the variance implying that including CP accounted for another 8% (see Figure 4.29).

Figure 4.29: Scatterplot showing the relationship between phonology and CP



Note: Plot for model using two groups (TD and combined clinical group).

The model that included only the clinical groups was reconsidered. A model including the main effects of group and CP and their interaction showed the interaction term to be non-significant (*p* = 0.807) so that term was eliminated and the model re-fit. There was no main effect of group (*p* = 0.807) but CP was a significant predictor of phonological skills (*p* = 0.023). Eliminating group from the model showed CP to account for 45% of the variance in the clinical groups.

4.7.3 Predicting reading skills

Finally, the experimental tasks were evaluated as predictors for reading skills (composite reading score). Control data was only available for the participants tested in Experiment 2 (n = 21).

4.7.3.1 NWR as a predictor

An attempt was made in the clinical groups alone to predict reading skills from NWR. There was no group difference between the SLI and SLI+SSD groups (*p* = 0.151), so these groups were combined in the analysis to form a single ‘clinical’ group. A linear regression model was used to assess the value of NWR and group (TD vs. clinical) in predicting reading skills. A model including both main effects and their interaction showed the interaction term to be non-significant (*p* = 0.258) so that term was eliminated and the model re-fit. Group did not predict reading skills (*p* = 0.503) however the main effect of NWR fell just short of significance (*p* = 0.055). This model accounted for 40% of the variance and eliminating group showed NWR to account for 39% of the variance. NWR therefore predicted reading skills irrespective of group membership.

4.7.3.2 NWD as a predictor

An attempt was made in the clinical groups alone to predict reading skills from NWD. There was a significant difference between the SLI and SLI+SSD groups (*p* = 0.002). A linear regression model was used to assess the value of NWD and group (TD, SLI, SLI+SSD) in predicting reading skills. A model including both main effects and their interaction showed the interaction term to be non-significant (*p* = 0.605) so that term was eliminated and the model re-fit. Both group (*p* = 0.003) and NWD (*p* = 0.001) were significant predictors of reading skills, the complete model accounting for 63% of the variance. Eliminating NWD from the model showed group membership to account for 45% of the variance, implying that including NWD accounted for another 18%.

The model that included only the clinical groups was reconsidered. A model including the main effects of group and NWD and their interaction showed the interaction term to be non-significant (*p* = 0.605) so that term was eliminated and the model re-fit. Both group (*p* = 0.002) and NWD were significant predictors of reading, the complete model accounting for 71% of the variance. Eliminating NWD from the model showed group membership to account for 36% of the variance, implying that including NWD accounted for a further 35% in the clinical groups.

4.7.3.3 CP as a predictor

When CP was also considered as a predictor for reading skills in the clinical groups alone, there was no difference between the SLI and SLI+SSD groups (*p* = 0.081). These groups were combined to form a single ‘clinical’ group. A linear regression model was used to assess the value of CP and group (TD vs. clinical) in predicting reading skills. A model including both main effects and their interaction showed the interaction term to be non-significant ($p = 0.389$), so that term was eliminated and the model re-fit. Both group ($p = 0.003$) and CP ($p < 0.020$) were significant predictors of reading skills and accounted for 43% of the variance. Eliminating CP from the model showed group membership to account for 33% of the variance implying that CP accounted for another 10%.

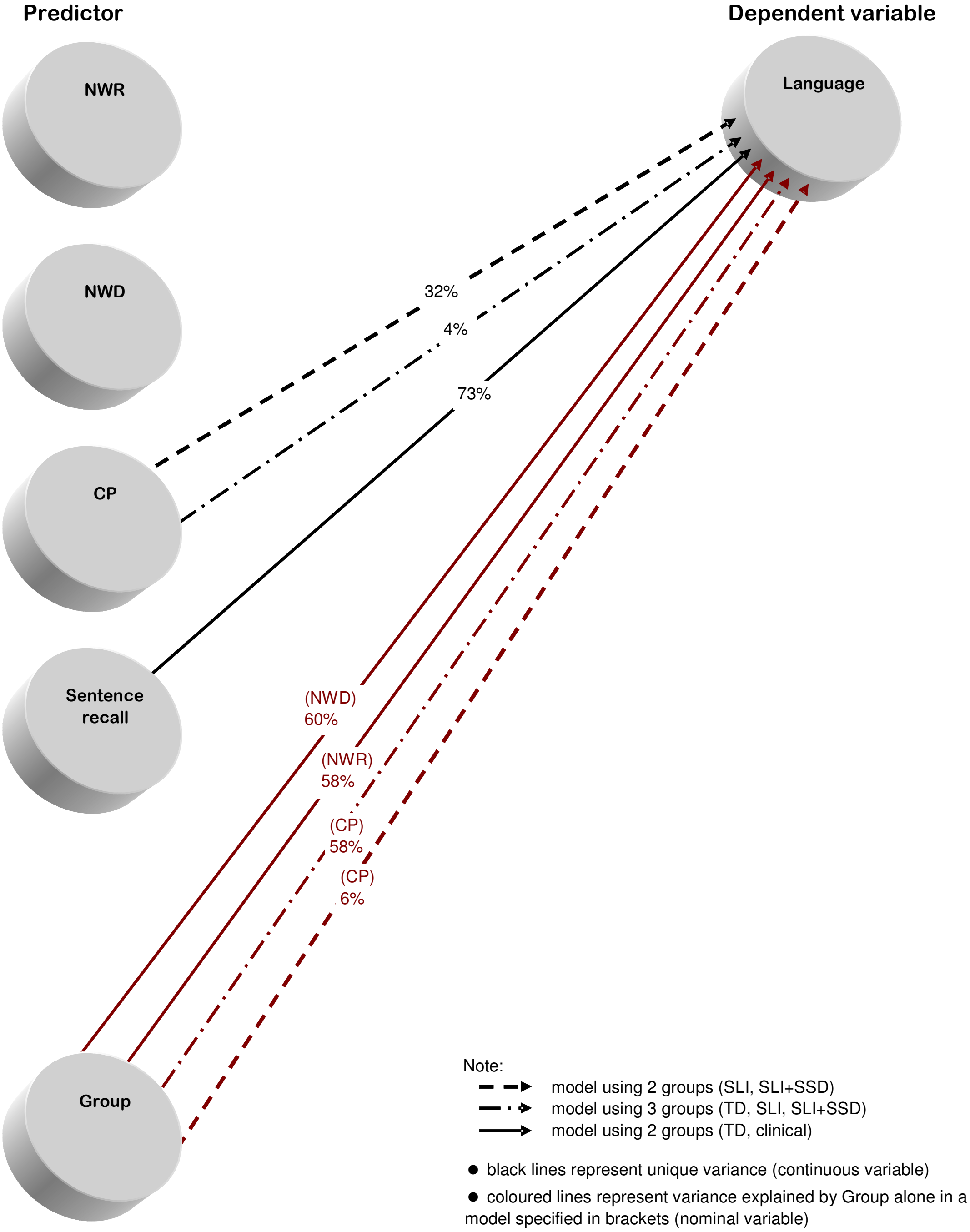
Finally, the model that included only the clinical groups was reconsidered. A model including the main effects of group and CP and their interaction showed the interaction term to be non-significant ($p = 0.880$) so that term was eliminated and the model re-fit. Neither group ($p = 0.081$) nor CP ($p = 0.224$) predicted reading in the clinical groups alone.

4.7.4 Summary

Predicting language skills (see also Figure 4.30):

- Using a quadratic term in the regression, sentence recall predicted language skills without regard for diagnostic group. Sentence recall accounted for 73% of the variance.
- In a model using two groups (TD vs. clinical), NWR did not predict language skills but group membership accounted for 58% of the variance. Neither NWR nor group membership predicted language skills in the clinical groups alone.
- In a model using two groups (TD vs. clinical), NWD did not predict language skills but group membership accounted for 60% of the variance. Neither NWD nor group membership predicted language skills in the clinical groups alone.
- In a model using three groups (TD, SLI, SLI+SSD), both group membership and CP predicted language skills. Group accounted for 58% of the variance and CP accounted for a further 4%. In the clinical groups alone, group membership accounted for 6% of the variance and CP accounted for a further 32%.

Figure 4.30: Diagram illustrating predictors of language skills.



Note:

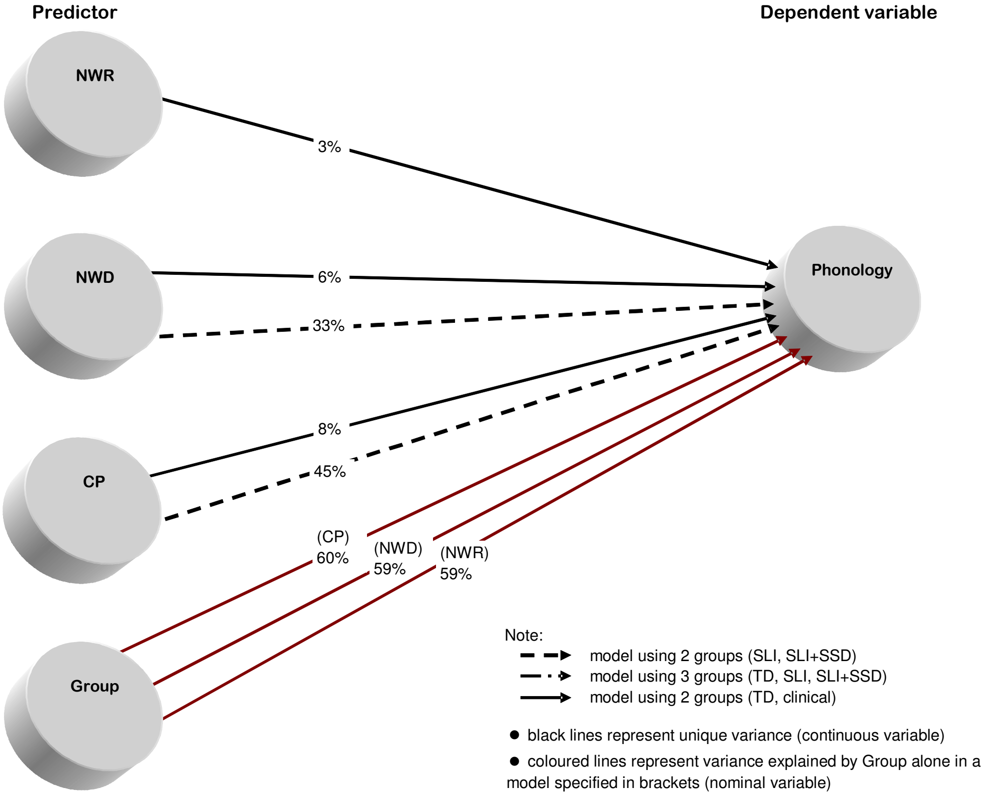
- – – ▸ model using 2 groups (SLI, SLI+SSD)
- —·—▸ model using 3 groups (TD, SLI, SLI+SSD)
- ——▸ model using 2 groups (TD, clinical)

- ● black lines represent unique variance (continuous variable)
- ● coloured lines represent variance explained by Group alone in a model specified in brackets (nominal variable)

Predicting phonological skills (see Figure 4.31):

- In a model using two groups (TD vs. clinical) both group and NWR predicted phonological skills. Group membership accounted for 59% of the variance and NWR accounted for a further 3%. Neither group nor NWR predicted phonological skills in the clinical groups alone.
- In a model using two groups (TD vs. clinical) group membership and NWD predicted phonological skills. Group accounted for 59% of the variance and NWD accounted for a further 6%. In the clinical groups alone, NWD (but not group) predicted phonological skills and accounted for 33% of the variance.
- In a model using two groups (TD vs. clinical), both group membership and CP predicted phonological skills. Group accounted for 60% of the variance and CP accounted for a unique 8%. Group did not predict phonological skills in the clinical groups alone, but CP accounted for a unique 45% of the variance.

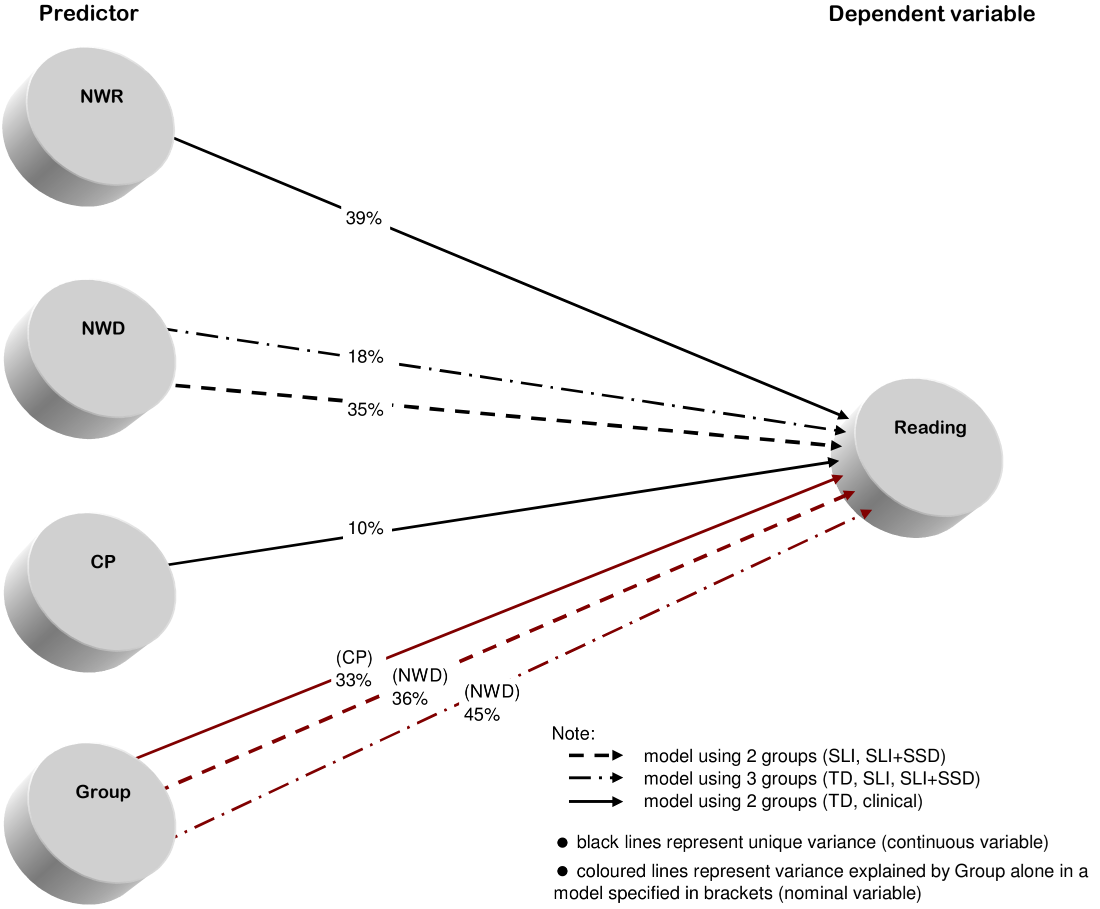
Figure 4.31: Diagram illustrating predictors of phonology skills.



Predicting reading skills (see Figure 4.32):

- In a model using two groups (TD vs. clinical) NWR predicted reading skills and accounted for 39% of the variance. There was no effect of group membership.
- In a model using three groups (TD, SLI, SLI+SSD) group membership and NWD predicted reading skills. Group accounted for 45% of the variance and NWD accounted for a unique 18%. In the clinical groups alone, group accounted for 36% of the variance and NWD accounted for a further 35%.
- In a model using two groups (TD vs. clinical) both group and CP predicted reading skills. Group membership accounted for 33% of the variance and CP accounted for a further 10%. Neither group nor CP predicted reading skills in the clinical groups alone.

Figure 4.32: Diagram illustrating predictors of reading skills.



- black lines represent unique variance (continuous variable)
- coloured lines represent variance explained by Group alone in a model specified in brackets (nominal variable)

4.7.5 Sensitivity and specificity

Most of the regression analyses showed group to be the primary predictor of each dependent variable. For this reason, a post-hoc analysis of sensitivity and specificity was carried out as an alternative assessment of the validity of the experimental tasks and sentence recall as markers of linguistic impairments.

In order to be considered a reliable binary marker, the behaviour being tested should be present in individuals who have a particular disorder and absent in those who do not. The validity of a diagnostic test is usually measured against reference ('gold') standards that are widely accepted as identifying the disorder. *Sensitivity* refers to the proportion of individuals who are correctly identified as having the disorder (positive diagnosis), and *specificity* refers to the proportion of individuals without the disorder who are correctly identified (negative diagnosis). There are no accepted thresholds for these measures, but some researchers suggest that both sensitivity and specificity should be above 80% (Plante & Vance, 1994).

The sensitivity and specificity of the experimental tasks and sentence recall in identifying participants with language and literacy deficits was investigated. Group membership was eliminated as a confound, the data instead being considered in terms of each participant's performance profile.

The composite scores and the z-scores for each proposed marker were converted to binary measures that indicated the presence or absence of an impairment. A participant was considered to be impaired on a particular measure if his/her composite or z-score fell more than 1.25 s.d. below the mean (equivalent to <10th percentile). This cut-off was used as it is often applied in research and clinical contexts (e.g. Tomblin, et al., 1997)[6]. Z-scores have a population mean of zero and a standard deviation of one, so in order to be considered

---

[6] A cut-off following the WHO criterion of -2.00 s.d. from the mean resulted in only 4 of the participants being identified as impaired on the composite language measure.

unimpaired a participant had to achieve a score above -1.25. The composite measures were considered to be the reference standards by which impairments could be identified reliably.

The sensitivity of the proposed markers was calculated using the equation: number impaired on proposed marker/ total number impaired on composite measure. The specificity of the proposed markers was calculated using the equation: (number unimpaired on proposed marker)/(total unimpaired on composite measure).

Finally, the likelihood ratio (LR) was calculated using the equation: sensitivity/(1 – specificity). This statistic reflects the ratio between the probability that a positive test comes from an individual with the disorder (true positive), and the probability that a positive test comes from an individual without the disorder (false positive). The LR does not depend on the prevalence of the disorder in the population. A large, positive LR indicates that a positive result is more likely to be true. LRs above 10 are considered to indicate an often conclusive increase in the likelihood of a disorder and LRs between 5 and 10 indicate a moderately strong increase (Jaeschke, Guyatt, & Lijmer, 2002).

Table 4.4: Sensitivity and specificity of experimental tasks and sentence recall (SR).

	Composite measure								
	+		-						
	Screening task								
	+	-	+	-	Total +	Total -	Sensitivity (%)	Specificity (%)	Likelihood
Language (without SR)									
NWR	9	1	13	42	10	55	90.0	76.4	3.8
NWD	7	3	9	46	10	55	70.0	83.6	4.3
CP	6	4	9	46	10	55	60.0	83.6	3.7
SR	9	1	9	46	10	55	90.0	83.6	5.5
Phonology									
NWR	11	0	11	43	11	54	100.0	79.6	4.9
NWD	9	2	7	47	11	54	81.8	87.0	6.3
CP	7	4	8	46	11	54	63.6	85.2	4.3
SR	11	0	7	47	11	54	100.0	87.0	7.7
Reading									
NWR	6	1	11	17	7	28	85.7	60.7	2.2
NWD	6	1	6	22	7	28	85.7	78.6	4.0
CP	4	3	9	19	7	28	57.1	67.8	1.8
SR	7	0	8	20	7	28	100.0	71.4	3.5

Note: + = positive result (presence of impairment)
- = negative result (absence of impairment)

Table 4.4 shows that for a cut-off of -1.25 s.d., only sentence recall (SR) was a moderately reliable identifier of participants with impaired language skills. SR had a sensitivity of 90%, a specificity of 84% and a likelihood ratio (LR) of 5.5. For NWR, sensitivity was equal to that for SR (90%), but specificity was weaker (76%), indicating that using NWR as a diagnostic tool would result in a number of false positives. NWD and CP, on the other hand, were equal to SR in specificity (84%), but both measures had sensitivity of only 70% and 60% respectively. This means that NWD and CP would result in a higher proportion of false negatives than the other screening measures.

A phonological impairment was moderately reliably identified by both NWD and SR (LRs of 6.31 and 7.71 respectively). The sensitivity of NWR was equal to that of SR (100%), but NWR resulted in more false positives (specificity of 80%) and its LR fell just short of reliability (4.91).

The specificity of most screening measures was relatively low for reading impairment, which resulted in a high proportion of false positives. NWD almost reached a specificity of 80%, and with a sensitivity of 86%, came close to being a reliable marker for reading impairment (LR = 4.00). NWR had the same sensitivity as NWD (86%), but was not as reliable at identifying children without a reading impairment. The sensitivity of Sentence Recall was 100% for reading impairment, but the reliability of this measure was compromised by specificity of only 71% (LR = 3.50).

In the sensitivity/specificity analysis, the criterion for impaired language, phonology or reading was a composite z-score of -1.25 s.d. from the mean. The analysis was carried out without regard to experimental group. Table 4.5 shows that when the -1.25 s.d. cut-off was applied to the TD group, one participant was identified as language impaired and one was identified as reading impaired. There were a total of 14 participants in the clinical groups, all of whom had a diagnosis of SLI (either as a single deficit or co-morbid with SSD), but only nine were considered to be language impaired according to the criterion used here. According to the measures used here, of the 14 clinical participants, 11 had a phonological impairment and six were reading impaired.

Table 4.5: Group breakdown of impairments identified by -1.25 s.d. criterion.

	Language		Phonology		Reading	
	+	-	+	-	+	-
Controls	1	50	0	51	1	20
SLI	4	2	4	2	1	5
SLI+SSD	5	3	7	1	5	3

Note: + = positive result (presence of impairment)
- = negative result (absence of impairment)

4.8 Discussion

This experiment aimed to investigate the speech processing skills of two groups of language impaired participants: those with SLI and those with SLI and a co-morbid speech sound disorder (SLI+SSD). Data from both clinical groups were compared to a group of typically developing (TD) participants. A range of analyses was undertaken to investigate the performance of each group on a battery of language and literacy measures and three experimental tasks: nonword repetition (NWR), nonword discrimination (NWD) and categorical perception (CP). The reliability of these tasks as markers of language, phonological and literacy impairments was evaluated alongside a published measure of sentence recall.

4.8.1 Linguistic measures

4.8.1.1 Typically developing (TD) participants

The TD data was typical of the general population for WISC, TROG and Rhyme, but this group performed above average on Recalling Sentences, BPVS, Digit Recall and Nonword Reading. When the TD participants were recruited, members of the school staff were given instructions to invite a random selection of children from the required age group. It was noted at the time of testing that only one head teacher overtly implemented randomised selection, namely by inviting every alternate child on each class register to participate. Thus selection in that school was based on alphabetic ordering of surname. It is possible that TD participants from other schools were invited to participate on the basis of academic performance, with class teachers judging higher achievers better able to miss time in class. This selection strategy is perhaps understandable when considered from the class teacher's perspective. Anecdotal evidence suggests this experience is common to many researchers in this field. It is possible that a sampling bias within the TD group may have had some effect on the comparative analyses reported here and in published studies.

4.8.1.2 Clinical groups

The speech of the SLI+SSD group, while impaired compared to TDs, was sufficiently intelligible to allow valid scoring on all published and informal linguistic measures. The language and phonology skills of participants in both clinical groups were similarly impaired when compared to TDs. The SLI+SSD group was impaired relative to the TD group on reading, but the SLI group was not.

The composite performance measures show that the language and phonology skills of the SLI and SLI+SSD groups were comparable. However the reading skills of the SLI group were better than those of the double deficit SLI+SSD group. Five out of eight participants in the SLI+SSD group achieved a composite reading score below -1.25 s.d. from the mean. This compares to only one SLI participant out of six who achieved a similarly low reading score.

At first glance, the difference in reading performance can be explained by an artefact of scoring. Any expressive speech error would likely cause a nonword reading response to be scored as incorrect even if the participant was able to read the nonword but the sounds it contained were not in his/her speech sound inventory. However, this explanation seems less likely when the composite score is broken down. There was no difference between the clinical groups on the published nonword reading task yet the SLI+SSD group performed worse than the SLI group on the informal reading screen, which presented real words and letters in addition to nonwords.

If the reading measures do reflect a true group difference, the data replicates that of studies that have found double deficit SLI+SSD children have worse literacy outcomes than children with a single deficit disorder (e.g. Lewis & Freebairn, 1992).

If scoring had been an issue in the nonword reading task, it should also have been reflected in a group difference on the grammar measures when speech errors could have been mistaken for morphological errors. However the clinical groups performed equally on all expressive grammar tasks. This finding did not replicate studies that have found a worse grammatical deficit in double deficit SLI+SSD children (Haskill & Tyler, 2007).

There was no difference between the clinical groups in terms of the published measures of phonology, yet the SLI+SSD group performed worse on NWR. Again, maybe an artefact of scoring was responsible for this pattern of results. It is possible that SLI+SSD responses on digit recall and rhyme were scored favourably despite speech errors. Target responses on those tasks were real words so speech errors would not have masked plausible linguistically correct attempts. There is no room for interpretation on NWR so all responses containing speech errors were scored incorrect. This identifies a need to describe all speech deficits in any future work so that speech errors can be teased apart from linguistic errors on nonword tasks.

4.8.2 Experimental measures (clinical groups)

At a task level, there were clear differences in group performance on CP, NWD and NWR. The SLI+SSD group was impaired relative to TDs at the level of CP while, as predicted, the SLI group was not. There was a trend for the SLI+SSD participants to perform worse than SLIs on CP. Both clinical groups were impaired on the NWD and NWR tasks, their impairment being comparable on NWD. Unsurprisingly, the SLI+SSD group was more impaired than the single deficit SLI group on NWR.

When task performance was compared within the clinical groups, each group had a distinct profile[7]. The SLI group's categorical perception was better than discrimination and repetition of nonwords. There was no difference in SLI performance on the latter two tasks. In contrast, the SLI+SSD group performed similarly on CP and NWD, and were significantly better at those tasks than NWR.

---

[7] Within group analyses of TD performance were not carried out. Z-scores had been calculated on the basis of TD raw scores, so the mean and standard deviation for each measure was constrained in the TD group.

It may be that participants in the SLI+SSD group were more sensitive to acoustic cues in the CP task, and were therefore less categorical than SLI participants in their perception. Alternatively, speech production may play a role in establishing phonological categories, so that the SLI+SSD group were less able to generate accurate representations of the contrasted phonemes in the CP task.

There is evidence that language impaired children with a reading deficit have phonological processing problems, while SLI-only children do not (Bishop, et al., 2009; Catts, et al., 2005). The SLI group was unimpaired on the reading composite reported here, which may explain the group's performance on the CP task. It does not explain, however, why the SLI children were impaired relative to TDs on NWD.

4.8.3 How did the clinical groups process nonwords?

CP is arguably the most basic linguistic task presented here, as it focuses on phonemes which are the smallest units of language (Coady, et al., 2005). Categorical perception of phonemes occurs at an early stage of speech processing and must be involved in more complex tasks such as NWD and NWR. The CP task presented here focused on the presence or absence of a cluster in the single syllable nonwords /kep/ and /klep/. The stimulus continuum was generated from natural speech tokens of the nonwords, which were also used in the NWD and NWR tasks.

NWD and NWR errors could be caused by a deficit at a basic processing level. The assumption was that if a participant was unimpaired on CP, any deficit observed in discrimination or repetition could not be due to difficulties in perception at the level of the phoneme contrast tested in the CP task (a contrast that also appeared in many of the NWD and NWR stimuli). It should be noted that only one phonological contrast was presented in the CP task, so general claims about auditory processing cannot be made.

The SLI+SSD group was impaired on the CP task relative to TDs and they performed marginally worse than the single deficit SLI group. The SLI+SSD and SLI groups performed comparably on NWD. These seem to be conflicting results: if the SLI participants were better at categorising phonemes than the SLI+SSDs, why did they not perform better than the dual deficit group on the NWD task?

The syllable and error data for the NWD task offer some insight into this pattern of performance. SLI participants performed comparably to TDs on one syllable stimuli while SLI+SSD participants performed significantly worse than TDs. In the case of NWD stimuli consisting of three syllables, the SLI group performed as poorly as the SLI+SSD group. It may be that a more severe, underlying perceptual deficit caused the SLI+SSD group to perform poorly on both CP and NWD. The perceptual skills of the SLI group, meanwhile, may have been stressed only by the longer, more complex nonwords in the NWD task. The single syllable nonwords in the CP task may not have overloaded the SLI processing system. Coady (2005) reports that task demands likely affect children with SLI in CP tasks, so it is reasonable to assume that the increased demands of NWD affected the SLI group in the same way. Both clinical groups showed comparably impaired phonological short term memory (PSTM), as indexed by Digit Recall scores, when compared to the TD group. It is unlikely that a qualitative difference in memory could explain the group differences in perception unless PSTM in the SLI group was stressed by increasing task demands.

When the data from NWR are considered, a similar picture emerges for the SLI group. As in the NWD task, the SLI group was able to repeat single syllable nonwords comparably to the TD group. In replication of previous studies (e.g. Bishop, et al., 1996; Dollaghan & Campbell, 1998), the SLI group only showed an impairment on the longer, three syllable stimuli. Just as Bishop et al. (1996) found that articulatory complexity affected NWR in children with SLI, the SLI group was unimpaired compared to TDs on three syllable stimuli containing no cluster but performed as poorly as the SLI+SSD group on nonwords with one or two clusters. It seems that SLI performance decreased with the increasing processing demands of length and complexity.

Length also affected the SLI+SSD group on NWR. This group was able to repeat single syllable nonwords more accurately than nonwords containing three syllables but, unlike the single deficit

SLI group, SLI+SSD participants were impaired compared to TDs even on the shorter stimuli. There was evidence that cluster number in three syllable nonwords affected the SLI+SSD group, but participants in this group also had difficulty repeating three syllable nonwords with no cluster. Clusters and polysyllables require more complex articulatory procedures, so taken in isolation, the NWR performance of the SLI+SSD participants could be explained purely by an output deficit. However this group also showed a perceptual deficit. Was there an interaction between input and output processing?

4.8.4 A perception-production link?

There are many reports of perception preceding production in the normal population. However, there is also evidence that impaired speech production affects perception. Bishop et al. (1990) found that children with cerebral palsy were impaired on phoneme discrimination in nonwords. The authors proposed that overt or covert repetition facilitates retention of nonword (therefore novel) stimuli, and that the inability to recode the input hinders performance on a discrimination task. Whitehill et al. (2003) reported similar findings with children with cleft palate, whereby Cantonese speaking participants had difficulty in a perceptual task involving velar sounds which they could not produce. Groenen et al. (1996) demonstrated the interdependence of perception and production by showing that the frequency of phoneme substitution errors made by children with dyspraxia was related to their performance on CP tasks. Moreover, Sahlen et al. (1999) found that output phonology was the most important predictor of NWR in five year old children with language impairment.

The data presented here seem consistent with a model in which production influences perception. In common with Experiment 1, NWD and NWR were related to one another, indicating that the two tasks shared underlying cognitive processes. Regression analyses showed that the perceptual task (NWD) did not predict the production task (NWR) to any convincing degree. NWD accounted for 4% of the variance when all groups were considered and 16% of the variance in the clinical groups alone. This suggests that a perceptual deficit would not significantly affect performance on NWR. However, when NWR was considered as a predictor for NWD, there was more robust evidence that output processing affected performance on the perceptual task. NWR accounted for 38% of the variance in NWD and group membership was non-significant in any model. The CP data showed a similar pattern, with NWR accounting for 25% of the variance in the CP task. CP did not predict performance on NWD and at best accounted for 4% of the variance in NWR.

It seems, then, that speech production skills influenced speech perception for all participants, regardless of group. Participants were instructed not to repeat the stimuli on the NWD and CP tasks, however there were instances when participants silently mouthed nonwords. This observation lends weight to the argument that motor programmes were accessed on at least some trials of the perceptual tasks.

The NWD and NWR error data seem to fit with this explanation. The only within group difference was for C2 movement errors in the SLI+SSD group. This group made more perceptual than repetition errors of this type. If a unidirectional transfer of degraded input phonological representations was leading to incorrect output representations, one would expect the same proportion or more repetition errors to occur. However if the links between input and output representations were bidirectional, the additional influence of degraded output representations would result in a higher proportion of discrimination than repetition errors, as occurred here.

An alternative explanation for the relationship between NWR and NWD follows. NWD did not predict NWR but NWR did predict NWD. It may be that NWR, as the more complex task, involving maximal sub-lexical stages of processing, encompassed NWD. This would likely explain the finding that NWR predicted performance on NWD. If it is assumed that NWD involves fewer (all input) stages of processing than NWR, NWD is unlikely to predict performance on NWR.

4.8.5 The locus of the processing deficit in SLI and SLI+SSD

The data suggest that while the clinical groups had comparable language, phonology and phonological short-term memory (PSTM) skills, their speech processing was qualitatively and quantitatively different. As far as identifying the locus of the groups' processing deficits is concerned, top-down lexical 'bootstrapping' can be ruled out of all three experimental tasks. There was no relationship between receptive vocabulary and performance on NWR, NWD or CP. Of particular interest, and in common with the results reported in Experiment 1, NWR did not predict receptive vocabulary in the TD group, and receptive vocabulary failed to predict NWR in the clinical groups. The latter finding replicates two studies that did not find a relationship between receptive vocabulary and NWR in SLI (Briscoe, et al., 2001; Edwards & Lahey, 1998). It seems that the current CP, NWD and NWR tasks all tapped sub-lexical processing.

A possible explanation for the pattern of performance in the SLI group follows. The SLI group seemed to be able to generate adequate input sublexical phonological representations when task demands were low as on the CP task. Past a critical level, task demands stressed the SLI perceptual system so that it was unable to generate or access adequate input phonological representations and performance on NWD and NWR suffered. Children with SLI often have word finding difficulties (WFDs) which means that they have difficulty generating output programmes for familiar items that do not yet have a precise output phonological representation (for a summary of WFDs in children, see Messer & Dockrell, 2006). It seems reasonable to assume, then, that this behaviour would manifest in an 'online' processing situation for novel stimuli. The transfer of such imprecise output information may have affected performance on the NWD task.

There are two plausible explanations for the pattern of performance observed in the SLI+SSD group. Firstly, an underlying perceptual deficit degraded both input and output phonological representations, and caused increasing impairment as the stimuli and task demands became greater (from CP to NWD to NWR). In this case, the number of NWR errors was likely exacerbated by articulation errors independently of input processing. Alternatively, the SLI+SSD group were affected by increasing task demands just like the SLI group, but their input phonological representations were degraded by inaccurate output representations to such an extent that performance on the simplest tasks and stimuli was affected. Given that the linguistic and phonological deficit in the SLI+SSD group was the same as that in the single deficit SLI group, the latter explanation seems more likely. An underlying perceptual deficit would likely have further impaired performance on the linguistic assessment battery too.

4.8.6 Implications for diagnosing and treating SLI

Before discussing results from the sensitivity/specificity analysis, it should be noted that the analysis was carried out without regard to experimental group and the criterion for impaired language, phonology or reading was set at a composite z-score of -1.25 s.d. from the mean. According to this criterion, only nine of the 14 participants in the clinical groups were identified as being language impaired. It is unlikely that this cut-off was set too low, as -1.25 s.d. is one of the more conservative criteria for diagnosing language impairment from composite performance scores. A more plausible explanation is that the composite used here included scores from both receptive and expressive assessments and may have masked participants with only expressive difficulties. Such children may have achieved higher scores on the receptive tasks which mitigated lower performance on expressive assessments when the composite was calculated. Alternatively, it is not unreasonable to assume that some of the clinical participants presented with a pattern of deficits whereby phonological skills were more affected than the language skills assessed here, or that others were ‘borderline’ impaired on the language measure.

Sentence recall (SR) was undoubtedly the most powerful of the diagnostic tools evaluated (NWD, NWR, CP and SR). This measure reliably identified most of the participants with impaired language skills and all of those with impaired phonological skills (90% and 100% respectively). SR also had moderately high specificity for the language and phonology composites (84% and 87% respectively). SR accounted for high unique variance in a regression model predicting language skills.

These findings are in line with previous reports that SR relies on both language and memory abilities, and can be used to identify deficits in both domains (Archibald & Joanisse, 2009).

Indeed, the results presented here suggest that SR was a better marker for phonological than language deficits. The phonological composite comprised Digit Recall (a relatively ‘pure’ measure of phonological short-term memory [PSTM]) and Rhyme, which has a high reliance on PSTM. Regression analyses confirmed the relationship between SR and phonological working memory. Digit Recall predicted SR, accounting for 9% of the variance in a two group model (TD vs. clinical) and 42% of the variance in the clinical groups alone.

SR also identified all participants who had impaired reading skills, but its reliability as a marker for literacy difficulties was compromised by a sensitivity of only 71%. There is consistent evidence for the presence of a phonological deficit in children with reading impairment. Poor phonological awareness and PSTM, the very measures included in the composite here, have been observed in children with dyslexia (e.g. Mann & Liberman, 1984). Given its sensitivity to phonological impairment in this analysis, it is likely that SR also tapped the phonological deficit that characterises reading impairment. A closer look at individual performance profiles adds weight to this argument. All clinical participants who were impaired on reading were also impaired on phonology.

Within the clinical groups at least, it seems that the reading composite measured underlying phonological skills. The only participant to be impaired on reading and not phonology was the youngest female participant who was a member of the typically developing (TD) group. At 4;1, she understandably achieved a low score on the informal reading screen, which included letter identification and real and nonword digraphs. The only other participant of a similar age (4;0) performed significantly better on the reading screen, which likely reflects more advanced experience in the learning environment and/or home. The TD participant who was impaired on the reading measure was unimpaired on phonology (and language).

The most reliable marker of reading impairment in the sensitivity/specificity analysis was NWD, which had a LR of 4.00. NWD was also the best marker of phonological impairment after SR. Linear regression showed that NWD predicted reading and phonological skills in the clinical groups, accounting for 35% and 33% of the variance respectively. However, while SR seemed to tap underlying phonological skills including PSTM, NWD was unrelated to Digit Recall.

Both length and complexity have implications for diagnostic tools that use nonwords to identify language impairment. It seems that words of more than one syllable, including clusters should be used in single word tasks such as nonword repetition (NWR). If simple, single syllable nonwords are used in an assessment, there may be a risk that a language impairment will be unidentified.

A nonword is a novel speech stimulus and must be processed in the same way as a new word. Consequently, nonwords are useful tools for teaching word learning skills. The current nonword set accessed skills at a sub-lexical level that are essential during the word learning process. Such materials could be used to develop input and output processing pathways that would improve the precision of sub-lexical phonological representations. These nonwords could not, however, be used to develop lexical access.

The three experimental tasks used nonword stimuli that were identical (in the case of NWD and NWR) or derived directly from members of the nonword corpus (in the case of CP). The stimuli were systematically generated and unlike those in commonly used NWR tasks. A different pattern of performance emerged to that described in much of the literature, which lends weight to an argument that all NWR tasks are not equal. With this in mind, it may be that NWR per se is not a marker for SLI, but that some NWR tasks accurately identify impairments in a sub-set of skill areas that are highly correlated with SLI.

4.8.7 Conclusions

Data from additional typically developing (TD) participants replicated the age effect observed in Experiment 1 on nonword repetition (NWR), nonword discrimination (NWD) and categorical perception (CP). This finding strengthened the claim that the experimental tasks tapped skills that develop in typically developing (TD) children between the ages of 4;0 and 9;11.

With respect to markers for SLI, sentence recall (specifically the Recalling Sentences subtest of the CELF) was confirmed as the most reliable measure both in a regression analysis and in terms of its sensitivity and specificity. This replicated existing literature identifying sentence recall as a potentially powerful diagnostic marker for SLI (Conti-Ramsden, et al., 2001). Nonword repetition (NWR) was moderately sensitive and specific in identifying children with low language scores, but knowing that a child was diagnosed with a speech or language impairment better predicted his/her language skills than NWR accuracy. Neither NWD nor CP reliably differentiated children with a language impairment from typically developing peers, so could not be proposed as alternative, non-verbal markers for SLI. NWD and CP did, however, predict phonological skills in children with SLI and SLI+SSD and NWD additionally predicted reading in the clinical groups. It should be noted that the overwhelming impression from the regression analyses was that once we know a child is diagnosed with SLI or SLI+SSD, the diagnosis itself is the major predictor of performance on many perceptual and linguistic tasks.

The data suggest a transfer of information from sub-lexical output speech processing stages to sub-lexical input processing stages in typically developing children and children with speech and language impairments, with NWR accuracy predicting performance on the NWD and CP tasks. This transfer of information was not quantified and merits further investigation.

Finally, the experimental tasks identified subtle underlying differences in speech processing between the SLI and SLI+SSD groups, with the latter group performing worse on a test of CP. It is difficult to state with any certainty the precise nature of the differences, but degraded phonological output representations may play a role in the SLI+SSD deficit.

5. General discussion

5.1 Summary of findings

The two experiments presented here were carried out to investigate speech processing in typically developing (TD) children and children with speech and language impairments (SLI or SLI+SSD). The main findings were as follows:

- Performance on the nonword discrimination (NWD), nonword repetition (NWR) and categorical perception (CP) tasks improved with age in TD children between the ages of 4;0 and 9;11.
- The data did not replicate commonly reported relationships between NWR and receptive vocabulary or NWR and phonological short-term memory (PSTM). This indicated that the NWR task used here measured different underlying skills to those measured by published NWR tests.
- Speech production seemed to influence performance on a test of nonword discrimination (NWD). Specifically, nonword repetition accuracy predicted performance on measures of nonword discrimination (NWD) and categorical perception (CP) regardless of group membership.
- Sentence recall was identified as the most reliable predictor of SLI and neither non-verbal experimental measure (NWD, CP) was a viable alternative as a clinical marker.
- Children with SLI+SSD seemed to show an underlying qualitative difference in categorical perception (CP) of nonword stimuli containing a singleton/cluster contrast when compared to TD children and children with a single deficit language impairment (SLI). It is possible that degraded sub-lexical output phonological representations affect children with SLI+SSD.
- The structure of nonword stimuli is crucial to the pattern of typical and atypical performance on NWR and significantly affects the underlying skills that are tapped by the task. Specifically, nonwords of three syllables containing clusters present the most challenge to language impaired children in NWR and NWD tasks, and one syllable nonwords with a singleton-cluster contrast challenge children with SLI+SSD in a CP task.

5.2 Recruitment and sampling issues

The sampling bias in the typically developing (TD) group was discussed in Chapter 4. The group performed above average on some tasks, and it was suggested that not all schools selected participants randomly as they had been asked to do. Anecdotal evidence suggests this experience is common to much research in this field. A sampling bias in favour of high performing TD participants may have affected the outcome of the comparative analyses reported here.

Sampling issues were not only relevant to the TD group. Also in common with many research projects within this field, participants for the SLI and SLI+SSD groups were recruited from a clinically served population, namely specialist language provisions. This recruitment strategy raised a number of practical and experimental issues.

Firstly, it was difficult to recruit sufficient numbers to the clinical groups. Many language provisions were already participating in research projects, and were – understandably – unable to commit to another. This restricted the pool of candidates for the experiment and data from only eight participants with SLI+SSD and six with single deficit SLI were included. Despite the apparent lack of statistical power, it is not uncommon for researchers to publish experiments with clinical groups of a similar size. For example, Gathercole and Baddeley's (1990) seminal study of phonological memory in SLI presented data from only six language impaired children. Moreover, recruiting from a population targeted by most clinical researchers increases the chance that participants have been tested in multiple projects and on potentially similar tasks. The effects of such an experimental confound may be minimal, but should not be ignored.

Selecting participants from within a clinically served population may introduce further confounds. Clinical populations are formed by help-seeking and do not result from the same processes that generate individual differences in the general population. There are natural biasing factors in a clinical population, such as severity of condition and access to services (Tomblin, 2010). It is likely that the most severe cases of SLI will be found in specific language provisions, from which the current cohort was recruited. The picture of SLI that emerges from research may not be representative of the range of deficits in the population at large. This does not mean that research drawn from clinical populations is not essential to drive diagnosis and treatment but it is important that researchers and clinicians recognise its limitations.

Within a clinical population, further selection criteria are often applied to participants. In general, these are necessary to eliminate significant confounds such as excluding children who fail a hearing screen from projects involving auditory tasks. However some criteria may result in data that is not representative of the population being tested.

In this project, for example, seven clinical participants were excluded after the first testing session because they did not achieve a standard score within one standard deviation below the mean on the WISC/WPPSI subtest. While it was important to ensure task performance reflected the participants' linguistic skills rather than other cognitive processes, there is increasing evidence that the distinction between language and conceptual knowledge is not clear-cut. Specific cognitive impairments have been found in children with SLI, for example in visuo-spatial short-term memory (Hick, et al., 2005) and the development of conceptual knowledge (McGregor, et al., 2002). Furthermore, children with SLI experience a significant drop in non-verbal IQ over time (Botting, 2005). Such findings have prompted some researchers to suggest that a 'Residual Normality' model, whereby cognitive skills are spared and there is no developmental interaction between systems over time, is not adequate to describe SLI (Botting, 2005; Karmiloff-Smith, 1998; Thomas & Karmiloff-Smith, 2002). Indeed, there may be variation in the degree to which linguistic and non-linguistic deficits co-occur in SLI and excluding children with obvious non-linguistic impairments from research may be biasing the picture of SLI that is emerging. Non-linguistic deficits have been excluded in some individuals with SLI (van der Lely, 1993), but it may be that not all deficits were probed for, or that other deficits were subtle and difficult to detect (Ullman & Pierpont, 2005). Children have differing patterns of strength and weakness and it is entirely possible that individuals with functionally different impairments may perform similarly on the sub-set of skills measured in a particular research project.

Co-morbidity of communication disorders may also be over-estimated in clinical populations. Chance alone dictates that some children will present with two (related or unrelated) disorders, if both disorders exist. Clinically identified samples are more likely to contain children with co-morbid conditions (Tomblin, 2010), and may result in the identification of false or exaggerated risk factors. For example, a population based study of pre-literacy skills in children with SSD failed to replicate the heightened risk of literacy difficulties reported in children that had been clinically identified (Tomblin, 2010). This suggests that clinical sampling may have biased the outcome of the study that pre-selected participants with SSD.

The results of Experiment 2 presented above showed that group membership accounted for much of the variance in performance on linguistic tasks. In essence, once you know a child has SLI, the diagnostic label itself (assuming the diagnosis is correct and current), predicts performance on a number of linguistic tasks. Test scores provide no information about the nature of the child’s deficit above what can be assumed by their diagnosis. If research participants are selected on the basis of their diagnosis, as in much of the developmental literature as well as the experiment reported here, data should be thoroughly scrutinised before being extrapolated to the population in general.

Finally, clinical identification itself can be problematic. Some studies rely solely on third party information to identify clinical participants, yet a study by Aram et al. (1993) found varying degrees of congruence (20% to 71.4%) between clinically identified children with SLI and those identified as SLI using psychometric criteria. A discrepancy such as this may be related to levels of experience and training on the part of the clinician and/or the researcher, or it could be related to the assessment materials used by either party. Researchers and clinicians need to be aware, for example, that nonword repetition accuracy seems to depend on the nature of the stimuli presented (see clinical implications below). Similarly, conditions such as SLI and SSD are not static, rather they change over time with maturation, intervention and environment (Conti-Ramsden & Botting, 1999); a clinical diagnosis may no longer be entirely appropriate when a child participates in a research project. In the experiment presented above, only a sub-set of the clinical participants were identified as language impaired on the sensitivity/specificity analysis. Indeed, one high achieving participant had to be removed from certain analyses because he skewed the SLI+SSD group data. It is entirely possible that some research participants were nearing the end of their specialist provision. This point raises a further potential confound: that of group assignment based on diagnosis rather than psychometric performance.

There is no simple solution to the issues presented here. Epidemiological research eliminates sampling bias and such confounds as artificially high rates of co-morbidity. However this type of research is costly, time-consuming and often impractical in the developmental field. It would be difficult, for example, in a small-scale epidemiological study to find sufficiently large groups of children with similar language and speech skills for the purpose of group comparison. It falls to the researcher to give a full account of the selection criteria and assessments used to characterise clinically selected participants, and it falls to the reviewer to thoroughly scrutinise the context of the experimental findings.

5.3 Experimental design

When designing an experiment involving a clinical population, extreme care should be taken to devise tasks and select published assessments that are appropriate for the age and ability of the participants to be tested. This basic rule is sometimes not easy to implement for several main reasons. Firstly, as was the case here, the age range of standardised assessments will often partially overlap, but not fully encompass, the ages of all participants. If the same task is to be administered to all participants and standardised scores extrapolated to the oldest (or youngest) individuals, a test must be selected that does not result in consistent floor or ceiling performance. Performance at particularly the lower extreme may not be a result of a task that overloads the skills targeted by the assessment. It may simply be that, for example, the method of response is too complex for younger individuals (e.g. typing on a keyboard) or that the task is insufficiently interesting to maintain attention. For children with SLI, response time data may be similarly confounded if a complex response is required, as language impaired children may have slower processing skills (Leonard, et al., 2007). Less engaging or long tasks may elicit poorer performance, especially in younger participants. Similarly, the size of the assessment battery and the length of each testing session should be considered carefully.

The assessment battery used in Experiment 2 (reported in Chapter 4), while carefully considered for children over the age of five years, was not wholly appropriate for the youngest children that were eventually tested. The age limit was lowered from five to four years during the course of testing and measures such as Sentence Recall were too complex for one child at the bottom end of the age range. The battery itself was also rather too long for the youngest child. Extrapolating age-normed scores was otherwise adequate for the published and informal assessments that were analysed.

5.4 Implications for clinical diagnosis and intervention

This research raises a number of issues relevant to clinical practice. First and foremost, it reinforces research identifying task-dependent differences in NWR accuracy. Archibald and Gathercole (2006a) compared the performance of SLI children on a NWR test commonly used in the UK, the Children's Test of Nonword Repetition (CNRep; Gathercole & Baddeley, 1996a), with their performance on a test commonly used in the US, the Nonword Repetition Test (NRT; Dollaghan & Campbell, 1998). Archibald and Gathercole found that children with SLI were more impaired on the CNRep task and they suggested this was due to the articulatory complexity of the nonword stimuli (the CNRep contains nonwords with consonant clusters; the NRT does not). The performance of children with SLI was likely mitigated by the finding that they benefitted from the lexical similarity of the CNRep stimuli to the same extent as typically developing (TD) controls. Furthermore, NRT scores for all children were closely related to performance on a measure of nonverbal reasoning. Regardless of the conclusions that the authors drew about the language deficits identified by each task, it is clear that the two assessments measure different underlying skills. On the one hand, the longer, more complex stimuli of the CNRep will likely present a greater challenge to children with SLI than the NRT stimuli. On the other hand, language impaired children can use top-down lexical processing to support NWR in the CNRep (which presents stimuli containing lexical items and morphemes) but not the NRT. Correlations with other measures point towards the underlying skills tapped by a task. In the case of the NRT, but not the CNRep, verbal reasoning affects performance. Finally, the commonly reported relationships between digit recall and receptive vocabulary were not replicated in the current study. This suggests that the NWR stimuli used here did not measure phonological short term memory (PSTM) or tap lexical knowledge, both of which have been implicated in previous NWR tasks. The skills measured by NWR and their relationships with other performance tests clearly depend on the nature of the NWR task and the nonword stimuli presented.

Clinicians should be cautious in interpreting assessment data. Assessments published by different authors do not necessarily target the same underlying skills. Informally devised screens may vary in a similar way. This is especially significant when one considers the resources within and across services. Given that a child is likely to access many parts of a service – or many services – over the course of his/her treatment, clinicians need to be very aware of the differences between overtly similar assessments. It is not sufficient to quantify a deficit in terms of age and performance norms; a deficit must be characterised and qualified. An alternative, but less practical solution would be to introduce a national framework prescribing a core set of specific assessment materials.

With respect to task administration, Wells (1995) noted that stimuli for NWR tests should be published in phonological notation with clear instructions about whether the subject is to imitate the tester exactly or to repeat the stimulus via their own pronunciation patterns. This is especially important if responses are to be compared to population norms. Without explicit guidelines for administration, or without adherence to explicit guidelines, the validity of standardisation can be questioned. This applies to any standardised performance assessment.

There is an obvious benefit in developing computer-based materials and scoring procedures, but high-tech solutions are not always practicable or financially viable in a clinical context.

This research also raised the issue of treatment planning. If production skills influence speech perception, the ability to discriminate certain speech sounds should not be a prerequisite for targeting the production of those speech sounds in either a lexical or a morphological context. Rather, this research lends support to therapeutic models that target perception and production in tandem (Rvachew, 1994).

Finally, with respect to the assessment and treatment of children with co-morbid SLI and SSD, this research has highlighted the need to consider that performance on discrimination and repetition tasks may be adversely affected by subtle differences in speech perception. A deficit on NWR or nonword discrimination (NWD) may not simply be a result of language or speech output difficulties, rather children with SLI+SSD may present with complex phonological profiles.

5.5 Future directions

The work presented here could be extended in a number of ways, to answer the following research questions.

5.5.1 How does each deficit affect speech perception in SLI+SSD?

Experiment 2 was originally designed to include a third clinical group consisting of children with only a speech sound disorder (SSD). Unfortunately, participants with this single deficit could not be recruited. It would be interesting to compare input and output speech processing in SSD children with no language impairment, with a view to further evaluating the transfer of information from output to input processing stages.

It would also be useful to evaluate the precise nature of the speech impairment in the participants with SLI+SSD (and any additional recruits with single deficit SSD). Leitao et al. (1997) reported that speech impaired children had weaker phonological awareness than typically developing (TD) controls, but that there were two distinct levels of performance in the speech impaired group. Children with a ‘deviant consistent’ speech impairment performed worse than those with a delayed or inconsistent speech disorder. In the experiment reported here, the speech deficits of the SLI+SSD participants were not analysed in detail and the group may not have been homogeneous in terms of the nature and/or severity of the speech impairments represented. Speech error data could be applied to nonword tasks (NWR and the reading measures) in order to tease apart underlying linguistic errors from surface articulation errors. It could also be compared with the experimental tasks to allow further experimental questions to be answered e.g. do nonword repetition errors represent underlying deficits in generating or storing phonological representations? Is there a relationship between the nature of a speech deficit and processing tasks such as nonword discrimination (NWD) or categorical perception (CP)? A cross-group comparison of performance could offer further insight into the nature of the language impairment in SLI+SSD, or identify any underlying linguistic deficits in ‘pure’ SSD.

5.5.2 What is the precise nature of phonological representation in SLI and SLI+SSD?

Further work could address the quality and accessibility of participants’ lexical phonological representations through tasks such as the Quality of Phonological Representations (QPR) task (Claessen, Heath, Fletcher, Hogben, & Leitao, 2009). This task presents an object picture with a recorded label and requires the child to decide if the object name has been accurately spoken or not. Foils are matched to an object name by varying a consonant or vowel, or by presenting a common misarticulation (such as /bəzgeti/ for ‘spaghetti’). In this way, the assessment targets phonological representations without requiring a verbal response and necessarily accessing output processes. Similarly, it would be interesting to collect clinical data from a rapid serial naming (RSN) task such as that from the Phonological Awareness Battery (PhAB; Fredrickson, et al., 1997). RSN has been shown to predict reading outcomes, and may be a protective factor in SSD (Raitano, et al., 2004). It is suggested that slow performance on RSN may result from imprecise phonological representations affecting accessibility or from a general slowing of access and processing speed (Wolf & Grieg Bowers, 1999). Additional information about participants' lexical representations and lexical access skills would likely allow further interpretation of the experimental data presented in both experiments here.

5.5.3 Are the data replicated in larger clinical groups?

With respect to recruitment, the validity of this research could be strengthened by analysing more clinical data. To improve statistical power, additional participants with SLI and SLI+SSD should be recruited so that the clinical groups are as close in size as possible to the typically developing (TD) group (n = 51). Alternatively or additionally, the data could be analysed with respect to individual differences within the clinical groups. There was clearly considerable variability in performance within the SLI and SLI+SSD groups. In a larger cohort, sub-groups of language impairment and/or speech disorder may be identified.

5.5.4 Precisely how does the nature of the nonword stimuli affect speech processing?

The stimuli used here elicit data that is somewhat different to that reported by other researchers. The experimental stimuli therefore warrant closer evaluation. Firstly, an analysis of wordlikeness and phonotactic probability could be carried out to further explain the lack of relationship between NWR and receptive vocabulary. Secondly, a direct comparison between the discrimination tasks used here and conventional measures of NWR such as the CNRep (Gathercole & Baddeley, 1996a) has merit. The affects of wordlikeness, length and complexity are well-documented with respect to the CNRep and a comparative analysis may shed light on the processes tapped by the NWR task described here. Thirdly, it would be interesting to compare the current data with that from a comparison of NWR, NWD and CP of identical stimuli derived from published measures of NWR. Do identical stimuli yield data that replicates documented relationships between speech perception and published NWR tasks? If not, do the data replicate those reported here?

Finally, in order to verify that no word learning took place over the course of testing, it would be useful to carry out a word recognition task (of the nonword stimuli) at the end of the experiment. If lexicalisation of these stimuli does take place, the challenge would be to design and administer a task that prevented or negated its effect. Alternatively, the possibility of lexicalisation of the stimuli could be eliminated by devising different sets of very carefully matched nonwords for each experimental task. The matching would have to take into account such factors as length, cluster position and stress and the frequency of specific phonemes and phoneme strings.

5.5.5 Concluding remarks

These experiments found subtle differences in the underlying deficits in SLI and SLI+SSD that could not readily be explained by the overt speech errors of the latter group. It seems that the SLI deficit was related to the length and complexity of the nonword stimuli and that a bidirectional transfer of information between sub-lexical output and input phonological representations may explain the more significant deficits observed in SLI+SSD. Of the tasks evaluated, sentence recall was the most powerful marker of SLI and no non-verbal experimental task was a reliable alternative.

When the data was considered in the context of existing developmental literature, it was clear that all nonword repetition (NWR) tasks are not equal. The nature of the nonword stimuli is crucial to NWR accuracy and should be borne in mind when designing, administering or interpreting NWR tasks. Nonwords do have the power to shed light on SLI and related disorders but, like all powerful tools, must be handled with care.

6. References

Adams, A. M., & Gathercole, S. E. (1995). Phonological Working-Memory and Speech Production in Preschool-Children. *Journal of Speech and Hearing Research, 38*(2), 403-414.

Adams, A. M., & Gathercole, S. E. (1996). Phonological Working Memory and Spoken Language Development in Young Children. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 49*(1), 216-233.

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and Visuospatial Short-Term and Working Memory in Children: Are They Separable? *Child Development, 77*(6), 1698-1716.

Aram, D. M., Morris, R., & Hall, N. E. (1993). Clinical and Research Congruence in Identifying Children with Specific Language Impairment. *Journal of Speech and Hearing Research, 36*(3), 580-591.

Archibald, L. M. D., & Gathercole, S. E. (2006a). Nonword Repetition: A Comparison of Tests. *Journal of Speech Language and Hearing Research, 49*(5), 970-983.

Archibald, L. M. D., & Gathercole, S. E. (2006b). Short-Term and Working Memory in Specific Language Impairment. *International Journal of Language & Communication Disorders, 41*(6), 675-693.

Archibald, L. M. D., & Joanisse, M. F. (2009). On the Sensitivity and Specificity of Nonword Repetition and Sentence Recall to Language and Memory Impairments in Children. *Journal of Speech Language and Hearing Research, 52*(4), 899-914.

Archibald, L. M. D., Joanisse, M. F., & Shepherd, M. (2008). Associations between Key Language-Related Measures in Typically Developing School-Age Children. *Zeitschrift Fur Psychologie-Journal of Psychology, 216*(3), 161-171.

Baddeley, A. (2000). The Episodic Buffer: A New Component of Working Memory? *Trends in Cognitive Sciences, 4*(11), 417-423.

Baddeley, A., Gathercole, S., & Papagno, C. (1998). The Phonological Loop as a Language Learning Device. [Review]. *Psychological Review, 105*(1), 158-173.

Baddeley, A., & Hitch, G. J. (1974). Working Memory. In G. A. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

Bird, J., & Bishop, D. V. M. (1992). Perception and Awareness of Phonemes in Phonologically Impaired Children. *European Journal of Disorders of Communication, 27*(4), 289-311.

Bird, J., Bishop, D. V. M., & Freeman, N. H. (1995). Phonological Awareness and Literacy Development in Children with Expressive Phonological Impairments. *Journal of Speech and Hearing Research, 38*(2), 446-462.

Bishop, D. V. M. (1997). *Uncommon Understanding: Development and Disorders of Language Comprehension in Children*. Hove: Psychology Press.

Bishop, D. V. M. (2002). Motor Immaturity and Specific Speech and Language Impairment: Evidence for a Common Genetic Basis. *American Journal of Medical Genetics (Neuropsychiatric Genetics), 114*(1), 56-63.

Bishop, D. V. M. (2003). Test for Reception of Grammar (Version 2 ed.). London: Psychological Corporation.

Bishop, D. V. M., Adams, C. V., Nation, K., & Rosen, S. (2005). Perception of Transient Nonspeech Stimuli Is Normal in Specific Language Impairment: Evidence from Glide Discrimination. [Article]. *Applied Psycholinguistics, 26*(2), 175-194.

Bishop, D. V. M., Adams, C. V., & 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*(2), 158-169.

Bishop, D. V. M., Bishop, S. J., Bright, P., James, C., Delaney, T., & Tallal, P. (1999). Different Origin of Auditory and Phonological Processing Problems in Children with Language Impairment: Evidence from a Twin Study. *Journal of Speech Language and Hearing Research, 42*(1), 155-168.

Bishop, D. V. M., Brown, B. B., & Robson, J. (1990). The Relationship between Phoneme Discrimination, Speech Production, and Language Comprehension in Cerebral-Palsied Individuals. *Journal-of-Speech-and-Hearing-Research., 33*(2), 210-219.

Bishop, D. V. M., & Edmundson, A. (1987). Language-Impaired 4-Year-Olds - Distinguishing Transient from Persistent Impairment. *Journal of Speech and Hearing Disorders, 52*(2), 156-173.

Bishop, D. V. M., Mcdonald, D., Bird, S., & Hayiou-Thomas, M. E. (2009). Children Who Read Words Accurately Despite Language Impairment: Who Are They and How Do They Do It? *Child Development, 80*(2), 593-605.

Bishop, D. V. M., North, T., & Donlan, C. (1996). Nonword Repetition as a Behavioural Marker for Inherited Language Impairment: Evidence from a Twin Study. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 37*(4), 391-403.

Bishop, D. V. M., & Snowling, M. J. (2004). Developmental Dyslexia and Specific Language Impairment: Same or Different? *Psychological Bulletin, 130*(6), 858-886.

Botting, N. (2005). Non-Verbal Cognitive Development and Language Impairment. *Journal of Child Psychology and Psychiatry, 46*(3), 317-326.

Botting, N., & Conti-Ramsden, G. (2003). Autism, Primary Pragmatic Difficulties, and Specific Language Impairment: Can We Distinguish Them Using Psycholinguistic Markers? *Developmental Medicine and Child Neurology, 45*(8), 515-524.

Bowey, J. A. (1996). On the Association between Phonological Memory and Receptive Vocabulary in Five-Year-Olds. *Journal of Experimental Child Psychology, 63*(1), 44-78.

Bowey, J. A. (2001). Nonword Repetition and Young Children's Receptive Vocabulary: A Longitudinal Study. *Applied Psycholinguistics, 22*(3), 441-469.

Brackenbury, T., & Pye, C. (2005). Semantic Deficits in Children with Language Impairments: Issues for Clinical Assessment. *Language Speech and Hearing Services in Schools, 36*(1), 5-16.

Briscoe, J., Bishop, D. V. M., & Norbury, C. F. (2001). Phonological Processing, Language, and Literacy: A Comparison of Children with Mild-to-Moderate Sensorineural Hearing Loss and Those with Specific Language Impairment. *Journal-of-Child-Psychology-and-Psychiatry., 42*(3), 329-340.

Briscoe, J., & Rankin, P. M. (2009). Exploration of a 'Double-Jeopardy' Hypothesis within Working Memory Profiles for Children with Specific Language Impairment. *International Journal of Language & Communication Disorders, 44*(2), 236-250.

Broomfield, J., & Dodd, B. (2004). The Nature of Referred Subtypes of Primary Speech Disability. *Child Language Teaching and Therapy, 20*(2), 135-151.

Bryan, K. (2004). Preliminary Study of the Prevalence of Speech and Language Difficulties in Young Offenders. *International Journal of Language and Communication Disorders, 39*(3), 391-400.

Catts, H. W., Adlof, S. M., Hogan, T. P., & Weismer, S. E. (2005). Are Specific Language Impairment and Dyslexia Distinct Disorders? *Journal of Speech Language and Hearing Research, 48*(6), 1378-1396.

Ceponiene, R., Service, E., Kurjenluoma, S., Cheour, M., & Naatanen, R. (1999). Children's Performance on Pseudoword Repetition Depends on Auditory Trace Quality: Evidence from Event-Related Potentials. *Developmental Psychology, 35*(3), 709-720.

Clegg, J., Hollis, C., Mawhood, L., & Rutter, M. (2005). Developmental Language Disorders a Follow-up in Later Adult Life. Cognitive, Language and Psychosocial Outcomes. *Journal of Child Psychology and Psychiatry, 46*(2), 128-149.

Clegg, J., Hollis, C., & Rutter, M. (1999). Life Sentence. *Bulletin, November*.

Coady, J. A., & Evans, J. L. (2008). Uses and Interpretations of Non-Word Repetition Tasks in Children with and without Specific Language Impairments (Sli). *International Journal of Language & Communication Disorders, 43*(1), 1-40.

Coady, J. A., Kluender, K. R., & Evans, J. L. (2005). Categorical Perception of Speech by Children with Specific Language Impairments. *Journal-of-Speech,-Language,-and-Hearing-Research, 48*(4), 944-959.

Conti-Ramsden, G., & Botting, N. (1999). Classification of Children with Specific Language Impairment: Longitudinal Considerations. *Journal of Speech Language and Hearing Research, 42*(5), 1195-1204.

Conti-Ramsden, G., & Botting, N. (2004). Social Difficulties and Victimization in Children with Sli at 11 Years of Age. *Journal of Speech Language and Hearing Research, 47*(1), 145-161.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic Markers for Specific Language Impairment (Sli). *Journal of Child Psychology and Psychiatry, 42*(6), 741-748.

Conti-Ramsden, G., Crutchley, A., & Botting, N. (1997). The Extent to Which Psychometric Tests Differentiate Subgroups of Children with Sli. *J Speech Lang Hear Res, 40*(4), 765-777.

Conti-Ramsden, G., & Durkin, K. (2007). Phonological Short-Term Memory, Language and Literacy: Developmental Relationships in Early Adolescence in Young People with Sli. *Journal of Child Psychology and Psychiatry, 48*(2), 147-156.

Conti-Ramsden, G., Durkin, K., Simkin, Z., & Knox, E. (2009). Specific Language Impairment and School Outcomes. I: Identifying and Explaining Variability at the End of Compulsory Education. *International Journal of Language & Communication Disorders, 44*(1), 15-35.

Cook, R., & Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. Chichester: John Wiley and Sons Inc.

Cowan, R., Donlan, C., Newton, E. J., & Lloyd, D. (2005). Number Skills and Knowledge in Children with Specific Language Impairment. *Journal of Educational Psychology, 97*(4), 732-744.

Diagnostic and Statistical Manual of Mental Disorders. (1994). (4th ed.). Washington, DC: American Psychiatric Association.

Dockrell, J. E., Messer, D., George, R., & Wilson, G. (1998). Children with Word-Finding Difficulties - Prevalence, Presentation and Naming Problems. *International Journal of Language & Communication Disorders, 33*(4), 445-454.

Dodd, B. (1995). *Differential Diagnosis and Treatment of Children with Speech Disorder*. London: Whurr.

Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2002). Diagnostic Evaluation of Articulation and Phonology. Hove: Psychological Corporation.

Dodd, B., & Iacano, T. (1989). Phonological Disorders in Children - Changes in Phonological Process Use During Treatment. *British Journal of Disorders of Communication, 24*(3), 333-352.

Dollaghan, C., Biber, M., & Campbell, T. (1993). Constituent Syllable Effects in a Nonsense-Word Repetition Task. *Journal of Speech and Hearing Research, 36*(5), 1051-1054.

Dollaghan, C., & Campbell, T. F. (1998). Nonword Repetition and Child Language Impairment. *J Speech Lang Hear Res, 41*(5), 1136-1146.

Dollaghan, C. A. (1987). Fast Mapping in Normal and Language-Impaired Children. *Journal of Speech and Hearing Disorders, 52*(3), 218-222.

Dollaghan, C. A., Biber, M. E., & Campbell, T. F. (1995). Lexical Influences on Nonword Repetition. *Applied Psycholinguistics, 16*(2), 211-222.

Dunn, L., Dunn, L., Whetton, C., & Burley, J. (1997). British Picture Vocabulary Scale (Second ed.). Windsor: NFER-Nelson.

Edwards, J., Beckman, M. E., & 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*(2), 421-436.

Edwards, J., Fox, R. A., & Rogers, C. L. (2002). Final Consonant Discrimination in Children: Effects of Phonological Disorder, Vocabulary Size and Articulatory Accuracy. *Journal of Speech Language and Hearing Research, 45*(2), 231-242.

Edwards, J., & Lahey, M. (1998). Nonword Repetitions of Children with Specific Language Impairment: Exploration of Some Explanations for Their Inaccuracies. *Applied Psycholinguistics, 19*(2), 279-309.

Enderby, P., Pickstone, C., John, A., Fryer, K., Cantrell, A., & Papaioannou, D. (2009). R.C.S.L.T. Resource Manual for Commissioning and Planning Services for Slcn. Retrieved 2nd May, 2010, from http://www.rcslt.org/speech_and_language_therapy/intro/sli

Fazio, B. B. (1999). Arithmetic Calculation, Short-Term Memory, and Language Performance in Children with Specific Language Impairment: A 5-Year Follow-Up. *Journal of Speech Language and Hearing Research, 42*(2), 420-431.

Fisher, S. E. (2005). Dissection of Molecular Mechanisms Underlying Speech and Language Disorders. *Applied Psycholinguistics, 26*(1), 111-128.

Fox, A. V., Dodd, B., & Howard, D. (2002). Risk Factors for Speech Disorders in Children. *International Journal of Language & Communication Disorders, 37*(2), 117-131.

Fredrickson, N., Frith, U., & Reason, R. (1997). Phonological Assessment Battery. Windsor: NFER-NELSON.

Frumkin, B., & Rapin, I. (1980). Perception of Vowels and Consonant-Vowels of Varying Duration in Language Impaired Children. [Article]. *Neuropsychologia, 18*(4-5), 443-454.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The Motor Theory of Speech Perception Reviewed. *Psychonomic Bulletin & Review, 13*(3), 361-377.

Gallon, N., Harris, J., & Van Der Lely, H. (2007). Non-Word Repetition: An Investigation of Phonological Complexity in Children with Grammatical Sli. [Article]. *Clinical Linguistics & Phonetics, 21*(6), 435-455.

Gathercole, S. E. (1995). Is Nonword Repetition a Test of Phonological Memory or Long-Term Knowledge - It All Depends on the Nonwords. *Memory & Cognition, 23*(1), 83-94.

Gathercole, S. E. (2006). Nonword Repetition and Word Learning: The Nature of the Relationship. *Applied Psycholinguistics, 27*(04), 513-543.

Gathercole, S. E., & Adams, A. M. (1994). Childrens Phonological Working-Memory - Contributions of Long-Term Knowledge and Rehearsal. *Journal of Memory and Language, 33*(5), 672-688.

Gathercole, S. E., & Baddeley, A. (1996a). *The Children's Test of Nonword Repetition*. London: The Psychological Corporation.

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the Role of Phonological Stm in the Development of Vocabulary in Children - a Longitudinal-Study. *Journal of Memory and Language, 28*(2), 200-213.

Gathercole, S. E., & Baddeley, A. D. (1990). Phonological Memory Deficits in Language Disordered Children - Is There a Causal Connection? *Journal of Memory and Language, 29*(3), 336-360.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic Influences on Short-Term Memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 25*(1), 84-95.

Gathercole, S. E., Service, E., Hitch, G. J., Adams, A. M., & Martin, A. J. (1999). Phonological Short-Term Memory and Vocabulary Development: Further Evidence on the Nature of the Relationship. *Applied Cognitive Psychology, 13*(1), 65-77.

Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: A Test of Phonological Working Memory. *Memory, 2*(2), 103 - 127.

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological Memory and Vocabulary Development During the Early School Years - a Longitudinal-Study. *Developmental Psychology, 28*(5), 887-898.

Girolametto, L., Wiigs, M., Smyth, R., Weitzman, E., & Pearce, P. S. (2001). Children with a History of Expressive Vocabulary Delay: Outcomes at 5 Years of Age. *American Journal of Speech-Language Pathology, 10*(4), 358-369.

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. *J Speech Lang Hear Res, 47*(5), 1117-1132.

Groenen, P., Maassen, B., Crul, T., & Thoonen, G. (1996). The Specific Relation between Perception and Production Errors for Place of Articulation Developmental Apraxia of Speech. *Journal of Speech and Hearing Research, 39*(3), 468-482.

Grunwell, P. (1992). Principled Decision Making in the Remediation of Children with Neurological Disorders. In P. Fletcher & D. Hall (Eds.), *Speech and Language Disorders in Children*. London: Whurr.

Haskill, A. M., & Tyler, A. A. (2007). A Comparison of Linguistic Profiles in Subgroups of Children with Specific Language Impairment. *American Journal of Speech-Language Pathology, 16*(3), 209-221.

Hauk, O., Johnsrude, I., & Pulvermuller, F. (2004). Somatotopic Representation of Action Words in Human Motor and Premotor Cortex. *Neuron, 41*(2), 301-307.

Hauner, K. K. Y., Shriberg, L. D., Kwiatkowski, J., & Allen, C. T. (2005). A Subtype of Speech Delay Associated with Developmental Psychosocial Involvement. [Article]. *Journal of Speech Language and Hearing Research, 48*(3), 635-650.

Hazan, V., & Barrett, S. (2000). The Development of Phonemic Categorization in Children Aged 6-12. *Journal of Phonetics, 28*, 377-396.

Hick, R., Botting, N., & Conti-Ramsden, G. (2005). Cognitive Abilities in Children with Specific Language Impairment: Consideration of Visuo-Spatial Skills. *International Journal of Language & Communication Disorders, 40*(2), 137-149.

Hill, E. L. (2001). Non-Specific Nature of Specific Language Impairment: A Review of the Literature with Regard to Concomitant Motor Impairments. *International Journal of Language & Communication Disorders, 36*(2), 149-171.

Hoffman, P. R., Stager, S., & Daniloff, R. G. (1983). Perception and Production of Misarticulated /R. *Journal of Speech and Hearing Disorders, 48*(2), 210-215.

Huckvale, M. (2004). Speech Filing System (Version 1.5 Release 4.6).

Huckvale, M., & Yanagisawa, K. (2007). *Spoken Language Conversion with Accent Morphing.* Paper presented at the Sixth ISCA Workshop on Speech Synthesis (SSW6-2007), Bonn.

Jaeschke, R., Guyatt, G., & Lijmer, J. (2002). Diagnostic Tests. In G. Guyatt & D. Rennie (Eds.), *User's Guide to the Medical Literature* (pp. 121-140). Chicago: American Medical Association Press.

James, D. (2006). *Hippopotamus Is Hard to Say: Children's Acquisition of Polysyllabic Words.* University of Sydney, Australia, Sydney.

James, D. G. H., Van Doorn, J., & Mcleod, S. (2008). The Contribution of Polysyllabic Words in Clinical Decision Making About Children's Speech. *Clinical Linguistics & Phonetics, 22*(4-5), 345-353.

Joanisse, M. F., & Seidenberg, M. S. (1998). Specific Language Impairment: A Deficit in Grammar or Processing? *Trends in Cognitive Sciences, 2*(7), 240-247.

Joanisse, M. F., & Seidenberg, M. S. (2003). Phonology and Syntax in Specific Language Impairment: Evidence from a Connectionist Model. *Brain and Language, 86*(1), 40-56.

Johnston, J. R., & Smith, L. B. (1989). Dimensional Thinking in Language Impaired Children. *Journal of Speech and Hearing Research, 32*(1), 33-38.

Johnston, J. R., & Weismer, S. E. (1983). Mental Rotation Abilities in Language-Disordered Children. *Journal of Speech and Hearing Research, 26*(3), 397-403.

Jusczyk, P. W., & Luce, P. A. (2002). Speech Perception and Spoken Word Recognition: Past and Present *Ear-and-Hearing, 23*(1), 2-40.

Kail, R. (1994). A Method for Studying the Generalized Slowing Hypothesis in Children with Specific Language Impairment. *Journal of Speech and Hearing Research, 37*(2), 418-421.

Karmiloff-Smith, A. (1998). Development Itself Is the Key to Understanding Developmental Disorders. *Trends in Cognitive Sciences, 2*(10), 389-398.

Kirk, C., & Demuth, K. (2005). Asymmetries in the Acquisition of Word-Initial and Word-Final Consonant Clusters. *Journal of Child Language, 32*(4), 709-734.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing Sounds, Understanding Actions: Action Representation in Mirror Neurons. *Science, 297*(5582), 846-848.

Larrivee, L. S., & Catts, H. W. (1999). Early Reading Achievement in Children with Expressive Phonological Disorders. *American Journal of Speech-Language Pathology, 8*(2), 118-128.

Law, J., Boyle, J., Harris, F., Harkness, A., & Nye, C. (2000). Prevalence and Natural History of Primary Speech and Language Delay: Findings from a Systematic Review of the Literature. *International Journal of Language & Communication Disorders, 35*(2), 165-188.

Leitao, S., Hogben, J., & Fletcher, J. (1997). Phonological Processing Skills in Speech and Language Impaired Children. *European Journal of Disorders of Communication, 32*(2), 91-111.

Leonard, L. (1998). *Children with Specific Language Impairment*. Cambridge, MA: MIT Press.

Leonard, L. B. (1982). Phonological Deficits in Children with Developmental Language Impairment. *Brain and Language, 16*(1), 73-86.

Leonard, L. B. (1995). Phonological Impairment. In P. Fletcher & B. MacWhinney (Eds.), *The Handbook of Child Language* (pp. 573-602). Oxford: Blackwell.

Leonard, L. B., Weismer, S. E., Miller, C. A., Francis, D. J., Tomblin, J. B., & Kail, R. V. (2007). Speed of Processing, Working Memory, and Language Impairment in Children. *Journal of Speech Language and Hearing Research, 50*(2), 408-428.

Levitt, H. (1971). Transformed up-Down Methods in Psychoacoustics. *Journal of the Acoustical Society of America, 49*(2), 467-&.

Lewis, B. A., & Freebairn, L. (1992). Residual Effects of Preschool Phonology Disorders in Grade School, Adolescence, and Adulthood. *J Speech Hear Res, 35*(4), 819-831.

Lewis, B. A., Freebairn, L. A., Hansen, A. J., Stein, C. M., Shriberg, L. D., Iyengar, S. K., et al. (2006). Dimensions of Early Speech Sound Disorders: A Factor Analytic Study. *Journal of Communication DIsorders, 39*(2), 139-157.

Lewis, B. A., Freebairn, L. A., & Taylor, H. G. (2000). Follow-up of Children with Early Expressive Phonology Disorders. *Journal of learning disabilities, 33*(5), 433-444.

Lewis, B. A., Shriberg, L. D., Freebairn, L. A., Hansen, A. J., Stein, C. M., Taylor, H. G., et al. (2006). The Genetic Bases of Speech Sound Disorders: Evidence from Spoken and Written Language. *Journal of Speech Language and Hearing Research, 49*(6), 1294-1312.

Liberman, A. M. (1957). Some Results of Research on Speech Perception. *Journal of the Acoustical Society of America, 29*(1), 117-123.

Locke, J. L., & Scott, K. K. (1979). Phonetically Mediated Recall in the Phonetically Disordered Child. *Journal of Communication DIsorders, 12*(2), 125-131.

Lum, J. A. G., Gelgic, C., & Conti-Ramsden, G. (2010). Procedural and Declarative Memory in Children with and without Specific Language Impairment. *International Journal of Language & Communication Disorders, 45*(1), 96-107.

Mann, V. A., & Liberman, I. Y. (1984). Phonological Awareness and Verbal Short-Term-Memory. [Article]. *Journal of learning disabilities, 17*(10), 592-599.

Marshall, C. R., & 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.

Marton, K., & Schwartz, R. G. (2003). Working Memory Capacity and Language Processes in Children with Specific Language Impairment. *Journal of Speech Language and Hearing Research, 46*(5), 1138-1153.

Masterson, J., Laxon, V., Carnegie, E., Wright, S., & Horslen, J. (2005). Nonword Recall and Phonemic Discrimination in Four- to Six-Year-Old Children. *Journal of Research in Reading, 28*(2), 183-201.

Mcarthur, G. M., & Bishop, D. V. M. (2001). Auditory Perceptual Processing in People with Reading and Oral Langauge Impairments: Current Issues and Recommendations. *Dyslexia, 7*, 150-170.

Mcarthur, G. M., & Bishop, D. V. M. (2005). Speech and Non-Speech Processing in People with Specific Language Impairment: A Behavioural and Electrophysiological Study. *Brain and Language, VOL 94; NUMBER 3*, 260-273.

Mcgrath, L. M., Hutaff-Lee, C., Scott, A., Boada, R., Shriberg, L. D., & Pennington, B. F. (2008). Children with Comorbid Speech Sound Disorder and Specific Language Impairment Are at Increased Risk for Attention-Deficit/Hyperactivity Disorder. *Journal of Abnormal Child Psychology, 36*(2), 151-163.

Mcgregor, K. K. (1997). The Nature of Word-Finding Errors of Preschoolers with and without Word-Finding Deficits. *Journal of Speech Language and Hearing Research, 40*(6), 1232-1244.

Mcgregor, K. K., Newman, R. M., Reilly, R. M., & Capone, N. C. (2002). Semantic Representation and Naming in Children with Specific Language Impairment. [Article]. *Journal of Speech Language and Hearing Research, 45*(5), 998-1015.

Mcgurk, H., & Macdonald, J. (1976). Hearing Lips and Seeing Voices. *Nature, 264*(5588), 746-748.

Mcleod, S., Van Doorn, J., & Reed, V. A. (2001). Normal Acquisition of Consonant Clusters. *American Journal of Speech-Language Pathology, 10*(2), 99-110.

Menyuk, P. (1964). Comparison of Grammar of Children with Functionally Deviant and Normal Speech. *Journal of Speech and Hearing Research, 7*(2), 109-121.

Messer, D., & Dockrell, J. E. (2006). Children's Naming and Word-Finding Difficulties: Descriptions and Explanations. *Journal of Speech Language and Hearing Research, 49*(2), 309-324.

Metsala, J. L. (1999). Young Children's Phonological Awareness and Nonword Repetition as a Function of Vocabulary Development. *Journal of Educational Psychology, 91*(1), 3-19.

Metsala, J. L., & Chisholm, G. M. (2010). The Influence of Lexical Status and Neighborhood Density on Children's Nonword Repetition. *Applied Psycholinguistics, 31*(3), 489-506.

Michas, I. C., & Henry, L. A. (1994). The Link between Phonological Memory and Vocabulary Acquisition. *British Journal of Developmental Psychology, 12*, 147-163.

Miller, C. A., Kail, R., Leonard, L. B., & Tomblin, J. B. (2001). Speed of Processing in Children with Specific Language Impairment. *Journal of Speech Language and Hearing Research, 44*(2), 416-433.

Montgomery, J. (1995). Examination of Phonological Working Memory in Specifically Language Impaired Children. *Applied Psycholinguistics, 16*, 355-378.

Montgomery, J. W. (2003). Working Memory and Comprehension in Children with Specific Language Impairment: What We Know So Far. *Journal of Communication DIsorders, 36*(3), 221-231.

Mortimer, J., & Rvachew, S. (2008). Morphosyntax and Phonological Awareness in Children with Speech Sound Disorders. In G. F. Eden & D. L. Flower (Eds.), *Learning, Skill Acquisition, Reading, and Dyslexia* (Vol. 1145, pp. 275-282).

Mortimer, J., & Rvachew, S. (2010). A Longitudinal Investigation of Morpho-Syntax in Children with Speech Sound Disorders. *Journal of Communication Disorders, 43*(1), 61-76.

Munson, B. (2006). Nonword Repetition and Levels of Abstraction in Phonological Knowledge. *Applied Psycholinguistics, 27*(04), 577-581.

Munson, B., Kurtz, B. A., & Windsor, J. (2005). The Influence of Vocabulary Size, Phonotactic Probability, and Wordlikeness on Nonword Repetitions of Children with and without Specific Language Impairment. *Journal of Speech Language and Hearing Research, 48*(5), 1033-1047.

Nathan, L., Stackhouse, J., Goulandris, N., & Snowling, M. J. (2004). The Development of Early Literacy Skills among Children with Speech Difficulties: A Test of The "Critical Age Hypothesis". *Journal of Speech Language and Hearing Research, 47*(2), 377-391.

Nation, K., Clarke, P., Marshall, C. M., & Durand, M. (2004). Hidden Language Impairments in Children: Parallels between Poor Reading Comprehension and Specific Language Impairment? *Journal of Speech Language and Hearing Research, 47*(1), 199-211.

Newton, E. J., Roberts, M. J., & Donlan, C. (2010). Deductive Reasoning in Children with Specific Language Impairment. *British Journal of Developmental Psychology, 28*(1), 71-87.

Nittrouer, S., & Studdert-Kennedy, M. (1987). The Role of Coarticulatory Effects in the Perception of Fricatives by Children and Adults. *J Speech Hear Res, 30*(3), 319-329.

Norrelgen, F., Lacerda, F., & Forssberg, H. (2002). Temporal Resolution of Auditory Perception and Verbal Working Memory in 15 Children with Language Impairment. *Journal-of-Learning-Disabilities., 35*(6), 540-546.

Pascoe, M., Stackhouse, J., & Wells, B. (2005). Phonological Therapy within a Psycholinguistic Framework: Promoting Change in a Child with Persisting Speech Difficulties. *International Journal of Language & Communication Disorders, 40*(2), 189-220.

Pennington, B. F., & Bishop, D. V. M. (2009). Relations among Speech, Language, and Reading Disorders. [Review]. *Annual Review of Psychology, 60*, 283-306.

Peterson, R. L., Pennington, B. F., Shriberg, L. D., & Boada, R. (2009). What Influences Literacy Outcome in Children with Speech Sound Disorder? *Journal of Speech Language and Hearing Research, 52*(5), 1175-1188.

Pickering, S., & Gathercole, S. (2001). The Working Memory Test Battery for Children. Hove: The Psychological Corporation.

Plante, E., & Vance, R. (1994). Selection of Preschool Language Tests: A Data-Based Approach. *Lang Speech Hear Serv Sch, 25*(1), 15-24.

Pulvermuller, F., Huss, M., Kherif, F., Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor Cortex Maps Articulatory Features of Speech Sounds. *Proceedings of the National Academy of Sciences of the United States of America, 103*(20), 7865-7870.

R.C.S.L.T. (2006). *Supporting Children with Speech, Language and Communication Needs within Integrated Children's Services*: Royal College of Speech and Language Therapists.

Raine, A., Hulme, C., Chadderton, H., & Bailey, P. (1991). Verbal Short-Term-Memory Span in Speech-Disordered Children - Implications for Articulatory Coding in Short-Term-Memory. *Child Development, 62*(2), 415-423.

Raitano, N. A., Pennington, B. F., Tunick, R. A., Boada, R., & Shriberg, L. D. (2004). Pre-Literacy Skills of Subgroups of Children with Speech Sound Disorders. *Journal of Child Psychology and Psychiatry, 45*(4), 821-835.

Ramus, F., Rosen, S., Dakin, S. C., Day, B. L., Castellote, J. M., White, S., et al. (2003). Theories of Developmental Dyslexia: Insights from a Multiple Case Study of Dyslexic Adults. *Brain, 126*, 841-865.

Rapin, I., & Allen, D. A. (1987). *Developmental Dysphasia and Autism in Preschool Children: Characteristics and Subtypes.* Paper presented at the The first international symposium on specific speech and language disorders in children, London.

Rechetnikov, R. P., & Maitra, K. (2009). Motor Impairments in Children Associated with Impairments of Speech or Language: A Meta-Analytic Review of Research Literature. *The American Journal of Occupational Therapy, 63*(3), 255-263.

Renfrew, C. (1995). Word Finding Vocabulary Test (Fourth ed.). Bicester: Winslow Press.

Renfrew, C. (1997). Action Picture Test (Fourth ed.). Bicester: Winslow.

Rice, M. L., Buhr, J. C., & Nemeth, M. (1990). Fast Mapping Word-Learning Abilities of Language-Delayed Preschoolers. *J Speech Hear Disord, 55*(1), 33-42.

Rice, M. L., & Wexler, K. (1996). Toward Tense as a Clinical Marker of Specific Language Impairment in English-Speaking Children. *Journal of Speech and Hearing Research, 39*(6), 1239-1257.

Rice, M. L., Wexler, K., & Hershberger, S. (1998). Tense over Time: The Longitudinal Course of Tense Acquisition in Children with Specific Language Impairment. *Journal of Speech Language and Hearing Research, 41*(6), 1412-1431.

Rosen, S. (2003). Auditory Processing in Dyslexia and Specific Language Impairment: Is There a Deficit? What Is Its Nature? Does It Explain Anything? *Journal of Phonetics, 31*(3-4), 509-527.

Rvachew, S. (1994). Speech Perception Training Can Facilitate Sound Production Learning. *Journal of Speech and Hearing Research, 37*, 347-357.

Rvachew, S. (2007). Phonological Processing and Reading in Children with Speech Sound Disorders. *American Journal of Speech-Language Pathology, 16*(3), 260-270.

Rvachew, S., Gaines, B. R., Cloutier, G., & Blanchet, N. (2005). Productive Morphology Skills of Children with Speech Delay. *Journal of Speech -Language Pathology and Audiology, 29*(2), 83-89.

Rvachew, S., & Jamieson, D. (1989). Perception of Voiceless Fricatives by Children with a Functional Articulation Disorder. *Journal of Speech and Hearing Disorders, 54*, 193-208.

Rvachew, S., Ohberg, A., Grawburg, M., & 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.

Sahlen, C. R.-W., Ulrika Nettelbladt, Karl Radeborg, Birgitta. (1999). Non-Word Repetition in Children with Language Impairment-Pitfalls and Possibilities. *International Journal of Language & Communication Disorders, 34*(3), 337-352.

Saxman, J. H., & Miller, J. F. (1973). Short-Term-Memory and Language Skills in Articulation-Deficient Children. *Journal of Speech and Hearing Research, 16*(4), 721-730.

Semel, E., Wiig, E., & Secord, W. (2000). Clinical Evaluation of Language Fundamentals (Third ed.). London: The Psychological Corporation.

Semel, E., Wiig, E., & Secord, W. (2006). Clinical Evaluation of Language Fundamentals (Fourth UK ed.). London: Harcourt Assessment.

Shriberg, L. D. (1993). 4 New Speech and Prosody-Voice Measures for Genetics Research and Other Studies in Developmental Phonological Disorders. *Journal of Speech and Hearing Research, 36*(1), 105-140.

Shriberg, L. D., Austin, D., Lewis, B. A., Mcsweeny, J. L., & Wilson, D. L. (1997). The Speech Disorders Classification System (Sdcs): Extensions and Lifespan Reference Data. *Journal of Speech Language and Hearing Research, 40*(4), 723-740.

Shriberg, L. D., Barbara, L., Tomblin, J., Mcsweeny, J. L., Karlsson, H. B., & Scheer, A. R. (2005). Toward Diagnostic and Phenotype Markers for Genetically Transmitted Speech Delay. *Journal of Speech Language and Hearing Research, 48*(4), 834-852.

Shriberg, L. D., Tomblin, J. B., & Mcsweeny, J. L. (1999). Prevalence of Speech Delay in 6-Year-Old Children and Comorbidity with Language Impairment. *Journal of Speech Language and Hearing Research, 42*(6), 1461-1481.

Smit, A. B., Hand, L., Freilinger, J. J., Bernthal, J. E., & Bird, A. (1990). The Iowa Articulation Norms Project and Its Nebraska Replication. *Journal of Speech and Hearing Disorders, 55*(4), 779-798.

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

Stein, C. M., Millard, C., Kluge, A., Miscimarra, L. E., Cartier, K. C., Freebairn, L. A., et al. (2006). Speech Sound Disorder Influenced by a Locus in 15q14 Region. *Behavior Genetics, 36*(6), 858-868.

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 of Speech Language and Hearing Research, 41*(2), 407-418.

Sussman, J.-E. (1993). Perception of Formant Transition Cues to Place of Articulation in Children with Language Impairments. *Journal-of-Speech-and-Hearing-Research, 36*(6), 1286-1299.

Szenkovits, G., & Ramus, F. (2005). Exploring Dyslexics' Phonological Deficit I: Lexical Vs Sub-Lexical and Input Vs Output Processes. *Dyslexia:-An-International-Journal-of-Research-and-Practice, 11*(4), 253-268.

Tallal, P., & Piercy, M. (1973). Developmental Aphasia - Impaired Rate of Nonverbal Processing as a Function of Sensory Modality. *Neuropsychologia, 11*(4), 389-398.

Thibodeau, L. M., & Sussman, H. M. (1979). Performance on a Test of Categorical Perception of Speech in Normal and Communication Disordered Children. *Journal of Phonetics, 7*(4), 375-391.

Tomblin, J. (2009). Co-Morbidly of Autism and Sli: Kinds, Kin or Confounds? *IJLCD Winter Lecture*, from http://www.city.ac.uk/lcs/ijlcd-winter-lecture.html

Tomblin, J. B. (2010). The Episli Database: A Publicly Available Database on Speech and Language. *Language, Speech, and Hearing Services in Schools, 41*, 108-117.

Tomblin, J. B., Freese, P. R., & Records, N. L. (1992). Diagnosing Specific Language Impairment in Adults for the Purpose of Pedigree Analysis. *Journal of Speech and Hearing Research, 35*(4), 832-843.

Tomblin, J. B., Mainela-Arnold, E., & Zhang, X. (2007). Procedural Learning in Adolescents with and without Specific Language Impairment. *Language Learning and Development, 3*(4), 269 - 293.

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.

Treiman, R., & Zukowski, A. (1996). Children's Sensitivity to Syllables, Onsets, Rimes, and Phonemes. [Article]. *Journal of Experimental Child Psychology, 61*(3), 193-215.

Ullman, M. T., & Pierpont, E. I. (2005). Specific Language Impairment Is Not Specific to Language: The Procedural Deficit Hypothesis. *Cortex, 41*(3), 399-433.

Van Der Lely, H. K. J. (1993). Specific Language Impairment in Children: Research Findings and Their Therapeutic Implications. *International Journal of Language & Communication Disorders, 28*(3), 247-261.

Van Der Lely, H. K. J. (1996). Specifically Language Impaired and Normally Developing Children: Verbal Passive Vs Adjectival Passive Sentence Interpretation. [Article]. *Lingua, 98*(4), 243-272.

Van Der Lely, H. K. J., Rosen, S., & Mcclelland, A. (1998). Evidence for a Grammar-Specific Deficit in Children. *Current Biology, 8*(23), 1253-1258.

Van Der Lely, H. K. J., & Stollwerck, L. (1997). Binding Theory and Grammatical Specific Language Impairment in Children. *Cognition, 62*(3), 245-290.

Vance, M., Rosen, S., & Coleman, M. (2009). Assessing Speech Perception in Young Children and Relationships with Language Skills. *International Journal of Audiology, 48*(10), 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.

Vernes, S. C., Newbury, D. F., Abrahams, B. S., Winchester, L., Nicod, J., Groszer, M., et al. (2008). A Functional Genetic Link between Distinct Developmental Language Disorders. *New England Journal of Medicine, 359*(22), 2337-2345.

Watson, M., & Scukanec, G. (1997). Profiling the Phonological Abilities of 2-Year-Olds: A Longitudinal Investigation. *Child Language Teaching and Therapy*, 3-14.

Wechsler, D. (1990). Wechsler Preschool and Primary Scale of Intelligence - Revised. Sidcup, Kent: The Psychological Corporation.

Wechsler, D. (1992). Wechsler Intelligence Scale for Children (Third ed.). Sidcup, Kent: The Psychological Corporation.

Weismer, S. E., & Hesketh, L. J. (1996). Lexical Learning by Children with Specific Language Impairment: Effects of Linguistic Input Presented at Varying Speaking Rates. *J Speech Hear Res, 39*(1), 177-190.

Wells, B. (1995). Phonological Considerations in Repetition Tests. *Cognitive Neuropsychology, 12*(8), 847-855.

Werker, J.-F., Fennell, C.-T., Corcoran, K.-M., & Stager, C.-L. (2002). Infants' Ability to Learn Phonetically Similar Words: Effects of Age and Vocabulary Size. *Infancy, 3*(1), 1-30.

Werker, J. F., & Tees, R. C. (1984). Cross-Language Speech-Perception - Evidence for Perceptual Reorganization During the 1st Year of Life. *Infant Behavior & Development, 7*(1), 49-63.

Whitehill, T. L., Francis, A. L., & Ching, C. K. Y. (2003). Perception of Place of Articulation by Children with Cleft Palate and Posterior Placement. *Journal of Speech Language and Hearing Research, 46*(2), 451-461.

Wolf, M., & Grieg Bowers, P. (1999). The Double-Deficit Hypothesis for the Developmental Dyslexias. *Journal of Educational Psychology, 91*(3), 415-438.

Wolf, M., & Segal, D. (1992). Word Finding and Reading in the Developmental Dyslexias. *Topics in Language Disorders, 13*(1), 51-65.