NONWORD REPETITION IN CHILDREN WITH DEVELOPMENTAL LANGUAGE DISORDER: REVISITING THE CASE OF CANTONESE

FU NGA CHING

MPHIL

The Hong Kong Polytechnic University

2024

The Hong Kong Polytechnic University

Department of Chinese and Bilingual Studies

NONWORD REPETITION IN CHILDREN WITH DEVELOPMENTAL LANGUAGE DISORDER: REVISITING THE CASE OF CANTONESE

FU NGA CHING

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

Jan 2023

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

Fu Nga Ching

DEDICATION

To my father, in loving memory. Thank you for encouraging me to go on this journey.

ABSTRACT

Background: Nonword repetition (NWR) has been advocated as a cross-linguistic clinical marker of Developmental Language Disorder (DLD; Bishop et al., 1996; Conti-Ramsden et al., 2001; Dollaghan & Campbell, 1998), due to its ability to discriminate children with DLD and those with typical development (TD) cross-linguistically, even in the context of bilingualism (Schwob et al., 2021). Cantonese has been the sole cross-linguistic exception, where NWR does not differentiate between TD children and those with DLD (Leung, 2010; Stokes et al., 2006). No studies have followed-up with this line of research, thus it remains unknown whether the divergent findings on Cantonese are based more heavily on typological differences between Cantonese and other previously examined languages, or methodological factors in Cantonese NWR studies.

This thesis revisits NWR in Cantonese in three studies.

Study I: Study I tested the hypothesis that the previous lack of significant findings on Cantonese NWR may be related to certain aspects of nonword stimuli design, including low levels of nonword lexicality and long nonword length. Three novel sets of NWR stimuli, which take into account factors known to affect NWR performance and group differentiation, were reported. Sixteen TD-DLD pairs of monolingual Cantonese-speaking children repeated two sets of High-Lexicality nonwords, and one set of Low-lexicality nonwords, which could be further analysed on sub-lexicality, based on CV combination attestedness (i.e., whether nonword syllables contained CV combinations that are attested in Cantonese). Children with DLD scored significantly below their TD peers, and effect sizes showed that nonwords with high levels of lexicality and sub-lexicality offered greater TD/DLD group differentiation, suggesting that Cantonese is not a true cross-linguistic exception in NWR. Future work could aim to replicate the present findings on a larger sample size, verify whether TD/DLD group differences are still captured by NWR in younger, Cantonese-speaking children, and examine the diagnostic accuracy of this NWR test.

Study II: Following positive findings on monolingual Cantonese-speaking children reported in Study I, Study II investigated whether Cantonese NWR stimuli are able to avoid disadvantaging bilingual second language (L2) TD children, compared to monolingual TD (MonTD) children, as L2-TD children are at risk of being

misclassified as children with language disorder in NWR, due to having reduced language knowledge and experience in the testing language to support NWR. NWR performance in 19 MonTD, 19 monolingual DLD (MonDLD) and 19 bilingual L1-Urdu-L2-Cantonese TD children (L2-TD) was examined on three sets of language-specific nonwords (reported in Study I) and one set of quasi-universal nonwords. When NWR accuracy was scored at whole-nonword level, language-specific, High-Lexicality nonwords captured group differences between DLD and TD groups (both monolingual and L2), while not disadvantaging the L2-TD group, compared to MonTD children. When NWR accuracy was scored at syllable level, quasi-universal, CL-NWR nonwords were the only set of stimuli that did not disadvantage L2-TD children relative to MonTD children, while still yielding significant group differences between the MonDLD and L2-TD groups. These findings provide evidence from a typologically distinct and understudied language that NWR has potential to disentangle the effects of language impairment (in MonDLD) and bilingualism (in L2-TD), and has potential to be further developed into clinically informative tools for DLD. Future studies can explore how an L2-DLD group performs relative to the examined groups, explore alternative bilingual groups of children acquiring other L1-L2 combinations, and explore sensitivity and specificity of Cantonese NWR in a bilingual context.

Study III: With a recent study suggesting that sub-lexical representations may have a facilitative effect on NWR, above and beyond that of lexical representations (Szewczyk et al., 2018), Study III examined whether sub-lexicality in Cantonese NWR stimuli affected Cantonese-speaking children's NWR performance and TD/DLD group differentiation. Data on NWR accuracy of 19 DLD-TD pairs of monolingual speaking children were analysed with two measures of sub-lexicality – CV combination attestedness, as positive first findings have been reported on this sub-lexical measure (both in Stokes et al., 2006, and Study I); and neighbourhood density (ND) of syllables, a newly proposed sub-lexical measure, as ND has been suggested as a strong cue to Cantonese word-likeness. Both measures significantly predicted NWR performance in Cantonese-speaking children, with ND of Syllables being a stronger predictor. CV combination attestedness interacted with participant group, where only syllables with attested CV combinations captured significant TD/DLD group differences. There was no evidence that ND of Syllables affected the degree of TD/DLD group differentiation in these Cantonese-speaking children. These findings suggest that sub-lexical

representations play a crucial role in Cantonese NWR, either alongside or above and beyond the influence of lexical representations. Future studies can examine whether ND of Syllables predict NWR performance in children acquiring other languages, to confirm whether the findings on this novel sub-lexical measure can be generalised cross-linguistically; and also examine an additional sub-lexical measure, phonotactic probability, as a predictor of NWR performance in Cantonese-speaking children, in addition to the two examined measures.

Conclusions: Cantonese is not a true cross-linguistic exception in NWR - NWR is able to capture significant group differences between Cantonese-speaking children with and without DLD, and Cantonese CL-NWR nonwords, as well as Cantonese language-specific High-Lexicality nonwords, demonstrated potential to disentangle the effects of language experience and language impairment. Factors affecting NWR accuracy in children acquiring other languages also affect children acquiring Cantonese, particularly in terms of nonword lexicality and sub-lexicality. The findings suggest that Cantonese NWR stimuli have the potential to be developed into clinically informative assessment tools for DLD in both monolingual and bilingual Cantonese-speaking children, but further research is needed to examine whether the present findings can be replicated in a larger sample size, generalised to younger children and bilingual children acquiring languages other than Urdu as their first language; and sensitivity and specificity of Cantonese NWR tests will also need to be examined for further development of NWR tasks into clinical assessment tools.

PUBLICATIONS ARISING FROM THE THESIS

Fu, N.C., Chan, A., Chen, S, Polišenská, K. & Chiat S. (in prep). The effect of neighbourhood density on nonword repetition performance in Cantonese-speaking children with and without Developmental Language Disorder. Target journal TBC.

Fu, N.C., Chen, S., Polišenská, K., Chan, A., Kan, R. & Chiat S. (resubmitted). Nonword repetition in children with Developmental Language Disorder: Revisiting the case of Cantonese. Submitted to *Journal of Speech, Language, and Hearing Research*

Fu, N.C., Chan, A., Chen, S, Polišenská, K. & Chiat S. (under revisions). Revisiting nonword repetition as a clinical marker of Developmental Language Disorder: Evidence from monolingual and bilingual L2 Cantonese. Research topic on "Developmental Language Disorder in Chinese: Status of the research landscape and new frontiers". *Brain and Language*.

With Polišenská, K., Chiat, S., Szewczyk, J. and researchers in the European bi-SLI consortium involving 18 research teams spanning 15 countries. (in prep). Is the Crosslinguistic Nonword Repetition Test valid crosslinguistically? Evidence from performance in different language groups and countries. Order of authorship and target journal TBC.

ACKNOWLEDGEMENTS

First of all, I would like to thank my chief supervisor, Dr. Angel Chan, for your continuous guidance and support throughout my MPhil studies. Thank you for always being available and taking the extra step to support me, whether for academic or personal matters; and for believing in me, even at my lows. I would also like to thank my co-supervisors, Dr. Sarah Chen, Prof. Shula Chiat, and Dr. Kamila Polišenská, for your invaluable patience and feedback. Thank you for generously providing your expertise throughout my studies. Thank you also to my board of examiners, Prof. Ute Bohnacker, Dr. Natalia Meir, and Dr. Yao Yao, for your very insightful and encouraging comments you have given me, which are incredibly helpful for this thesis and publications arising from this MPhil study.

Thanks also goes to speech therapists, Bonnie Chan, Carman Lai, and Sylvia Wan; student speech therapists, Hazel Chau, Christy Chu, Eva Kwan, Annie Lam, Angel Wo, Jovial Wong, and Joanne Yiu; research assistants, Ashley Chan, Sonia Hui, Circle Kai, Mars Kwan, Maggie Lee, Cheri Li, Imran Muhammad, and Katrina Wong; and programmer, Roland Shum; who supported the development of project materials and the running of the experiments.

I would also like to extend my gratitude to all children and families who took part in the studies.

To fellow and ex- graduate students, Jane, Saboor, Saira, Vasundhara and Yi, thank you for your warm company and kind advice – I wholeheartedly enjoyed the laughter (and food) we shared along the way.

Finally, I would like to express my deepest thanks to family and friends around me, particularly to my mother, who always puts in her all to take care of the family, and acts as the best role model in my life; to my partner, Joshua, for his unwavering believe in me, and his love and support throughout the years; and to my dearest friends Corey, Czarina, John, Kimberley, Liwen, Tobie, Wilson, and Winton, thank you for the emotional support you have continuously given me and for the fun you add to my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I – Literature review

1.1 Introduction

Developmental Language Disorder (DLD), where children have difficulties understanding and using language with no associated biomedical conditions, is a neurodevelopmental condition that has been documented since the early 1800s (Leonard, 2020), but has only gained substantial attention in research in the recent decades. Such increase in attention, both in the research and clinical fields, has taken place in parallel to societal changes over time. In the 19th and early 20th centuries, individuals with language impairments were less likely to be considered disabled, due to an abundance of "blue collar" jobs (e.g. production process workers and labourers) in the labour market. As the nature of dominant industries shifted from "blue collar" to "white collar" in the 20th century, oral and written language skills became a requirement in the work force (Ruben, 2000), making language skills critically important for economic prosperity and the mental and social well-being of individuals. The rising importance of language in everyday life has motivated the increase in research efforts on language disorders such as DLD, which is now considered to be a form of hidden disability.

One important area of research under the topic of DLD is early identification of this disorder in young children, such that children with DLD could benefit from early intervention from speech and language therapists and educators. In particular, a relatively simple task of *Nonword Repetition* (NWR) has gained great interest in the field of assessments for DLD, as it has demonstrated potential to be a clinical marker of DLD (Bishop et al., 1996; Conti-Ramsden et al., 2001; Dollaghan & Campbell, 1998). In NWR, children simply listen to and immediately repeat nonsense words (i.e. novel phonetic sequences), such as *bamudi* and *duligasumu* – the task has been shown to be effective at differentiating between age-matched groups of children with DLD and those with typical development (TD) cross-linguistically (Schwob et al., 2021). However, most of the relevant research has focused solely on the population of monolingual English-speaking children, and only in the 2010s, has more research attention been given to culturally and linguistically diverse populations, including non-English-speaking and bilingual children, in the context of evaluating the usefulness of NWR tasks within an assessment tool kit for assessing children for DLD. This thesis will fill this gap through investigating the potential for NWR to capture group differences in children with and without DLD among Cantonese-speaking children,

adding a unique typological perspective to the topic by examining a largely understudied cultural and linguistic population of Cantonese-speaking children in NWR.

This chapter first provides an introduction on the history, definitions, prevalence, and outcomes of DLD. It then reviews the literature on the potential for NWR to be a clinical marker of DLD and the mechanisms behind this potential, focussing on the strengths of NWR as an assessment tool for DLD, followed by an overview of the underlying processes tapped in NWR tests, and an outline of the factors affecting NWR performance. It also covers in detail, previous studies on NWR in monolingual Cantonese-speaking children, monolingual Mandarin-speaking children (as Mandarin is typologically similar to Cantonese), and bilingual children, as these will be the linguistic populations of focus in this MPhil study. Finally, the chapter will introduce the research questions of this thesis, together with descriptions of three studies that will be used to address the topic of whether NWR has the potential to capture group differences in children with and without DLD among a Cantonese-speaking population.

1.2 Developmental Language Disorder

1.2.1 Terminology and Diagnostic Criteria for Developmental Language Disorder

Developmental Language Disorder (DLD) is a common neurodevelopmental disorder involving significant and unexplained difficulties in learning, understanding and using language (Bishop et al., 2016, 2017). Previously, the condition has been labeled with a variety of terminology, such as language impairment, language disorder, language delay, developmental aphasia, and most commonly, specific language impairment (SLI). Historically, the term, SLI, has been used to depict a significant language impairment occurring in the absence of other developmental deficits, particularly in non-verbal intelligence quotient (NVIQ). That is, a child with SLI should have a discrepancy between their verbal and non-verbal abilities, commonly reflected by a discrepancy in standardised test scores, where language scores fall at least 1.25 standard deviations (SD) below the mean, but NVIQ scores fall no more than 1 SD below the mean. While such diagnostic criteria was initially favoured for the purpose of narrowing participant groups in experimental research (i.e. those with a “specific” language impairment), more recently, the exclusionary criterion has been criticized by different scholars to be largely discredited, conceptually unsound, and misinformed (Ebbels, 2014b).

In a 2014 journal special issue on SLI (Ebbels, 2014a), articles and commentaries thoroughly discussed the problematic nature of the term, *SLI,* and the exclusionary criteria associated with the term. Firstly, the term, SLI, has been coined in a time when standardised testing has yet to be available and classic epidemiological studies on the condition have yet to be published. Rather than being informed by data from large populations, descriptions of the condition were based on early clinical observations on single or group cases studies (Reilly et al., 2014). The exclusionary criteria for SLI were also criticized to be conceptually unsound – taking data from two epidemiological studies (Reilly et al., 2010; Tomblin et al., 1996), when verbal and nonverbal scores of children were plotted against each other, many data points were clustered around the cut-off lines, and that moving the cut-off criteria by as little as one point would lead to a change of group classification for a large number of children (Reilly et al., 2014). There was also no consensus to the exact cut-off point for NVIQ scores in the classification of SLI, with many researchers using cut-off points between NVIQ scores of 75 and 80 (i.e. 1 or 2 SD below the mean; Ebbels, 2014b). Furthermore, there is no evidence suggesting that children with higher or lower NVIQ scores respond differently to language interventions (Ebbels et al., 2014; Reilly et al., 2014). Therefore, applying an exclusionary criterion for the classification of SLI is unproductive and may have a profound impact on a large number of children who could benefit from relevant services that they are unable to access due to a stringent and arbitrary exclusionary criterion.

In addition, *SLI* was not the only label used to describe unexplained language problems in children – Bishop's (2014) survey on labels used to describe such language difficulties found 132 different terms in use, with 33 of which resulting in over 600 returns on Google Scholar. The lack of consensus in both the terminology and diagnostic criteria created confusion, hindered research progress, and interfered communication between researchers, practitioners, and the public.

These discussions laid the groundwork for two major studies by the CATALISE consortium led by Bishop (2016 & 2017), where the Delphi technique is used to achieve consensus on the terminology and criteria for identifying children with unexplained language difficulties. With the Delphi technique, a panel of experts from various disciplines, including speech and language therapy, psychology, education, pediatrics, psychiatry, audiology, as well as charity representatives from organizations with a primary focus on supporting families affected by children's language impairments, evaluated statements based on the articles and commentaries from the 2014 journal

special issue on SLI (Ebbels, 2014a). With reference to ratings of and comments on the relevance and validity of each statement, moderators removed, combined, and modified the statements for further round of evaluations, until consensus (i.e. when 75% or more of the panel rated each statement either as "strongly agree" or "agree") was reached among the panel. In both Delphi studies (the first targeting criteria and the second targeting terminology for children's language problems), consensus among the panel was achieved in two rounds of evaluations. The research team endorsed the use of the terms, *Language Disorder*, for referring to a profile of language difficulties associated with poor prognosis and causing negative functional impact in everyday life; and *Developmental Language Disorder* or *DLD*, for describing the language disorder when it is not associated with a known biological etiology.

Reflecting on the current understanding and recent discussions of the disorder, the main outcomes of the Delphi study on the criteria for DLD included that: the diagnostic criteria have transitioned from being solely reliant on imposed cut-off points in standardised testing, to taking a combined approach, where standardised test scores are considered together with caregiver reports and clinical observations to form an overall judgment on the functional impact and prognosis of a language difficulty. The new criterion also takes into account the language learning context of a child, such as whether a child is acquiring an additional language bilingually, and recognises lack of familiarity with the additional language(s) as a *speech, language and communication need* that is independent of language disorders. Importantly, the new criteria for DLD no longer require a discrepancy between verbal and non-verbal abilities. Based on the new consensus, Bishop and her colleagues proposed the following pathway for classifying children with various speech, language and communication needs and language disorders (see Figure 1).

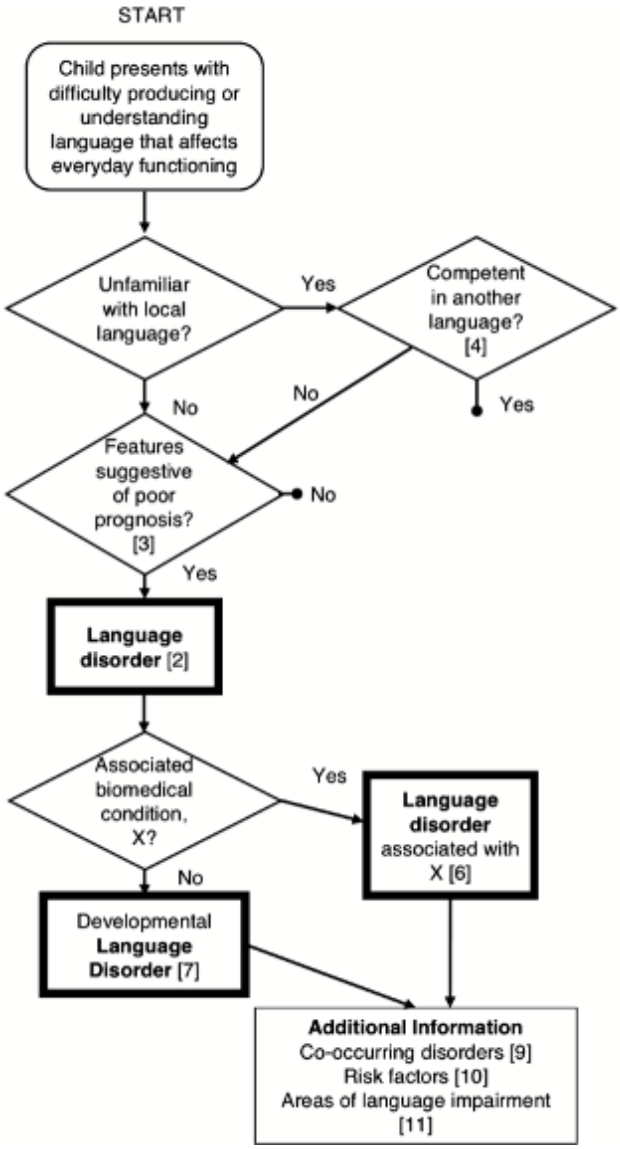

Figure 1. *Flow chart illustrating pathways to diagnosis of language disorder. Reprinted from "Phase 2 of CATALISE: a multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology," by Bishop, D., Snowling, M., Thompson, P., Greenhalgh, T. & the CATALISE-2 consortium, 2017, Journal of Child Psychology and Psychiatry, 58(10), p.1075. 2017, Copyright 2017 by Wiley-Blackwell.*

1.2.2 Prevalence, Associated Conditions and Outcomes of DLD

DLD is a high-incidence disability that is estimated to have a prevalence rate of 7.4% (Tomblin et al., 1997) and 7.58% (Norbury et al., 2016) in children at school entry in English-speaking countries.

DLD can co-occur with other common developmental disorders such as attention-deficit and hyperactivity disorder (ADHD), dyslexia, or learning disabilities (Young et al., 2002), as well as subtle weaknesses in motor development in preschool children (Cheng et al., 2009). DLD alone, or together with its co-occurring conditions, increase the chances for poor literacy and academic attainment in children, as language is the foundation to learning in a school environment.

Additionally, language is also fundamental to building meaningful relationships with others, thus DLD is also associated with increased chances of socioemotional difficulties. Relative to their TD peers, children with DLD are reported to be more likely to have social and peer problems, such as bullying or being victimized by others (Forrest et al., 2020; van den Bedem et al., 2018). Children with DLD are also more likely to demonstrate both internalizing problems, such as anxiety and depression, and externalizing behaviours, such as aggression, and such problems tend to be more pronounced as children with DLD grow older (Curtis et al., 2018). With DLD being a life-long condition, DLD can continue to have downstream effects on individuals in their adulthood. Adults with DLD have been reported to have lower academic and vocational qualifications, less skilled employment and less full-time employment than adults without a history of DLD (Clegg et al., 2005), although educational opportunities have improved for young adults with DLD in the recent years, and other studies have reported no differences between the annual income for adults with and without DLD (Conti-Ramsden et al., 2018).

1.3 Nonword Repetition as a Potential Clinical Marker of DLD

Nonword Repetition (NWR) tests require children to repeat novel phonetic sequences (i.e. nonwords; e.g. *bamudi*) immediately after hearing them. NWR has been advocated as a potential clinical marker for DLD (Bishop et al., 1996; Boerma et al., 2015; Conti-Ramsden et al., 2001; Dollaghan & Campbell, 1998), as it has been shown to be able to differentiate between TD children and those with DLD cross-linguistically (for example, English: Conti-Ramsden et al., 2001; Dutch: de Bree et al., 2010; Italian: Dispaldro et al., 2013; French: Ferré & Santos, 2015; Swedish: Kalnak et al., 2014;

Vietnamese: Pham & Ebert, 2020; Mandarin: Wang & Huang, 2016, amongst others). Meta-analyses on NWR have demonstrated that children with DLD scored, on average, 1.27 standard deviations below their TD peers (Graf-Estes et al., 2007), and that NWR is able to produce large effect sizes for group differences, on both monolingual and bilingual populations (Schwob et al., 2021).

1.4 What Does Nonword Repetition Measure?

Despite being a deceptively straightforward test, the successful repetition of nonwords involves a range of underlying processes. Prior to NWR being established as a potential clinical marker of DLD, researchers first started using NWR tests, in the late 1980-1990s, as a tool to explore the deficits that underlie language impairments in children. To answer such question, researchers must first come to an understanding of what NWR tasks actually measure.

1.4.1 Phonological Encoding

Early accounts suggest that NWR is a measure of phonological encoding, with researchers like Kamhi and her colleagues (1986, 1988) claiming that there are limited storage and retrieval demands in a task requiring immediate repetition.

1.4.2 Phonological working memory

NWR was later proposed as a strict measure of phonological working memory (Gathercole & Baddeley, 1990), as more studies reported a decrease in NWR performance when children repeated longer, multi-syllabic words.

1.4.3 Phonological Processing

Both of the accounts above were later criticized by Snowling et al. (1991), who suggested that NWR involves complex mechanisms of speech perception, speech segmentation, creation of a robust phonological representation in working memory, as well as the formulation and execution of an articulation plan; while these processes tap on phonological encoding and phonological working memory, they should also be considered individually, thus NWR should not be considered as a simple measure of storage capacity of phonological working memory.

Snowling's commentary has prompted further studies that have examined the

relative contributions of phonological memory and phonological sensitivity to NWR, and there were competing accounts on whether NWR is supported by phonological sensitivity (Metsala, 1999) or general phonological processing (i.e. both phonological sensitivity and phonological memory; Bowey, 1996, 1997).

1.4.4 Vocabulary Knowledge

Aside from the criticism that NWR should not be taken as a simple measure of phonological working memory capacity, Snowling et al. (1991) also suggested that NWR should not be taken as a content-free measure of phonological memory, as implied in the original account from Gathercole & Baddeley (1990). This is because, during NWR, the knowledge of phonological, morphological and prosodic regularities in long-term lexical knowledge could be used to support the repetition of unfamiliar, novel phonetic sequences, as the memory trace of the target (i.e. nonword), decays rapidly within working memory. Over time, many studies have provided substantial evidence that NWR performance is indeed mediated by lexical knowledge (Chiat & Polišenská, 2016; Chiat & Roy, 2007; Dollaghan et al., 1993; Dollaghan et al., 1995; Munson, 2001; Thordardottir & Brandeker, 2013; Zamuner et al., 2004).

1.4.5 Summary of Skills Measured by Nonword Repetition

Throughout the 1990s when NWR was mainly used as a window to explore deficits underlying the condition of DLD, NWR has been suggested to be a measure of phonological encoding, phonological working memory, phonological sensitivity, general phonological processing, and vocabulary, with researchers failing to reach agreement over which specific skill NWR measures. Later in the 2000s, there was greater consensus among researchers that NWR likely involves all of the above skills and more (Briscoe et al., 2001; Coady & Evans, 2008; Snowling et al., 1991), thus there has been a change in the direction of focus in studies on NWR, from attempting to pinpoint the specific skill(s) the task measures, to examining the practical utility of the task for identifying DLD or a risk of DLD in children, by capitalizing on the multi-process nature of the task. Because children with DLD have heterogeneous profiles with a diverse range of language difficulties (Bishop et al., 2017), deficits in any component processes of NWR, including auditory perception, phonological encoding and assembly, storage of phonological representations, motor planning, and articulation (Coady & Evans, 2008), may be reflected in lower NWR scores, making NWR an

informative tool for detecting potential language disorders in children.

1.5 Strengths of NWR tasks

The use of nonwords in NWR tasks is associated with a number of unique strengths, when compared to more traditional linguistic assessments. NWR tasks have been described to be less heavily influenced by prior linguistic knowledge (although it is not free of linguistic-content, as established above) because it assesses the ability to process novel information (Archibald, 2008), making it appropriate for testing very young children (Guiberson & Rodríguez, 2013), or bilingual children who have weaker language skills in the target language due to limited exposure (Boerma et al., 2015; Chiat, 2015). Findings from meta-analyses also suggest that TD/DLD group differences could be yielded by NWR across a wide age range (Graf-Estes et al., 2007; Schwob et al., 2021), making NWR a dependable tool for assessing children to young adolescents, across different stages of development. Most evidence also suggest that NWR performance is not influenced by other internal and external factors, including gender and socioeconomic status (Boerma et al., 2015; Chiat & Roy, 2007; Washington & Craig, 2004), hence NWR is suitable for testing a wide population with heterogeneous characteristics.

1.6 Clinical Accuracy of Nonword Repetition

Furthermore, studies have reported adequate levels of sensitivity and specificity values for NWR tasks, where most older studies reported values ranging from about 80% to 100% (Plante & Vance, 1994). Some recent studies have reported lower sensitivity values of approximately 40% when NWR was used in a monolingual context, but have also reported that the diagnostic accuracy could be significantly improved when combined with other assessment tools, such as tests of receptive vocabulary, expressive vocabulary, and intelligence (Thal et al., 2005). Similarly, some studies have also reported low sensitivity levels of NWR when used on a bilingual population (Santos & Ferré, 2018), but the use of a quasi-universal, cross-linguistic NWR task, rather than language-specific NWR tasks (see further discussions on the distinction between quasi-universal and language-specific NWR tasks in Chapter III, Sections 3.1.2-3.1.3), allows for NWR to be diagnostically accurate even in a bilingual population, at least at age five to six years (Boerma et al., 2015). Less clear-cut findings have been reported on children above six years of age, where longitudinal data from

the same children included in Boerma et al. (2015), when tested at six to seven years (one year after initial testing) and seven to eight years (two years after initial testing), suggested that both sensitivity and specificity of NWR dropped below clinically acceptable levels (<80%). With that said, when NWR was combined with a narrative assessment measure (with components of narrative production, including story-telling and retelling, and comprehension), clinically accuracy improved to acceptable levels (>80%; Boerma & Blom, 2021). Overall, it is well established that NWR is an informative and recommendable tool that deserves a fixed position in language assessment batteries for children (Schwob et al., 2021), although it may have limited diagnostic value when used as a standalone assessment tool (Ortiz, 2021).

1.7 Factors Affecting Nonword Repetition Performance: Item-Related Factors

As more NWR tests have been developed, studies have identified a number of nonword characteristics that can influence NWR accuracy.

1.7.1 Length

Given the importance of phonological working memory capacity in NWR, nonword length has been very consistently found to affect NWR performance, where accuracy declines as nonword length increases and gradually raises processing demands within children's limited working memory capacity (e.g. Bortolini et al., 2006; Chiat, 2015; Chiat & Roy, 2007; Dispaldro et al., 2013; Jones et al., 2010).

1.7.2 Lexicality

With studies suggesting that lexical knowledge mediates NWR performance, it is also unsurprising that the lexicality of nonwords (i.e. the degree to which nonwords resemble real words in an ambient language, in terms of whether morphemic elements are incorporated into nonword stimuli) can influence NWR accuracy, as nonwords with phonological, morphological and prosodic patterns that resemble real words could maximise the support provided by long-term lexical knowledge during NWR through redintegration (Gathercole, 1999; Schweickert, 1993), where lexical representations are drawn on, to restore the incomplete trace of the stimuli. Studies have found better NWR performance on nonwords that contain real lexical items or morphological markings within items than those that do not (Casalini et al., 2007; Graf-Estes et al., 2007; Jones

et al., 2010). To illustrate with examples in English, compare items in Gathercole and Baddeley's (1996) test (e.g. *defermication*), which contain morphemes /di/, /fɜm/ and /eɪʃən/, with items in Dollaghan and Campbell's (1998) test, which do not incorporate real morphemes (e.g. /nɑɪtʃɔɪtɑʊvub/) by design – items from Gathercole and Baddeley's (1996) test can be described as having higher lexicality and better performance has been reported on such type of nonwords.

1.7.3 Sub-Lexicality

In the same way that lexical representations from long-term memory support NWR, sub-lexical representations also support redintegration processes during the repetition of nonwords. At a sub-lexical level, higher NWR accuracy was seen on nonwords containing syllables with high phonotactic probability (PP) compared to those with low PP (McKean et al., 2013; Szewczyk et al., 2018), and on nonword syllables containing attested consonant-vowel (CV) combinations compared to those with CV combinations that never appear in the language (Stokes et al., 2006). Evidence has even been reported that the sub-lexical measure, mean ngram frequency of all phonemes (which measures phonotactic probability of phoneme sequences at all grain sizes within a nonword, such as bigrams, trigrams, and so on) is a stronger predictor of NWR performance than other lexical factors such as neighbourhood density (which measures the degree to which a nonword, in its entirety, phonologically resembled real words in a given language), pointing to the primacy of sub-lexical representations in successful NWR (Szewczyk et al., 2018).

1.7.4 Segment Complexity

Related to the importance of phonological sensitivity, general phonological processing, and motor planning abilities in NWR, segmental complexity of nonwords have also been found to affect NWR accuracy, where children were reported to have more difficulty repeating nonwords containing consonant clusters than those that do not (Archibald & Gathercole, 2006b; Jones et al., 2010), as well as nonwords containing a higher number of consonants than those with fewer consonants (Szewczyk et al., 2018), as higher levels of segmental complexity of the nonwords pose greater processing demands with more complex phonological analysis and motor planning.

1.7.5 Prosodic Effects

Furthermore, prosodic features have also been found to affect NWR accuracy – Chiat and Roy (2007) reported that children were more likely to omit unstressed than stressed syllables, suggesting that nonwords with elements that are less perceptually salient can be more challenging in the context of NWR. Furthermore, Sahlén et al. (1999) also reported greater NWR accuracy when nonwords had stress patterns that were typical in the ambient language.

1.8 Nonword Repetition in Children with Developmental Language Disorder

For children with DLD, NWR is particularly challenging because DLD is associated with various impairments that are relevant to the underlying processes necessary for successful NWR. Children with DLD have been suggested to have more limited phonological working memory capacity (Gathercole & Baddeley, 1990; Montgomery, 2002), limited general processing capacity (Marton & Schwartz, 2003), weaker processing of phonetically non-salient features (Leonard, 1989), limitations in the construction of accurate phonological representations (Edwards & Lahey, 1998), poorer vocabulary skills (Gray et al., 1999; McGregor, 2009; Watkins et al., 1995), and motor planning deficits (Stark & Blackwell, 1997), all of which pose challenges to children with DLD in some or all of the component processes in NWR. Therefore, factors known to affect adversely NWR performance are often found to be disproportionally challenging for children with DLD.

Due to limitations in working memory storage and processing capacity, children with DLD have been reported in some studies to be more affected by increases in nonword length than TD children, as seen from greater drops in NWR performance as nonword length increases, compared to their TD peers (Dispaldro et al., 2013; Graf-Estes et al., 2007; McKean et al., 2013). Children with DLD were also found to be more negatively affected by the presence of consonant clusters in nonwords than TD children (Bishop et al., 1996; Briscoe et al., 2001; Jones et al., 2010; Santos & Ferré, 2018), perhaps related to both weaker phonological analysis and motor planning skills. Also, children with DLD were found to be more likely to omit non-accented syllables, compared to TD children (Chiat & Roy, 2007), possibly indicating weaker processing of phonetically non-salient features.

The disproportionate challenges posed by NWR on children with DLD are what allows the task to differentiate between children with DLD and their TD peers, giving

NWR the potential to be a clinical marker of DLD.

1.9 Nonword Repetition in Cantonese-Speaking Children

Interestingly, Cantonese has been a rare exception in NWR, where NWR has been shown to be unable to differentiate between TD children and those with DLD. In a study by Stokes et al. (2006), native Cantonese-speaking children aged four to five years, who were diagnosed with DLD, were compared with age-matched TD children and younger, language-matched TD children (aged two to three years) in NWR performance. While an age effect was found between the groups, where younger TD children achieved significantly lower accuracies in NWR than the remaining groups, no significant differences were found between the DLD group and TD age-matched group (although the directionality of the results was consistent with expectations, where TD children repeated nonwords with higher accuracy than children with DLD, albeit non-significantly).

Similarly, in an unpublished thesis by Leung (2010), which tested five- to seven-year-old children's NWR performance using the pseudo-morpheme subtest from the Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS; T'sou et al., 2006), no significant differences were found between Cantonese-speaking TD children and those with DLD.

Stokes et al. (2006) proposed that one possible reason behind the discrepant findings with NWR in Cantonese speakers is that Cantonese has relatively simple phonotactic structures, with a lack of consonant clusters and a syllable-timed prosodic pattern. When Cantonese NWR stimuli were designed with reference to such phonotactic constraints, Cantonese nonwords would naturally also have lower levels of segmental and suprasegmental complexity, making them less taxing on working memory than nonwords designed with reference to other examined languages, like English and Swedish, which have more complex phonotactic structures. This may have resulted in Cantonese nonword stimuli being unable to capture TD/DLD group differences.

1.10 Nonword Repetition in Mandarin-Speaking Children

More recent studies have also examined NWR in Mandarin, which is typologically similar to Cantonese and also has relatively simple phonotactic structures (Lin & Wang, 1992), but have found good differentiation between TD and DLD groups (Chi, 2007;

Wang & Huang, 2016).

On the other hand, there are also Mandarin findings that depart from cross-linguistic patterns in other ways. In Chi (2007), children with DLD (specific participant ages were not stated; all children were studying Grade 1 at primary school, thus expected to be approximately six to seven years of age) showed weaker phonetic discrimination (in addition to poorer NWR) than the TD group, contradicting claims (based on studies where there were no such differences, e.g. Edwards & Lahey, 1998) that poor encoding rather than discrimination was the source of difficulty for children with SLI in NWR measures.

In Li and Cheung (2014), who tested TD children speaking Taiwanese Mandarin, children were examined in the repetition of single and multi-syllabic non-words. A productive phonology task (measuring the quality/accuracy of producing phonemes and syllable shapes elicited through a picture-naming task) was included as a measure of phonological analysis, where poor analysis meant children were less accurate at encoding phonological information. For single-word stimuli, phonological analysis better predicted NWR performance than storage capacity (as measured by a digit span task), while digit span was the better predictor on stimuli of two or three syllables. Phonological analysis had a bigger effect than storage for repeating non-words containing unattested syllables (which requires more processing resources) than attested syllables, suggesting an interaction between the effects of sub-lexicality and the reliance on phonological processing.

1.11 Nonword Repetition in Bilingual Children

Aside from monolingual children with DLD, NWR has also been suggested to be a useful tool for assessing bilingual children with and without DLD. Given that bilingual children acquire multiple languages, they inevitably receive less exposure to each language they are acquiring, when compared to their monolingual counterparts. This is also further exacerbated in children who may be acquiring a minority language in society, where they are put under especially reduced input conditions for learning the language (such as only being able to receive input from the language at home, when another majority language is typically used in schools and in the community), or when they acquire the target language as a second language. Therefore, bilingual children are likely to have weaker knowledge of a language, compared to monolingual children of the same age, which puts bilingual children in a similar position to monolingual

children with DLD in the context of NWR, where they are less able to draw on long-term, lexical and sub-lexical representations during NWR, compared to monolingual TD children.

Indeed, some studies have reported that bilingual TD children do not perform as well as monolingual TD children in NWR (Boerma et al., 2015; Kohnert et al., 2006; Sharp & Gathercole, 2013); although others also found no differences between monolingual and bilingual TD groups (Armon-Lotem & Chiat, 2012; Lee & Gorman, 2013; Messer et al., 2010), suggesting that only in certain L1-L2 combinations and certain NWR tasks can the gap between monolingual and bilingual TD children be narrowed (Armon-Lotem, 2018). Importantly, bilingual children with DLD were reported to perform at an even lower level of accuracy in NWR than bilingual TD children (Armon-Lotem, 2018; Schwob et al., 2021), thus bilingual TD children do seem to be relatively unimpaired in NWR, despite having reduced language experience and weaker language knowledge. Therefore, NWR has demonstrated some potential to disentangle the effects of language impairment and reduced language experience. To date, no studies have examined NWR performance in bilingual Cantonese-speaking children, and it is yet to be determined whether bilingual Cantonese-speaking TD children perform more similarly to monolingual TD Cantonese-speaking children, or monolingual Cantonese-speaking children with DLD in NWR.

1.12 This Thesis

Chapter I has established that cross-linguistically, NWR can be a promising tool for identifying DLD in both monolingual and bilingual children, due to the nature that it taps a wide range of underlying processes that are often impaired in children with DLD. Most of these findings, however, were based on studies examining speakers of Indo-European languages like English, French, and Dutch etc., with much less work being done on other language families. In fact, in two studies on NWR in Cantonese-speakers, NWR was found to be unable to differentiate between four- to six-year-old TD children and children with DLD (Stokes et al., 2006), and between five- to seven-year-old children with and without DLD (Leung, 2010). Therefore, it remains unknown whether the appropriateness of using NWR as an assessment tool for identifying DLD can truly be generalized cross-linguistically.

As no studies have followed-up with this line of research, the present thesis sought to revisit the topic of NWR in Cantonese. Specifically, this thesis aims to address three

main research questions:

1) Whether NWR can capture significant TD/DLD group differences in Cantonese-speaking children, consistent with cross-linguistic findings, when an alternative set of nonword stimuli to those used previously is adopted;

2) Whether Cantonese-based nonword stimuli can minimize disadvantaging children with TD who are acquiring Cantonese as a second language (L2) bilingually, compared to children with TD and children with DLD acquiring Cantonese monolingually, thus evaluating the potential of Cantonese NWR to be developed as an accurate and informative assessment tool for differentiation between TD and DLD, even for bilingual Cantonese-speaking children;

3) How sub-lexical factors in Cantonese nonword stimuli affect NWR performance and group differentiation, given that recent studies suggest sub-lexical representations to be fundamental to successful NWR.

Three studies addressed each of these research questions.

Study I, titled "**Is Cantonese a true cross-linguistic exception in nonword repetition?**", explored whether previous inconsistent findings between NWR in speakers of Cantonese and other examined languages were more heavily influenced by typological differences between these languages, or methodological differences in studies on Cantonese vs. other examined languages, specifically concerning certain nonword characteristics of the stimuli used in previous Cantonese NWR studies. Three sets of novel Cantonese-based nonword stimuli with varying levels of lexicality would be presented, and they are evaluated on their ability to generate group differences between monolingual Cantonese-speaking TD and DLD groups. The design of these newly developed NWR stimulus sets and the study findings will be presented in Chapter II of this thesis.

Study II, titled "**Disentangling effects of language impairment and bilingualism with Cantonese nonword repetition stimuli**", built on positive findings from Study I, and extended the analysis on TD and DLD groups among monolingual Cantonese-speaking children to also include bilingual second language TD (L2-TD) learners of Cantonese. Due to having reduced language experience and proficiency associated with bilingual language acquisition, some L2-TD children may demonstrate language difficulties similar to that in monolingual children with DLD, and are at risk of being disadvantaged relative to monolingual TD children in NWR tests. Study II

evaluates whether L1-Urdu-L2-Cantonese bilingual TD children will be disadvantaged by the three sets of novel Cantonese-based (i.e. language-specific) nonword stimuli presented in Study I, and an additional, fourth set of Cantonese-adapted, quasi-universal nonword stimuli. Study II also investigates the effects of using different scoring methods in NWR on the pattern of group differentiation. The findings are presented in Chapter III of this thesis.

Study III, titled "**Sub-lexical effects on nonword repetition in Cantonese-speaking children**", examined two sub-lexical measures of nonwords and their effects on Cantonese-speaking children's NWR performance and TD/DLD group differentiation, given that recent studies suggest sub-lexical representations to be fundamental to successful NWR, alongside lexical representations. The two measures examined were CV combination attestedness, as some positive first findings with this measure have been reported in Stokes et al. (2006); and neighbourhood density (ND) for syllables, a novel measure of sub-lexicality presented in this study, and developed based on claims by Cantonese phonologists that ND is a strong cue to Cantonese word-likeness, due to the unique phonotactic properties of Cantonese. Findings on these two sub-lexical measures will be presented in Chapter IV.

Finally, Chapter V of this thesis provides a summary and discussion of findings from each of the three studies, and discusses their implications on research on the clinical utility of NWR for assessing DLD, in a Cantonese context and beyond.

CHAPTER II – Study I
Is Cantonese a True Cross-Linguistic Exception in Nonword Repetition?

2.1 Introduction

NWR has been advocated as a potential clinical marker of DLD, as it has consistently been shown to be able to differentiate between TD children and those with DLD cross-linguistically, with Cantonese being the only reported exception (Leung, 2010; Stokes et al., 2006). These findings have important implications, suggesting that NWR is not a potential indicator of DLD in Cantonese-speaking children, unlike what has been proposed for children speaking other languages, which also indicate that current findings on the potential of NWR to be an informative assessment tool for DLD may not be generalizable across languages. This chapter (Chapter II) presents a study re-examining whether NWR can capture significant TD/DLD group differences among Cantonese-speaking children, or in other words, whether Cantonese is a true cross-linguistic exception in NWR.

2.1.1 Lexical Phonological Properties of Cantonese

In considering the possible reasons behind the discrepant findings on Cantonese NWR, it is important to understand the unique lexical phonological properties of Cantonese, in relation to other previously studied languages. Cantonese morphemes are commonly monosyllabic (Bauer & Benedict, 1997), with each syllable taking relatively simple forms of either (C)V(V) or (C)V(C) structure, and consonant clusters are not permitted. Cantonese is also a tonal language, with each syllable being marked by one of six contrastive lexical tones. Furthermore, Cantonese has been proposed as a syllable-timed language, which does not have the variable stress patterns that occur in stress-timed languages like English (Mok, 2009). Together, these characteristics mean that the phonotactic constraints are rather simple at segmental and suprasegmental levels in Cantonese.

Accordingly, Stokes et al. (2006) proposed that NWR may not be able to differentiate between Cantonese-speaking children with TD and DLD, due to Cantonese-based nonword stimuli lacking complex phonotactic structures and/or variable stress patterns, rendering them less taxing on working memory than nonwords created with reference to other examined languages, thus even young children with DLD had little difficulty coping with the demands of the task. However, more recent

studies of NWR in Mandarin (which is typologically similar to Cantonese) and Vietnamese, which both have relatively simple phonotactic structures and are described as syllable-timed languages, have found good differentiation between TD and DLD groups (Chi, 2007; Pham & Ebert, 2020; Wang & Huang, 2016), suggesting that other reasons may offer better explanations to Stokes et al.'s (2006) discrepant findings on Cantonese.

2.1.2 Nonword Characteristics of Previously Examined Cantonese NWR Stimuli

If the simple phonotactic properties of Cantonese cannot explain the previous lack of significant findings, other aspects of the design of nonword stimuli adopted in Stokes et al. (2006) may be a possible culprit of the null findings, as extensive evidence from cross-linguistic studies suggests that NWR accuracy and TD/DLD group differentiation could be affected by characteristics of the nonword stimuli, as discussed in Chapter I.

To recap, nonword-related factors affecting NWR performance and TD/DLD group differentiation include: 1) Nonword length, where NWR accuracy declines with increasing nonword length, with some studies reporting that children with DLD are disproportionally affected by increasing nonword length, while others reporting no interaction between participant group (TD/DLD) and nonword length; 2) Nonword lexicality, where better NWR performance and greater TD/DLD group differences were reported on nonwords with higher lexicality (in terms of incorporating real morphemes within nonwords); 3) Nonword sub-lexicality, where higher NWR accuracy was seen on nonwords with higher phonotactic probability and attested CV combinations (compared to unattested CV combinations), with non-significant findings suggesting greater TD/DLD group differences to be captured by attested than unattested CV combinations; 4) Segmental Complexity, where lower NWR scores have been reported on nonwords with more complex segments, such as those containing consonant clusters and a higher number of consonants, and an increase in segmental complexity is also disproportionally challenging for children with DLD, compared to their TD peers; and 5) Prosodic features, where unstressed syllables were found to be more difficult for children to repeat than stressed syllables – as Cantonese is a syllable-timed language, where all syllables are equally stressed, this characteristic is not applicable to Cantonese nonword stimuli.

2.1.2.1 Characteristics of Nonword Stimuli used in Stokes et al. (2006). Several aspects of nonword characteristics of the stimulus set used by Stokes et al. (2006) may have contributed to the pattern of findings yielded in their study. Most notably, Stokes et al. (2006) intentionally avoided the use of real morphemes across all of their nonwords to reduce the influence of prior lexical knowledge, on the grounds that NWR tasks were proposed as strict measures of working memory capacity at the time when the study was conducted. Therefore, the stimulus set used by Stokes et al. (2006) had low levels of lexicality, which has later been found in other studies to be disadvantageous for capturing significant TD/DLD group differences in NWR.

The authors did, however, manipulate a sub-lexical factor within their stimuli – consonant-vowel (CV) combination attestedness – where two types of syllables, labelled as IN and OUT, were compared. IN syllables contained CV combinations that are attested in Cantonese, but do not necessarily constitute real morphemes (e.g., the IN syllable, /tɛː/, is attested in Cantonese because it can combine with a final consonant /-ŋ/ to form a real word /tɛːŋ/, meaning 'to listen', although /tɛː/ itself is not a real morpheme), while OUT syllables contained CV combinations that never occur in Cantonese (e.g., /ŋuːt/, where neither the syllable itself nor the CV combination /ŋuː/ occur in Cantonese). Although better NWR performance was seen on IN syllables than OUT syllables across both TD and DLD groups, which was consistent with expectations, the study reported non-significant findings on larger TD/DLD group differences being captured by IN syllables, compared to OUT syllables. This was suggested to be potentially related to the small sample size of the study.

Also interestingly, nonwords used in this study consisted of both syllables with CV structure and CVC structure, meaning that segmental complexity was not controlled in this stimulus set. As later NWR studies on other languages revealed that segmental complexity does affect children's NWR performance, it is likely that varying the structure of nonword segments, specifically syllables in this case, may have resulted in confounding effects with other factors, such as CV combination-attestedness.

Concerning nonword length, this set of nonwords ranged from two to four syllables in length, which is typical compared to stimulus sets used in other NWR studies.

2.1.2.2 Characteristics of Nonword Stimuli used in Leung (2010). In an unpublished undergraduate thesis, Leung (2010) used examined NWR performance in

Cantonese-speaking children with and without DLD, using nonword stimuli from The Test of Nonword Repetition with Pseudomorphemes from the Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS; T'sou et al., 2006). In this NWR test, nonwords were created through linking real monosyllabic Cantonese morphemes that have no meaning when combined. An English equivalent example would be linking together three words, *cup*, *knee*, and *pin*, to form a nonword /kʌpniːpɪn/. Such approach creates nonwords with high levels of lexicality, given that each nonword consists of only morphemic elements and no non-morphemic elements at all. This also naturally results in the nonwords having high levels of sub-lexicality, in terms of CV combination attestedness, as CV combinations within real morphemes are by definition, attested in the language.

Despite this set of nonword stimuli having higher levels of lexicality and sub-lexicality than the stimuli used by Stokes et al. (2006), which have been suggested to be particularly important for allowing NWR tests to capture significant TD/DLD group differences in the cross-linguistic literature, this study also reported no significant differences in NWR performance between Cantonese-speaking children with and without DLD. A possible reason for the lack of significant findings in this study may be related to the length of nonwords used in this study, together with the scoring approach used. Unlike most NWR studies, this set of nonwords ranged from one to nine syllables in length, when some studies suggest that even four- and five-syllable nonwords could result in floor effects in pre-school aged children (Guiberson & Rodríguez, 2013). Although children included in this study were slightly older (with a mean age of 5;5 for the TD group and 6;0 for the DLD group), nonwords of six to nine syllables in length are still extraordinarily long for these young children to repeat.

Furthermore, this NWR task is designed to be scored in accuracy at syllable level (i.e. children score one point for each correctly repeated syllable). Considering that there was only one item at each nonword length, the majority of syllables in the test would be concentrated on the unusually long nonword items. Therefore, the design of this NWR task deviates greatly from most other NWR task examined, and it can be argued that it is inappropriate to compare findings generated from this NWR task to those from other NWR tasks.

Similarly to Stokes et al. (2006), segmental complexity is also not controlled in this NWR task, with CV and CVC syllables being mixed together within nonwords, resulting in possible confounding effects with other factors such as nonword length.

2.1.3 The Present Study

Taken together, although previous studies on NWR in Cantonese-speaking children seem to consistently suggest that NWR could not capture group differences between TD children and those with DLD, this study argues that the previous lack of significant findings may be related to some potential flaws in the design of NWR stimuli adopted in these studies. These include low levels of lexicality in stimuli from Stokes et al. (2006), overly long nonword length in stimuli used by Leung (2010), and possible confounding effects between syllable complexity and other factors of interest in both studies.

Despite the growing understanding of how nonword characteristics may affect NWR performance and TD/DLD group differentiation, to my knowledge, no study to date has examined the use of any alternative nonword stimuli to those used by Stokes et al. (2006) and Leung (2010), for capturing TD/DLD group differences in Cantonese-speaking children. To address whether Cantonese is a true cross-linguistic exception in NWR, this study reports on a newly designed set of nonword stimuli, which takes into account each of the factors known to affect NWR performance and group differentiation from the cross-linguistic findings, as highlighted in Chapter I of this thesis.

With this novel stimulus set, the objectives of this study are: 1) to revisit whether NWR is able to generate TD/DLD group differences in Cantonese-speaking children; and 2) to identify whether nonwords at particular levels of lexicality, sub-lexicality, length, and syllable complexity are best suited to capturing TD/DLD group differences in Cantonese.

2.1.3.1 Research Questions. Specifically, four research questions (RQ) are addressed:

RQ1: How does lexicality of NWR stimuli, in terms of morphemicity, affect NWR performance and TD/DLD group differentiation in Cantonese-speaking children?

RQ2: As an extension of Stokes et al.'s (2006) analysis of IN and OUT syllables, how does sub-lexicality, in terms of CV combination attestedness, affect NWR performance and TD/DLD group differentiation in Cantonese-speaking children?

RQ3: How does nonword length, in number of syllables, affect NWR performance and TD/DLD group differentiation in Cantonese-speaking children?

RQ4: How does syllable complexity in nonwords, in terms of CV vs. CVC syllable structure, affect NWR performance and TD/DLD group differentiation in Cantonese-speaking children?

2.1.3.2 A Note on Participant Age. Unlike most previous studies looking at younger children (Schwob et al., 2021), this study examine older, eight to eleven year-old children, as younger children could not be accessed during the pandemic due to reluctance from parents to enroll their children in research studies. Given that TD/DLD group differences in NWR have been reported on children and adolescents across different ages, at least up to age 15;4 (Riches et al., 2011; Schwob et al., 2021), examining children from this older age range should still allow the research objectives to be appropriately addressed.

2.1.3.3 Significance of the Study. Findings on these research questions would have important implications on the topic of whether NWR could be a potential cross-linguistic clinical marker of DLD. This study would also establish important groundwork for future studies on the utilisation of NWR for clinical identification of Cantonese-speaking children with DLD. Finally, the findings from this study can add to the understanding of optimal NWR stimuli design, from a language with very different lexical-phonological properties from most languages studied.

2.2 Methods

2.2.1 Participants

Thirty-two predominantly monolingual Cantonese-speaking children from Hong Kong participated. They were either recruited online or invited to take part in this study after participating in other projects. These children are described as "predominantly monolingual", as they acquire Cantonese (the majority community language of Hong Kong) as first language at home and attended local schools where Cantonese was the medium of instruction, whilst also being exposed to English and Mandarin in second language classes at school, given the language education policies in Hong Kong. Unlike other bilingual children, e.g. heritage speakers of Cantonese living in an English-speaking country or children attending international schools in Hong Kong, the children included in this study are only exposed to their second languages, English and Mandarin, for less than 20% of their awake time. Therefore, they do not have extensive and

intensive exposure to languages other than Cantonese, and are described as predominantly monolingual, rather than bi-/ multi-lingual, following common operational definitions of mono-/bi-lingualism in other studies in terms of relative exposure to languages (see e.g., Paradis, 2023).

2.2.1.1 Participant Selection Criteria. Sixteen children (eleven male), aged 8;1 to 11;0 (M = 9;7, SD = 0;11), met the criteria for a DLD diagnosis following the CATALISE criteria (Bishop et al., 2017). These children demonstrated poor language skills in Cantonese – 14 children scored at 1.25 SD below age means in two or more out of six subtests of the standardized norm-referenced Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS; T'sou et al., 2006), and two scored at -1.25 SD in one subtest and -1.0 SD in another subtest. Parents and/or school personnel expressed concerns providing evidence of negative functional impact of their language difficulties, affecting daily social interactions or educational progress. Two children had co-occurring attention deficit hyperactivity disorder and dyslexia respectively.

The remaining sixteen TD children, aged 8;0 to 11;3 (M = 9;7, SD = 1;0), were individually matched to each child in the DLD group in age (within four months of age difference on the day of testing), gender, and grade in school. Parents reported no concerns over language or other aspects of development in the questionnaire. All these children were confirmed to have age-appropriate language skills under HKCOLAS.

No participants reported having any hearing impairments, and their hearing status was ascertained from passing a pure tone audiometry hearing screening test. All children also completed Raven's Progressive Matrices (Raven et al., 1996) and had standard scores above 70, screening out the likelihood of intellectual disability. The standard scores of the DLD group ranged from 82 to 127 (M = 102.7, SD = 12.5), significantly below standard scores of the TD group, which ranged from 99 to >135 (M = 113.6, SD = 11.4), $t(30) = 2.27$, $p = .02$. None of the participants were suspected to have autism spectrum disorder (ASD) by their parents and school personnel, and none have ever undergone assessments for or received a diagnosis of ASD.

2.2.1.2 Ethics Statement. This study was carried out in accordance with the recommendations of the Human Subjects Ethics Sub-committee at the Hong Kong Polytechnic University (reference number: HSEARS20161230004). Written informed consent was given by the parents of each participant.

2.2.2 Materials

2.2.2.1 Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS). HKCOLAS (T'sou et al., 2006) is a norm-referenced language assessment tool designed to examine Cantonese oral language abilities of five- to twelve- year-old children in Hong Kong. HKCOLAS has six subtests, targeting vocabulary (Lexical-Semantic Relations Test, Word Definition Test, and Expressive Nominal Vocabulary Test), morpho-syntax (Test of Hong Kong Cantonese Grammar), and narratives (Textual Comprehension Test and Narrative Test). Children who score 1.25 SDs below age means in two or more subtests qualify for a diagnosis of language disorder; at this diagnostic cut-off, HKCOLAS has a sensitivity of 0.95 and specificity of 0.98 respectively. HKCOLAS also has high test reliabilities based on coefficient alpha (0.80-0.97 across all subtests) and standardised error of measurement.

2.2.2.2 Pure Tone Audiometry Hearing Screening Test. Pure tone audiometry hearing screening test was performed using an Interacoustics AD226 diagnostic audiometer. Children are asked to raise their hands when they hear a beep (i.e. pure tones), which are presented at 25 dB hearing levels (HL) at frequencies of 500 Hz, 1000 Hz, 2000 Hz and 4000 Hz. To pass the hearing screening, children have to respond to pure tones at all test frequencies at 25dB HL in both left and right ears.

2.2.2.3 Raven's Progressive Matrices. A Hong Kong Chinese adapted version of Raven's Progressive Matrices (Chan, 1984; Raven et al., 1996) was used as a measure of non-verbal intelligence quotient, to screen out the possibility of intellectual disability when assessing children suspected of DLD. Raven's Progressive Matrices include 60 multiple choice questions, where examinees identify a missing piece from six to eight options that completes a pattern. Children are considered to be within the normal range if they gain standard scores of 70 or above.

2.2.2.4 NWR Stimuli. Three sets of NWR stimuli, which varied in lexicality levels in relation to Cantonese, were used (See Table 1 for a comparison of the nonword sets and examples, and Appendix for the full list of nonword items).

Table 1. *Comparison of nonword sets with examples (morphemic syllables in examples are marked with *)*

Nonwords	Lexicality	Sub-lexicality	Vowel Range	Examples
High-Lexicality	**High**: 100% syllables are morphemic in tone 1, but meaningless when combined	Not examined (100% IN by definition)	Wider	fe* ji* maa*
High-Lexicality-Vowel-Matched	**High**: 100% syllables are morphemic in tone 1, but meaningless when combined	Not examined (100% IN by definition)	More restricted, matched with Low-Lexicality	lo* fo*
Low-Lexicality	**Low**: 0% syllables are morphemic across all six contrastive lexical tones	50% IN (attested CV combinations), 50% OUT (unattested CV combinations)	More restricted, matched with High-Lexicality-Vowel-Matched	ngu fi hu
Stokes et al. (2006)	**Low**: 0% syllables are morphemic	50% IN (attested CV combinations), 50% OUT (unattested CV combinations)	N/A	nu pim
Leung (2010)	**High**: 100% syllables are morphemic in tone 1, but meaningless when combined	Not examined (100% IN by definition)	N/A	lei* ding* pai*

2.2.2.4.1 High-Lexicality Nonwords. High-Lexicality nonwords had the highest lexicality level, where all constituent syllables are morphemic in Cantonese, but are meaningless when combined. The items ranged from two to five syllables in length. Syllable complexity was also manipulated, in terms of rime structure (rime refers to the sequence of all phonemes following the onset, i.e. the initial consonant, within a syllable) – half of the High-Lexicality items were constructed solely with relatively simple CV syllables (i.e. rime structure being V), while the other half were constructed solely with relatively complex CVC syllables (i.e. rime structure being VC). Consonant clusters and diphthongs were not included as candidates for complex syllable structures, because the former do not occur in Cantonese (Matthews & Yip, 2011) and the latter are typically acquired early, around the same time as monophthongs (To et al., 2013).

2.2.2.4.2 Low-Lexicality Nonwords. Low-Lexicality nonwords had the lowest lexicality level, where all constituent syllables are non-morphemic across all the six contrastive lexical tones. Low-Lexicality nonwords are similar to those used in Stokes et al. (2006), in that all syllables are non-morphemic, although the syllable selection criteria are more relaxed in Stokes et al. (2006), where syllables were only non-morphemic in the tones they were presented in; in contrast, Low-Lexicality nonwords in this study consist of syllables that are non-morphemic across all six contrastive lexical tones. Due to such stringent syllable selection criteria, Low-Lexicality nonwords had a smaller vowel range than High-Lexicality nonwords. Like High-Lexicality nonwords, Low-Lexicality items also ranged from two to five syllables in length, and half of the items were constructed with CV syllables, while the other half were created with CVC syllables.

2.2.2.4.3 IN vs. OUT Syllables. Within Low-Lexicality nonwords, where all syllables were non-morphemic, the constituent syllables could be further divided into two subtypes – IN and OUT – based on their sub-lexical characteristics. Following the design in Stokes et al. (2006), IN syllables were CV or CVC structures containing attested CV combinations (e.g., *hik*, where the syllable in its entirety does not occur in Cantonese, but the CV combination *hi* does occur in Cantonese in other phonological contexts, as in *hing*), while OUT syllables contain unattested CV combinations (e.g., *ngut*, where neither itself nor the CV combination *ngu* occur in Cantonese). Half of the constituent syllables within Low-Lexicality nonwords were IN syllables, the other half

OUT, allowing for comparisons to be made on NWR performance based on this sub-lexical feature of NWR stimuli.

2.2.2.4.4 High-Lexicality-Vowel-Matched Nonwords. To match the smaller vowel range of Low-Lexicality nonwords, an additional set of High-Lexicality nonwords, labelled as High-Lexicality-Vowel-Matched nonwords, was created. Like High-Lexicality nonwords, High-Lexicality-Vowel-Matched items also had constituent syllables that are morphemic in Cantonese, and therefore also had the highest lexicality level, but they matched the more restricted vowel range in Low-Lexicality nonwords. High-Lexicality-Vowel-Matched nonwords also ranged from two to five syllables in length, and half of the items were constructed with CV syllables, while the other half were created with CVC syllables.

2.2.2.4.5 Other Considerations. Consonants and vowels used across all nonword sets were expected to be acquired by age 4;0 in speech production by monolingual Cantonese-speaking children (To et al., 2013). Syllables that sounded like real English words (e.g., *wet* or *fit*) and nonwords with syllable combinations as subparts that sounded like real multi-syllabic, Cantonese words were avoided. To control for prosodic effects, all nonwords were set to be articulated with Cantonese tone one, with even length and stress on each syllable. All items were recorded by a female native Cantonese-speaking student speech and language therapist (SLT).

2.2.3 Procedures

All experimental tasks were administered by native Cantonese-speaking student SLTs, in a quiet clinic room at our Speech Therapy Unit. The testing session, consisting of a hearing screening, NWR task, standardised language assessment and non-verbal intelligence quotient test, lasted for about two hours.

The procedures of the NWR task were modelled after those from Polišenská and Kapalková (2014). The computerised NWR task was presented as a picture story through PowerPoint slides. Participants listened to pre-recorded instructions and stimuli that were embedded into the slides through noise cancelling headphones in a quiet room. In the two practice trials, children were instructed to listen to and repeat magic words (i.e., nonwords) exactly as they heard them, and a bead would appear on a thread on screen when an attempt has been made. Replays of the practice stimuli were

permitted, and feedback on accuracy was given to ensure the participant understood the task requirements. The experimental block was embedded into a story about helping story characters repair a broken necklace for their mother, by repeating nonwords exactly as they heard them. Nonwords from the three stimulus sets were pooled together, and the order of nonword presentation was randomised. With every attempt, a bead would appear on screen, until the necklace was fully repaired at the end of the experimental block. Replays were not permitted in the experimental block, unless the presentation of stimuli was interrupted by transient distractions (e.g., talking), and feedback on accuracy was not provided.

2.2.3.1 Scoring. Responses were audio recorded and transcribed. Performance on all nonwords was scored on whole-nonword correctness (i.e., responses must contain all and only the target segments in the correct order to be scored as correct). Low-Lexicality nonwords were further scored on syllable-level accuracy, to allow for analysis of NWR performance based on sub-lexical characteristics. In Hong Kong Cantonese, there are two well documented free variants that are prevalent even among adult native speakers, which are the omission of initial /ŋ/ consonant, and substitutions between final /k/ and final /t/ consonants (To et al., 2013). Therefore, responses with such variations were not regarded as incorrect. Changes in prosody (e.g. tone) were not penalised, and such changes were rarely observed.

2.2.4 Inter-rater Reliability

Five native-Cantonese speakers with linguistic training on Cantonese phonetics and phonology and phonetic transcription of normal and disordered speech samples transcribed and scored the data. One completed the first round of transcriptions and scoring for all data (both by whole-nonword and by syllable scoring). Two independently transcribed 31.3% of all data and scored NWR accuracy by whole-nonword correctness. The remaining two independently transcribed 37.5% of responses to Low-Lexicality nonwords and scored NWR accuracy at syllable level. At whole-nonword level scoring, the average measure Intra-class Coefficient (ICC) using a two-way mixed model and absolute agreement was .98 for High-Lexicality-Vowel-Matched nonwords (95% Confidence intervals or CI of .90 to .99); .90 for High-Lexicality nonwords (95% CI of .63 to .98); and .93 for Low-Lexicality nonwords (95% CI of .70 to .98). At syllable level scoring, the average ICC using a two-way mixed model and

absolute agreement was .89 (95% CI of .88 to .91), indicating good to excellent levels of reliability between raters at both levels of scoring.

2.2.5 Data Analysis

NWR scores were analysed with Logistics Mixed Effects Models, using the R package lme4 (Bates & Maechler, 2010) in R (version 4.1.3, R Core Development Team, 2021). Four models were used to address each of the four RQs respectively.

RQ1, concerning the effects of lexicality of NWR stimuli on NWR performance and TD/DLD group differentiation, was addressed with Model 1. Model 1 had a dependent variable of NWR accuracy at whole-nonword level (as a categorical variable of Correct vs. Incorrect for each trial), while lexicality (High-Lexicality vs. Low-Lexicality vs. High-Lexicality-Vowel-Matched), participant group (TD vs. DLD) and their interaction were added to the model as independent variables. Participants and nonword items were added as random effects.

RQ2, concerning the effects of CV combination attestedness on NWR performance and TD/DLD group differentiation, was addressed with Model 2. In Model 2, NWR accuracy at syllable level was the dependent variable (measured categorically as Correct vs. Incorrect for each trial), with independent variables of CV combination attestedness (IN vs. OUT), participant group (TD vs. DLD), and their interaction. The dependent variable in Model 2 was NWR accuracy at syllable level, as opposed to whole-nonword level, because IN syllables and OUT syllables co-occurred within items, i.e. each nonword consisted of both IN and OUT syllables, so nonwords had to be scored at syllable level for the effects of CV combination attestedness to be examined. Participant and nonword items were added as random effects.

RQ3, concerning the effects of nonword length on NWR performance and TD/DLD group differentiation, was addressed with Model 3. Model 3 had the dependent variable of NWR accuracy at whole-nonword level (as a categorical variable of Correct vs. Incorrect for each trial), while length (in number of syllables, as a continuous variable), participant group (TD vs. DLD) and their interaction were added to the model as independent variables. Participants and nonword items were added as random effects.

RQ4, concerning the effects of syllable complexity on NWR performance and TD/DLD group differentiation, was addressed with Model 4. Model 4 had the dependent variable of NWR accuracy at whole-nonword level (as a categorical variable

of Correct vs. Incorrect for each trial), while syllable complexity (CV vs. CVC), participant group (TD vs. DLD) and their interaction were added to the model as independent variables. Participants and nonword items were added as random effects.

2.3 Results

2.3.1 RQ1: How does lexicality of NWR stimuli, in terms of morphemicity, affect NWR performance and TD/DLD group differentiation in Cantonese-speaking children?

The specification of Model 1 was as follows: NWRaccuracy_NonwordLevel ~ Lexicality + Group + Lexicality*Group + (1|participant) + (1|Nonword_Item).

The results of Model 1 are shown in Table 2. There was a significant main effect of Lexicality, with significantly lower scores on Low-Lexicality nonwords compared to High-Lexicality-Vowel-Matched nonwords (p <.001), but no significant difference between performance on High-Lexicality and High-Lexicality-Vowel-Matched nonwords, which shared the same level of lexicality. There was also a significant main effect of Group (p <.001), where the TD group scored significantly higher than the DLD group in the NWR task, at whole-nonword level scoring.

The Group x Lexicality interaction was not statistically significant, with nonwords at each lexicality level capturing significant TD/DLD group differences in NWR performance, although effect sizes in odds ratios (OR) showed that greater TD/DLD group differentiation occurred with nonwords of the higher lexicality level (i.e., High-Lexicality: OR = 5.06, medium effect size; and High-Lexicality-Vowel-matched: OR = 4.04, medium effect size), compared to Low-Lexicality nonwords (OR = 3.10, small effect size).

Table 2. *Results of Model 1 on effects of lexicality on NWR accuracy at whole-nonword level and TD/DLD group differentiation*

Fixed Effect	β	*SE*	*z*	*p*
(Intercept)	0.52	0.41	1.25	.21
Lexicality (Low-Lexicality)	-2.54	0.50	-5.05	<.001
Lexicality (High-Lexicality)	-0.01	0.49	-0.02	.98
Group (TD)	1.53	0.38	4.01	<.001
Group(TD) : Lexicality(Low-Lexicality)	-0.49	0.30	-1.65	.10
Group(TD) : Lexicality(High-Lexicality)	0.07	0.28	0.24	.81
Random Effect	Variance	*SD*		
Item	2.44	1.56		
Participant	0.85	0.92		

2.3.2 RQ2: How does sub-lexicality, in terms of CV combination attestedness, affect NWR performance and TD/DLD group differentiation in Cantonese-speaking children?

The specification of Model 2 was as follows: NWRaccuracy_SyllableLevel ~ CVcombination_attestedness + Group + CVcombination_attestedness*Group + (1|Participant) + (1|Nonword_Item).

The results of Model 2 are shown in Table 3. There was a significant main effect of CV combination attestedness, with significantly lower scores on OUT syllables (i.e. those containing unattested CV combinations) compared to IN syllables (i.e. those containing attested CV combinations; p <.001). There was also a significant main effect of Group (p <.001), where the TD group scored significantly higher than the DLD group in the NWR task, at syllable level scoring. The interaction between CV combination attestedness and group was significant (p <.001).

To assist the interpretation of this interaction, we used likelihood ratio tests to examine the effects of group on NWR accuracy at syllable level in IN syllables and OUT syllables separately. When IN syllables were analysed, the fixed effect of Group was highly significant (p <.001) with a medium effect size (OR = 3.08), and when OUT syllables were analysed, the fixed effect of Group was significant (p = .04), although with a small effect size (OR = 1.65). We also plotted predicted probabilities of NWR accuracy (with 95% confidence intervals) for children on IN and OUT syllables, for the DLD and TD groups separately (see Figure 2). Figure 2 shows that in children with

DLD, the gap between performance on IN syllables and OUT syllables was small, when the 95% confidence intervals (CI) were taken into consideration. On the other hand, in TD children, the gap was large, indicating a more consistent and prominent improvement in performance on IN syllables compared to OUT syllables, even when the 95% CI were factored in.

Table 3. *Results of Model 2 on effects of sub-lexicality on NWR accuracy at syllable level and TD/DLD group differentiation*

Fixed Effect	β	*SE*	*z*	*p*
(Intercept)	1.18	0.25	4.76	<.001
CVcombination_Attestedness(OUT)	-1.03	0.13	-7.98	<.001
Group (TD)	1.11	0.25	4.56	<.001
Group(TD) : CVcombination_Attestedness(OUT)	-0.66	0.19	-3.46	<.001
Random Effect	Variance	*SD*		
Item	0.32	0.56		
Participant	0.78	0.88		

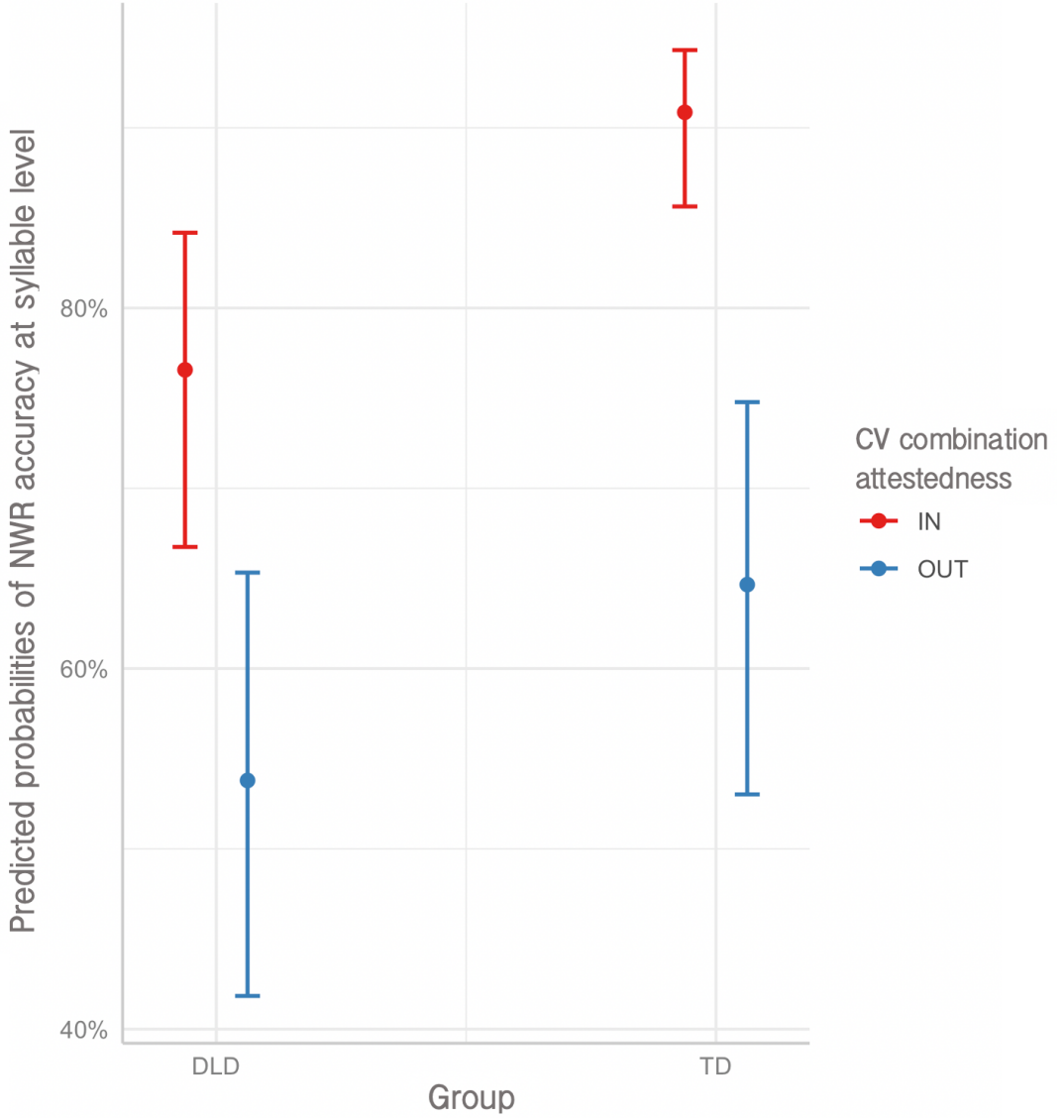

Figure 2. *Predicted probability of NWR accuracy at syllable level (with 95% confidence intervals) on IN and OUT syllables within Low-Lexicality nonwords, in the DLD vs. TD groups.*

2.3.3 RQ3: How does nonword length, in number of syllables, affect NWR performance and TD/DLD group differentiation in Cantonese-speaking children?

The specification of Model 3 was as follows: NWRaccuracy_NonwordLevel ~ Length + Group + Length*Group + (1|Participant) + (1|Nonword_Item).

The results of Model 3 are shown in Table 4. There was a significant main effect of Length, where nonword accuracy decreased as nonwords increased in number of syllables (p <.001), but the model did not register a significant main effect of Group (p

=.12)[1]. The model also did not register a significant interaction between Length and Group.

Table 4. *Results of Model 3 on effects of nonword length on NWR accuracy at whole-nonword level and TD/DLD group differentiation*

Fixed Effect	β	*SE*	*z*	*p*
(Intercept)	3.65	0.74	4.93	<.001
Length	-1.14	0.19	-5.91	<.001
Group (TD)	0.83	0.53	1.55	.12
Group(TD) : Length	0.16	0.11	1.43	.15
Random Effect	Variance	*SD*		
Item	2.76	1.66		
Participant	0.85	0.92		

2.3.4 RQ4: How does syllable complexity in nonwords, in terms of CV vs. CVC syllable structure, affect NWR performance and TD/DLD group differentiation in Cantonese-speaking children?

The specification of Model 4 was as follows: NWRaccuracy_NonwordLevel ~ Syllable Complexity + Group + Syllable Complexity*Group + (1|Participant) + (1|Nonword_Item).

The results of Model 4 are shown in Table 5. The model did not register a significant main effect of Syllable Complexity (p =.08), but the main effect of Group was significant (p <.001), indicating that the TD group scored significantly higher than the DLD group. The model also did not register a significant interaction between Syllable Complexity and Group.

[1] While this maximal model did not register a significant main effect of Group, the effect of Group was significant (TD > DLD) when an alternative stepwise forward selection analysis approach was used, where two models, one including and one excluding a factor (i.e. Group in this case), are compared using likelihood ratio tests to identify significant effects when building up a model step-by-step.

Table 5. *Results of Model 4 on effects of syllable complexity on NWR accuracy at whole-nonword level and TD/DLD group differentiation*

Fixed Effect	β	*SE*	*z*	*p*
(Intercept)	0.09	0.43	0.21	.84
Syllable Complexity(CVC)	-0.90	0.51	-1.78	.08
Group (TD)	1.24	0.37	3.39	<.001
Group(TD) : Syllable Complexity(CVC)	0.37	0.24	1.55	.12
Random Effect	Variance	*SD*		
Item	4.13	2.03		
Participant	0.85	0.92		

2.4 Discussion

Study I revisited whether NWR is able to capture TD/DLD group differences in predominantly monolingual Cantonese-speaking children, and whether nonwords with certain characteristics are better able to capture such group differences than others. Specifically, Study I addressed four RQs, which asked whether NWR performance and TD/DLD group differentiation in Cantonese-speaking children are affected by: 1) lexicality, in terms of morphemicity; 2) sub-lexicality, in terms of CV combination attestedness; 3) nonword length, in number of syllables; and 4) syllable complexity, by comparing syllables of CV and CVC structure.

This study is the first to document that Cantonese-speaking TD children do perform significantly above their DLD peers in NWR, and our findings suggest that nonwords with higher levels of lexicality and sub-lexicality maximise the degree of TD/DLD group differentiation.

2.4.1 Potential Utility of NWR in Identifying DLD in Cantonese-Speaking Children

The present findings suggest that Cantonese is not a true cross-linguistic exception in NWR, despite previous suggestions that its simple phonotactic structure may mean that Cantonese nonword stimuli are less taxing on working memory, and therefore unable to generate TD/DLD group differences in children acquiring Cantonese. Using the newly designed NWR stimulus set, TD/DLD group differences were captured in all statistical models, except in Model 3, which focused on the factor, nonword length; given that significant group differences were otherwise consistently registered, there is

a possibility that the exceptional findings in Model 3 may be a Type II error, stemming from the relatively small sample size in this study.

Moreover, significant TD/DLD group differences were even found in NWR performance on the Low-Lexicality items, which were similar to the stimuli used by Stokes et al. (2006), who previously reported non-significant findings. The present results, therefore, point to NWR having a potential to be further developed as an assessment tool to aid the identification of DLD in Cantonese-speaking children. This will certainly require future work to replicate the present findings on a larger sample size, as well as to examine the diagnostic accuracy of this NWR test for Cantonese-speaking children. Despite a relatively small sample size in the current study that is comparable to that of the widely cited study by Stokes et al. (2006; 14 DLDs and 15 age-matched TDs), the present findings have established that Cantonese is not a cross-linguistic exception in NWR, providing further support for the utility of NWR in identifying DLD alongside other forms of assessment in Cantonese, as well as cross-linguistically.

2.4.2 Considerations of Nonword Factors in Designing NWR Tests for TD/DLD Group Differentiation

The present findings also suggest that most nonword-related factors known to affect NWR performance and group differentiation in children acquiring other languages affect Cantonese-speaking children in similar ways. Therefore, the utility of NWR tasks for generating TD/DLD group differences depends on careful consideration of these nonword-related factors, especially those that affect TD and DLD groups differently. In the following, we will discuss the findings on each of the nonword-related factors examined.

2.4.2.1 Lexicality. Significant findings were found on lexicality, where Cantonese-speaking children, regardless of TD or DLD status, repeated nonwords more accurately in High-Lexicality and High-Lexicality-Vowel-Matched items (where all constituent syllables are morphemic in Cantonese), compared to Low-Lexicality items (where all constituent syllables are non-morphemic in Cantonese). These findings imply that children do draw on their long-term lexical-phonological knowledge at a morphemic level to support the repetition of nonwords, when the nonwords allowed them to do so.

Furthermore, the degree of TD/DLD group differentiation differed depending on the lexicality level of nonwords, as indexed by differences in effect sizes – larger differences were captured by High-Lexicality and High-Lexicality-Vowel-Matched nonwords, compared to Low-Lexicality nonwords. A possible reason behind this pattern of findings is that since TD children are expected to have stronger lexical representations (and sub-lexical representations, see the next section on Sub-lexicality), TD children can rely on their stronger lexical (and sub-lexical) knowledge for greater use of redintegration strategies to support NWR, resulting in a more prominent improvement in NWR accuracy when nonwords increased in lexicality levels, relative to children with DLD. Combined with the lack of significant findings previously reported by Stokes et al. (2006), NWR data on Cantonese-speaking children suggest that nonwords with low lexicality levels are not optimal for detecting TD/DLD group differences, and that lexical factors must be carefully considered in the design of NWR stimuli, to allow for children to draw on their lexical-phonological knowledge to support the repetition of nonwords.

2.4.2.2 Sub-lexicality. The data suggest that even within Low-Lexicality items, where all constituent syllables are non-morphemic, children achieve better performance on syllables that contain CV combinations attested in Cantonese (i.e., IN syllables) than those that do not (i.e., OUT syllables). Furthermore, although both IN and OUT syllables generated TD/DLD group differences, greater differentiation was observed on IN syllables than OUT syllables, as shown by effect size differences, and the plot (Figure 2) also indicated that children with DLD showed less benefit on IN syllables than TD children. This pattern of findings mirrored those on lexicality, suggesting that sub-lexicality in nonwords affected NWR performance and TD/DLD group differentiation in an analogous manner to lexicality, and that sub-lexicality factors must also be carefully considered in designing nonwords for the purpose of generating TD/DLD group differences. This is consistent with recent findings on other languages that children draw from multiple streams – both lexical and sub-lexical representations – to support NWR, and particularly that sub-lexical representations may be fundamental to successful NWR (Szewczyk et al., 2018).

2.4.2.3 Length. A significant effect of length was also found on NWR accuracy, which is consistent with most, if not all previous studies, in finding a decrease in NWR

accuracy as nonwords increase in length and become gradually more taxing on working memory (Schwob et al., 2021). The present data did not register an interaction between group and length, suggesting that these children with DLD were not disproportionally affected by increasing nonword length.

2.4.2.4 Syllable complexity. The data did not register a significant effect of syllable complexity, nor an interaction between group and syllable complexity. While this may seem inconsistent with previous studies that have reported lower NWR accuracy on nonwords with higher segmental complexity (e.g. nonwords containing consonant clusters; Jones et al., 2010), this is unsurprising for two reasons. First, as our participant sample was older than previously examined groups, an increase in syllable complexity from CV to CVC structure may not pose significantly more challenge to our participants, including those with DLD. Second, as consonant clusters are not permitted in Cantonese, even nonwords with relatively complex (specifically CVC) syllables are simpler than segmentally complex nonwords examined in other languages previously.

2.4.3 Limitations and Future Directions

One limitation of the present study is its relatively small sample size, which meant that there is a possibility for statistical models to miss effects that would emerge as significant in a larger sample. On the other hand, significant TD/DLD group differences were still registered in the current study, demonstrating the potential clinical utility of NWR for identifying DLD in Cantonese-speaking children. The present findings would need to be replicated in a larger sample, which would allow for the diagnostic accuracy of this NWR test to be examined at an individual, rather than group level, in terms of sensitivity and specificity.

Moreover, future work is needed to confirm whether significant group differences in NWR accuracy between Cantonese-speaking children with and without DLD carries over to younger children, as the lack of positive findings in Stokes et al. (2006) and Leung (2010) was reported on a younger participant group at pre-school age, and the development of clinical screening tools should strive to allow for early identification of developmental disorders.

The current findings also have important clinical implications for further research on using NWR measures to improve the identification of DLD in Cantonese-speaking

bilingual children, an area largely lacking in assessment tools.

2.5 Conclusion

The results of Study I suggest that, by improving the design of Cantonese NWR stimuli, NWR is able to capture TD/DLD group differences in Cantonese-speaking children, providing novel findings that are consistent with the cross-linguistic literature and confirm that NWR tasks have the potential to be used as informative assessment tools for DLD in Cantonese and cross-linguistically. Both lexical and sub-lexical factors must be considered in the design of NWR stimuli for generating TD/DLD group differences, as children draw on lexical and sub-lexical representations to support NWR.

2.6 Summary of Chapter II

While NWR has been described as a potential cross-linguistic clinical maker of DLD, Cantonese has been the only reported exception, where NWR does not differentiate between TD/DLD groups (Leung, 2010; Stokes et al. 2006). Study I argues that the previous lack of significant findings may be related to certain aspects of the stimuli design in previous studies, and reports on a novel set of NWR stimuli which take into account factors known to affect NWR performance and group differentiation. Sixteen TD-DLD pairs of Cantonese-speaking children repeated two sets of High-Lexicality nonwords, and one set of Low-lexicality nonwords, which could be further analysed on sub-lexicality, based on CV combination attestedness. Children with DLD scored significantly below their TD peers, and effect sizes showed that nonwords with high lexicality and sub-lexicality (i.e., attested CV combinations) offered the greater TD/DLD group differentiation, suggesting that Cantonese is not a true cross-linguistic exception in NWR. Future work could aim to replicate the present findings on a larger sample size, verify whether TD/DLD group differences are still captured by NWR in younger, Cantonese-speaking children, and examine the diagnostic accuracy of this NWR test.

CHAPTER III – Study II
Disentangling Effects of Language Impairment and Bilingualism with Cantonese Nonword Repetition Stimuli

3.1 Introduction

The task of identifying DLD in bilingual children has been described as a major challenge (Armon-Lotem, 2012; Schwob et al., 2021), due to similarities in language limitations and errors observed in children with transient difficulties that arise from insufficient exposure to a language (in bilingual children), and children experiencing more persistent language difficulties from a language disorder (in children, monolingual or bilingual, with DLD; Camilleri & Law, 2007), where language difficulties are associated with poor prognosis and unlikely to resolve on its own, and require specialist support. NWR has been advocated as an important tool within test batteries for assessing DLD in bilingual children, as children with DLD consistently perform below TD children in NWR, even in a bilingual context (Schwob et al., 2021).

Given the promising findings on novel Cantonese nonword stimuli reported in Study I of this thesis (see Chapter II), that TD/DLD group differences can be captured among monolingual Cantonese-speaking children, this chapter presents a study that extends research on Cantonese NWR to bilingual Cantonese-speaking children. Specifically, Study II will address whether bilingual TD children acquiring Cantonese as a second language (L2) will be disadvantaged by Cantonese NWR stimuli.

3.1.1 NWR in Bilingual Children

Cross-linguistically, NWR is able to discriminate between monolingual children with and without DLD, because children with DLD often have deficits in some or all of the skills required to support the accurate repetition of nonwords, including impaired phonological working memory capacity (Gathercole & Baddeley, 1990; Montgomery, 2002), vocabulary skills (Gray et al., 1999; McGregor, 2009; Watkins et al., 1995), and motor planning deficits (Stark & Blackwell, 1997), which meant that the demands of NWR are disproportionally challenging for children with DLD.

In the case of bilingual children without a language disorder (i.e. bilingual TD children), NWR performance is affected in similar ways to both monolingual TD (MonTD) children and monolingual children with DLD (MonDLD), in that bilingual

TD children are expected to have processing capabilities (e.g. working memory capacity, motor planning) comparable to that of MonTD children, but they may also have weaker lexical and sub-lexical representations to support NWR, similar to MonDLD children, due to reduced language experience and proficiency associated with bilingual language acquisition. While NWR involves the repetition of nonsense phonemic sequences, vocabulary knowledge plays an important role in successful NWR, as children draw on their existing lexical-phonological knowledge to reconstruct a target nonword, as the memory trace decays in working memory. Therefore, bilingual TD children are at risk of being disadvantaged relative to MonTD children and misclassified into disordered groups in NWR, due to having weaker lexical and sub-lexical support associated with reduced language experience.

Empirically, some studies have reported that bilingual TD children perform just as well as MonTD children in NWR (Armon-Lotem, 2018; Armon-Lotem & Chiat, 2012; Lee & Gorman, 2013; Messer et al., 2010), while others reported that bilingual TD children perform below MonTD children (Boerma et al., 2015; Kohnert et al., 2006; Sharp & Gathercole, 2013). Importantly, bilingual TD groups have consistently been reported to perform above MonDLD groups (Boerma et al., 2015; Thordardottir & Brandeker, 2013), as well as bilingual DLD groups (Schwob et al., 2021).

3.1.2 Language-Specific vs. Quasi-Universal NWR Tests

Whether bilingual TD children are disadvantaged in NWR, in comparison to MonTD children, appears to be partly related to the characteristics of the nonword stimuli used. A meta-analysis (Schwob et al., 2021) reported that differences in NWR performance between monolingual and bilingual TD groups tend to be minimized in studies that used quasi-universal NWR tests, rather than language-specific NWR tests. Unlike traditional NWR tests that are created with reference to phonotactic constraints of a given language (i.e. language-specific NWR tests), quasi-universal NWR tests are designed to be maximally compatible with the lexical phonology of diverse languages, hence being "quasi-universal". One such example is the crosslinguistic NWR task (CL-NWR; Chiat, 2015), where nonword stimuli (e.g. *bamudi*) were designed to contain only crosslinguistically frequent consonants and vowels, and are set to be articulated with neutral prosody, through applying even length, stress, and pitch on all syllables equally. By removing elements of nonwords that are relatively uncommon across languages and may only be familiar to speakers of certain languages (e.g. lexical tones

in Mandarin and Cantonese, consonant clusters in English, Polish and French), nonword stimuli in CL-NWR are able to reduce the likelihood of disadvantaging bilingual children with reduced language experience in a given language, therefore minimising the gap between monolingual and bilingual TD groups.

3.1.3 The Crosslinguistic Nonword Repetition Test (CL-NWR)

The CL-NWR test was evaluated by Boerma et al. (2015) in its ability to differentiate TD and DLD children in both monolingual and bilingual groups. The study compared four groups of children – monolingual Dutch-speaking TD children, monolingual Dutch-speaking children with DLD, bilingual Dutch-speaking TD children, and bilingual Dutch-speaking children with DLD (all bilingual children were second language (L2) learners of Dutch with different first languages (L1s)) – on their NWR performance in the Dutch version[2] of the CL-NWR task versus a Dutch language-specific NWR task. Results indicated that children with DLD, regardless of monolingual or bilingual status, performed worse than TD children on both the language-specific and quasi-universal NWR tasks. Bilingual TD children performed worse than monolingual TD children in the language-specific NWR task, but not the quasi-universal CL-NWR task; and monolingual and bilingual children with DLD performed at a similar level in both language-specific and quasi-universal NWR tasks, suggesting that quasi-universal NWR tasks such as the CL-NWR test is able to reduce disadvantage for bilingual children, despite their reduced language experience.

Aside from group comparisons, the study also examined the diagnostic accuracy of the language-specific NWR task and the Dutch CL-NWR task, which found that both tasks had adequate levels of sensitivity and specificity in the monolingual sample, but the language-specific task fell short in classifying L2 learners of Dutch with and without DLD. These results suggest that the quasi-universal CL-NWR test may have

---

[2] Even though the CL-NWR test is designed to be cross-linguistic in nature, the test still requires adaptations to different language versions, to 1) ensure that all consonants and vowels in the test are present in the phonetic inventory of a given test language (e.g., Dutch in this case) and familiar to children acquiring that language; and 2) avoid the possibility of real words or inflections in a given test language being included in the nonword stimuli. The adaptation process involves selecting one out of four to six options for each of the 16 nonword items included in the test.

higher clinical accuracy than language-specific NWR tasks when used in a bilingual population, providing support that a quasi-universal NWR task may be better able at disentangling the effects of language impairment and bilingualism, making CL-NWR a promising tool for identifying DLD in bilingual children.

More recent studies on the CL-NWR task have reported less clear-cut findings. In a follow-up longitudinal study (Boerma & Blom, 2021), the same groups of children included in Boerma et al. (2015) were retested using the Dutch CL-NWR task at one-year intervals over a two-year period. At six- to seven- years (second wave of testing) and seven- to eight- years (third wave of testing), sensitivity and specificity of the Dutch CL-NWR task were found to drop to clinically unacceptable levels (<80% accuracy) for both monolingual and bilingual children, although significant and large group differences were still yielded between TD and DLD groups, while bilingual L2 children were not found to perform below their monolingual counterparts. Furthermore, diagnostic accuracy reached clinically acceptable levels when the Dutch CL-NWR task was used in combination with a narrative assessment, suggesting that while the CL-NWR task may not be suitable to be used as a standalone diagnostic tool, especially for children aged six years or older, it can be an informative measure when used within test batteries to assess monolingual and bilingual children.

In another study (Öberg & Bohnacker, 2022), NWR performance on the Swedish CL-NWR task was examined in four- to seven-year-old bilingual children acquiring L1-Arabic and L2-Swedish. Due to a large discrepancy in the sample size of the L2-TD (N = 99) and L2-DLD (N = 11) groups in this study, group means could not be compared, but z-scores (derived using the TD group as the reference population) revealed substantial overlap in performance between the bilingual-TD and bilingual-DLD groups, thus it was concluded that NWR performance on the Swedish CL-NWR task could not reliably distinguish bilingual children with DLD from bilingual TD children. Therefore, the degree of TD/DLD group differentiation and clinical accuracy of the CL-NWR task, particularly on bilingual children, may depend on the specific languages (and L1-L2 language combinations) examined. Other factors, such as criteria for recruitment and identification of the DLD sample, might also contribute to different findings. As such, it is important for the CL-NWR task to be examined in typologically diverse languages.

3.1.4 Using Cantonese NWR stimuli to assess Bilingual Cantonese-Speaking Children

Following the positive findings on monolingual Cantonese-speaking children reported in Study I, Study II further examines the potential utility of NWR (both language-specific and quasi-universal NWR tests) for identifying DLD in Cantonese-speaking children, by exploring whether NWR is still able to accurately capture TD/DLD group differences in Cantonese-speaking children, when bilingual second-language (L2) learners are also taken into account. Two particularly important issues will need to be addressed.

The first issue concerns whether Cantonese NWR stimuli will disadvantage bilingual TD children, compared to monolingual TD children. When assessing bilingual children using NWR, an ideal scenario would be to have nonwords that could maximize the gap between TD and DLD groups (i.e. between MonTD vs. MonDLD, L2-TD vs. L2-DLD, and particularly L2-TD vs. MonDLD groups), while also minimizing the gap between monolingual and bilingual L2-TD children, despite L2-TD children having reduced exposure to and knowledge of the target language of testing. In other words, for NWR to be developed into a clinically informative assessment tool for bilingual Cantonese-speaking children, the NWR task should demonstrate capabilities in disentangling the effects of language impairment and language experience.

The second issue concerns the influence of nonword characteristics, particularly lexicality, on the degree of TD/DLD group differentiation. While previous findings on predominantly monolingual Cantonese-speaking children suggest that high lexicality and sub-lexicality nonwords were better suited towards capturing greater TD/DLD group differences, this may not be the case when bilingual L2 learners are taken into account, as L2 learners may not be able to benefit from the increase in lexicality, due to weaker lexical and sub-lexical representations in their L2, compared to MonTD children. Therefore, nonwords that are too similar to real words in the ambient language may disadvantage, rather than support L2-TD children. Thus, it is also important to examine the influence of lexicality of nonwords on the pattern and degree of group differentiation.

3.1.5 The Present Study

This study explores the potential of Cantonese NWR stimuli to be developed into

clinically informative assessment tools for identifying DLD in bilingual Cantonese-speaking children, by examining whether bilingual L2-TD children, particularly L1-Urdu-L2-Cantonese-speaking children, will be disadvantaged against monolingual Cantonese-speaking TD children, on language-specific, Cantonese NWR stimuli (reported in Study I), as well as quasi-universal, Cantonese-adapted CL-NWR stimuli (Chiat, 2015).

An additional group of L2 Cantonese-speaking children with DLD was not included, as assessment tools for L2 Cantonese-speaking children are currently under development – these assessment tools have been demonstrated to be useful for identifying DLD in a case study (Hamdani et al., under revisions), but findings have to be replicated in larger scale studies to confirm the diagnostic potential of these new assessment tools. Hence, it is currently very difficult to identify a group of L2 Cantonese-speaking children with DLD, who also have to have comparable language experience to the L2-TD group.

The objective of this study is to first examine whether Cantonese nonword stimuli, both language-specific and quasi-universal, is able to minimize the differences between MonTD and L2-TD children, whilst maximizing the differences between MonTD and MonDLD groups, and between L2-TD and MonDLD groups.

3.1.5.1 Selection of Target Participant Group. This study will focus on a group of TD, L1-Urdu-L2-Cantonese-speaking children residing in Hong Kong, in addition to predominantly-monolingual Cantonese-speaking TD children and their peers with DLD. The choice of this particular bilingual group was influenced by three factors.

First, people of Pakistani origin constitute 23.9% of the large South Asian communities residing in Hong Kong, according to the Hong Kong 2021 Population Census. Poon (2010) noted that Urdu, rather than the local official languages of Hong Kong (i.e. Cantonese, English, and Mandarin), is the preferred language in these Pakistani families, as many parents are not proficient in Cantonese. As a result, most of these Pakistani children are bilingual learners who are using a minority language of Urdu as first language (L1) at home, while acquiring Cantonese as L2 in a school and community context, with reduced exposure to both languages.

Second, assessing these children in their L1 is challenging, given that little is known about the developmental expectations of this bilingual group, with language

assessment tools in Urdu being unavailable or inaccessible for local speech and language therapists (SLTs).

Third, these children often come from families with low socio-economic status, where parents may sometimes lack sensitivity to the possibility of language disorder when their child presents with language difficulties, and may be less likely to seek support from professionals for assessment. Therefore, relative to other bilingual groups of children in Hong Kong, such as English-Cantonese and Mandarin-Cantonese bilingual children, testing L1-Urdu-L2-Cantonese-speaking bilingual children provides a particular motivation and opportunity to examine whether our novel Cantonese NWR stimuli disadvantage bilingual L2 children likely to be socioeconomically disadvantaged and have weak language skills in the testing language (i.e. Cantonese). Moreover, there is a greater and more urgent need for suitable assessment tools to be developed for L1-Urdu-L2-Cantonese-speaking bilingual children, given the current lack of suitable assessment tools.

3.1.5.2 Secondary Objective: An Analysis of NWR Scoring Method. As a further extension to the line of work on Cantonese NWR, this study will also examine the effects of using different scoring methods in NWR on group differentiation.

Previous studies on other languages have found different levels of clinical accuracy when different scoring methods were used. The two most commonly used approaches were percentage of items correct (PIC; i.e. scoring NWR at whole-item level), and percentage of phonemes correct (PPC; i.e. scoring NWR at phoneme level). Studies have found that both methods discriminated between TD and DLD groups in both monolingual and bilingual children (Schwob et al., 2021), with some studies finding that PIC generated higher levels of diagnostic accuracy than PPC (Dispaldro et al., 2013; Guiberson & Rodríguez, 2013), and others reporting no difference between PIC and PPC, especially in bilingual populations (Boerma et al., 2015; le Clercq et al., 2017). Practically speaking, studies have also noted that PIC is faster and easier to score, making it more ideal for use in speech and language therapy clinics than PPC (Dispaldro et al., 2013; Pham & Ebert, 2020).

It is hard to conclude from these findings, whether PIC or PPC scoring in NWR is definitively better than the other, but it is clear that the scoring methods do affect the power of TD/DLD group differentiation (see Ortiz, 2021; and Schwob et al., 2021). As no studies have examined how scoring approaches in Cantonese NWR tasks affected

group differentiation between TD children and those with DLD, this study aims to also examine whether two scoring approaches – scoring Cantonese NWR at a syllable level vs. whole-item level, produces a different pattern or degree of group differentiation in Cantonese MonTD, MonDLD, and L2-TD children.

3.1.5.3 Research Questions. Taken together, this study addresses two research questions (RQ):

RQ1: When NWR accuracy is scored on whole-nonwords, can language-specific Cantonese nonwords (including High-Lexicality, Low-Lexicality, High-lexicality-Vowel-Matched nonwords), and Cross-linguistic nonwords capture significant group differences between MonTD and MonDLD children, and between L2-TD and MonDLD children, while minimising group differences between MonTD and L2-TD children?

RQ2: When NWR accuracy is scored on each nonword syllable, can language-specific Cantonese nonwords (including High-Lexicality, Low-Lexicality, High-lexicality-Vowel-Matched nonwords), and Cross-linguistic nonwords capture significant group differences between MonTD and MonDLD children, and between L2-TD and MonDLD children, while minimising group differences between MonTD and L2-TD children?

3.1.5.4 Significance of the Study. Addressing these RQs would lay important groundwork for future studies examining the clinical utility and diagnostic accuracy of Cantonese NWR tests for identifying DLD in bilingual L2 Cantonese-speaking children. The findings could also add to the understanding of how lexicality and language-specificity of nonword stimuli, as well as scoring approaches, affects TD/DLD group differentiation when bilingual children are also taken into consideration.

3.2 Methods

3.2.1 Participants

Fifty-seven Cantonese-speaking children participated in this study[3]. The children were either recruited online or re-invited to take part in this research study, after

[3] Of these 57 children, 32 (16 MonTD and 16 MonDLD) were also included in the participant sample of Study I.

previously participating in other projects. All children attended local schools in Hong Kong, where Cantonese was the medium of instruction (MOI).

3.2.1.1 Monolingual DLD Group. The first group of children (N = 19, fourteen male), aged 8;1 to 11;10 (M = 9;8, SD = 1;0), were predominantly monolingual Cantonese-speaking children, who met the criteria for DLD (MonDLD), on the basis that they demonstrated poor language skills in the norm-referenced, Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS; T'sou et al., 2006), and that their language difficulties had a negative functional impact on daily social interactions or educational progress, as reported by parents and/or school personnel. In HKCOLAS, seventeen of these children scored at 1.25 SD below age means in two or more out of six subtests, and two scored at 1.25 SD below age means in one subtest and 1 SD below age means in another. One child had co-occurring attention deficit hyperactivity disorder (ADHD), and another child had co-occurring dyslexia.

3.2.1.2 Monolingual TD Group. The second group of children (N = 19, fourteen male), were predominantly monolingual TD children (MonTD), who were individually matched to each child in the DLD group in age (within four months of age difference on the day of testing, M = 9;8, SD = 1;1, range = 8;0 to 11;9), gender, and grade in school. These children scored age-appropriately in HKCOLAS and there were no parental concerns over any areas of development.

3.2.1.3 Bilingual TD Group. The third group of children (N = 19, seven male) were L1-Urdu-L2-Cantonese-speaking TD children (L2-TD), who were acquiring Cantonese as a second, school and community language, while using Urdu as a heritage language at home. These children were not individually matched to each monolingual DLD-TD pair in age, gender, and grade in school, but as a group, they had a comparable age range with the two groups of predominantly monolingual children (M = 9;7, SD = 1;1, range = 8;0 to 11;7). As there are no available norm-referenced tests to assess the language profiles of these L2-Cantonese-speaking children, their TD status was established through a parental questionnaire, which was adapted from the Language Impairment Testing in Multilingual Settings – Parents of Bilingual Children Questionnaire (LITMUS-PaBiQ; Tuller, 2015). Responses on the questionnaire confirmed that all participants were born at full-term, did not have significantly delayed

one-word and word-combination stages, did not have any other developmental delays, did not have a history of receiving speech therapy, did not have hearing impairments, have not been diagnosed with any other developmental disorders (including ADHD, dyslexia, ASD) and did not have family history of language impairments. Parents also expressed no concerns over the children's development in any areas, and Urdu was reported to be the strongest language of all children. Of these children, twelve had previously participated in another research study, where they had also been assessed by an experienced Urdu-speaking SLT in a range of language assessment tasks tapping lexical, morphosyntactic and narrative competence in their strongest language (i.e. Urdu). Their performance on these language tasks was considered developmentally appropriate based on the clinical judgment of the experienced SLT.

3.2.1.4 Other Considerations. No participants from any of the three groups were reported to have hearing impairments or ASD. All children from the MonDLD and MonTD groups passed a pure tone audiometry hearing screening test. They also scored standard scores of above 70 in Raven's Progressive Matrices (Raven et al., 1996), screening out the likelihood of intellectual disability, although the MonTD group (M = 110.4, SD = 13.1, range = 85 to >135) scored significantly higher than the MonDLD group (M = 101.6, SD = 12.2, range = 82 to 127), $t(36) = 2.14$, $p = .04$. The L2-TD group did not undergo pure tone audiometry hearing screening and tests for non-verbal intelligence, as the likelihood of having hearing impairments or intellectual disability was deemed to be very low, with no parent suspicion reported through the parental questionnaire for any participant, and all children were studying in mainstream schools with no expressed concerns from schools over the children's development in any areas either.

3.2.1.5 Ethics Statement. This study was carried out in accordance with the recommendations of the Human Subjects Ethics Sub-committee at the Hong Kong Polytechnic University (reference number: HSEARS20161230004). All parents gave written consent for their children to participate in the study.

3.2.2 Materials

Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS), Pure Tone Audiometry Hearing Screening Test, and Raven's Progressive Matrices were used to ascertain clinical status (TD/DLD) in monolingual children. These tests have already

been described in detail in Chapter II of this thesis, see Sections 2.2.2.1-2.2.2.3.

3.2.2.1 Parental Questionnaire for Establishing TD Status in L2-TD Children. The parental questionnaire used was an adapted version of the LITMUS-PaBiQ (Tuller, 2015). The following questions were asked: 1) whether the child was born at full term; 2) whether there were significant delays in the child's early language milestones; 3) whether the child has received speech and language assessments; 4) whether the child has received a diagnosis of any developmental disorders or language impairments; 5) whether the child has hearing impairments or had frequent ear infections; 6) whether the parent has concerns over the child's development in any areas; 7) whether there is a history of speech and language impairments in any family members; 8) years living in Hong Kong; 9) age at which regular exposure to Cantonese had begun; and 10) parent's subjective ratings on the child's ability to speak and understand Urdu and Cantonese (rated on a scale of 1 to 7, with 1 being "poor" and 7 being "excellent"). The parental questionnaire was administered in the form of a phone interview by an Urdu-speaking SLT, such that parents were able to use a language they are familiar with (Urdu in this case) when providing responses in the questionnaire.

3.2.2.2 NWR Stimuli. Three sets of language-specific nonwords (i.e. High-lexicality, Low-lexicality and High-Lexicality-Vowel-Matched), described in detail in Chapter II of this thesis (see Sections 2.2.2.4.1-2.2.2.4.5), and one set of quasi-universal, Cantonese-adapted cross-linguistic nonwords were used (see Appendix Materials for full list of NWR stimuli).

3.2.2.2.1 Cross-linguistic (CL-NWR) Nonwords. Cantonese adapted nonwords from the quasi-universal, CL-NWR test (Chiat, 2015) contained only cross-linguistically frequent consonants and vowels, had neutral prosodic features, and simple CV structures (there were no CVC nonwords). Like the other three sets of nonwords, CL-NWR nonwords ranged from two to five syllables in length. When compared with the three sets of language-specific nonwords in terms of lexicality, CL-NWR nonwords had a medium level of lexicality when adapted to Cantonese, with 57% (12/21) of its constituent syllables being morphemic in Cantonese, e.g. /si:bu/ (*si* meaning poem or silk in Cantonese, and *bu* being meaningless across all lexical tones). All syllables within CL-NWR nonwords were set to be articulated in Cantonese tone

one (i.e. high, level tone), except for the final syllables of each nonword, which were lengthened and articulated with a lower pitch (similar to tone six, low-mid level tone, in Cantonese) to mark the end of an utterance. This prosodic pattern was kept consistent across cross-linguistic adaptations of CL-NWR nonwords, in order to facilitate cross-linguistic comparisons. All consonants and vowels included were expected to be acquired by age 4;0 in speech production by monolingual Cantonese-speaking children according to developmental norms (To et al., 2013). All nonwords were recorded by a female native Cantonese-speaking student SLT.

3.2.3 Procedures

All experimental tasks were administered by trained native Cantonese-speaking experimenters, and diagnostic testing was conducted by Cantonese-speaking SLTs or student SLTs under the supervision of an experienced SLT. Children from the MonDLD and MonTD groups attended the testing session at a SLT clinic, where they completed a hearing screening, followed by the NWR task, then a standardised language assessment and finally a test of non-verbal intelligence quotient; the session lasted for approximately two hours. Children from the L2-TD group were tested through a home-visit session, as they did not have to undergo diagnostic testing at a clinic, with a lack of norm-referenced language assessment tools. The session started with a warm-up task, where children verbally named items, locations, occupations and items of clothing shown in pictures, before moving on to the NWR task. The session lasted for about 45 minutes.

The NWR task was separated into two experimental blocks, which were the quasi-universal block (consisting of Cantonese adapted CL-NWR items) and the language-specific block (consisting of High-Lexicality, Low-Lexicality and High-Lexicality-Vowel-Matched items); CL-NWR items were presented independently to facilitate future cross-linguistic comparisons. The order of presentation of items within each block was randomised, and the order of presentation of the two experimental blocks was counterbalanced across participants. Further details on the presentation of the NWR task have already been described in Chapter II of this thesis (See Section 2.2.3).

3.2.3.1 Scoring. Responses were audio recorded and transcribed. Performance on all nonwords was scored on both whole-nonword-level accuracy (i.e. responses must contain all and only the target segments in the correct order to be regarded as correct)

and syllable-level accuracy (i.e. each correctly repeated syllable within a response gets a score of one). Two sound variations in the responses were not regarded as incorrect, which are the omission of initial /ŋ/ consonant, and substitutions between final /k/ and final /t/ consonants, because these are well documented free variants that are prevalent even in adult native Cantonese speakers. Changes in prosody, such as tone, were also not regarded as incorrect, and there were very few instances of such changes in children's responses.

3.2.4 Inter-rater Reliability

3.2.4.1 MonTD and MonDLD Groups. Data from the two groups of monolingual Cantonese-speaking children (MonTD and MonDLD) were transcribed and scored by seven-native Cantonese speakers with linguistics training. Five completed the first round of transcriptions and scoring, and three independently transcribed 36.8% of the data (with one transcriber acting as the first transcriber for some children and the second transcriber for other children).

For NWR scores at whole-nonword-level of accuracy, the average measure Intra-class Coefficient (ICC) using a two-way mixed model and absolute agreement was .94 for High-Lexicality-Vowel-Matched nonwords (95% CI of .87 to .91); .89 for High-Lexicality nonwords (95% CI of .86 to .91); .89 for Low-Lexicality nonwords (95% CI of .70 to .98), and .93 for CL-NWR nonwords (95% CI of .91 to .95), indicating good to excellent levels of reliability between raters.

For NWR scores at syllable-level of accuracy, the average measure ICC using a two-way mixed model and absolute agreement was .89 for High-Lexicality-Vowel-Matched nonwords (95% CI of .88 to .90); .88 for High-Lexicality nonwords (95% CI of .87 to .89); .85 for Low-Lexicality nonwords (95% CI of .84 to .87), and .93 for CL-NWR nonwords (95% CI of .92 to .94), indicating good to excellent levels of reliability between raters.

3.2.4.2 L2-TD Group. The same approach was used to compute inter-rater reliability in the data from L2-TD group. Five native-Cantonese speakers with linguistic training transcribed and scored the data. One completed the first round of transcriptions and scoring, and four independently transcribed 36.8% of the data and scored NWR accuracy both at whole-item level and syllable level.

When NWR was scored at whole-nonword level, the average measure ICC using

a two-way mixed model and absolute agreement was .89 for High-Lexicality-Vowel-Matched nonwords (95% CI of .84 to .92); .87 for High-Lexicality nonwords (95% CI of .82 to .91); .79 for Low-Lexicality nonwords (95% CI of .71 to .85), and .92 for CL-NWR nonwords (95% CI of .89 to .95), indicating good to excellent levels of reliability between raters.

When NWR was scored at syllable level, the average measure ICC using a two-way mixed model and absolute agreement was .88 for High-Lexicality-Vowel-Matched nonwords (95% CI of .83 to .88); .85 for High-Lexicality nonwords (95% CI of .82 to .87); .83 for Low-Lexicality nonwords (95% CI of .80 to .86), and .86 for CL-NWR nonwords (95% CI of .82 to .88), indicating good to excellent levels of reliability between raters.

3.2.5 Data Analysis

NWR scores were analysed with Logistics Mixed Effect Models, using the R package lme4 (Bates & Maechler, 2010) in R (version 4.1.3, R Core Development Team, 2021). All assumptions required by logistic mixed effect models were met in the dataset.

Two statistical model were created to address each of the RQs, the first focussing on NWR scoring at a whole-nonword level; and the second focussing on NWR scoring at a syllable level. As both RQs asked whether each of the nonword sets were able to capture significant differences in NWR performance between MonDLD and MonTD groups, and between MonDLD and L2-TD groups, whilst minimising differences MonTD and L2-TD groups, the two models had the same independent variables and random effects. The independent variables were, Group (MonDLD vs. MonTD vs. L2-TD), Nonword Set (High-Lexicality vs. Low-Lexicality vs. High-Lexicality-Vowel-Matched vs. CL-NWR), and their interactions. The random effects added were Participant and Nonword Item. The dependent variable of Model 5 was NWR accuracy at whole-nonword level (scored as a categorical variable, correct vs. incorrect repetition of whole-nonword); and for Model 6, NWR accuracy at syllable level (scored as a categorical variable, correct vs. incorrect repetition of each nonword syllable).

3.3 Results

3.3.1 Model 5: NWR Accuracy at Whole-Nonword Level

The specification of Model 5 was as follows: NWRaccuracy_WholeNonword ~ Group + Nonword_Set + Group*Nonword_Set + (1|Participant) + (1|Nonword_Item).

The results of Model 5 are shown in Table 6. When NWR was scored at whole-nonword level, there was a significant main effect of Group, where MonTD children performed significantly better than the MonDLD group (p <.001), but there was no significant difference in NWR accuracy between MonDLD and L2-TD children (p =.15). There was also a significant main effect of Nonword Set, where children scored significantly lower on Low-Lexicality nonwords, compared to High-Lexicality-Vowel-Matched nonwords (p <.001), and there were no significant differences between performance on High-Lexicality and High-Lexicality-Vowel-Matched nonwords (p =.92). Children also scored significantly higher on CL-NWR nonwords than High-Lexicality-Vowel-Matched nonwords (p =.01), despite High-Lexicality-Vowel-Matched nonwords having a higher lexicality level.

Table 6. *Results of Model 5 on analysis of NWR accuracy at whole-nonword level in MonDLD, MonTD and L2-TD groups*

Fixed Effect	β	*SE*	*z*	*p*
(Intercept)	0.64	0.40	1.61	.11
Group (MonTD)	1.66	0.34	4.93	<.001
Group (L2-TD)	0.47	0.32	1.45	0.15
Nonword_Set (CL-NWR)	1.38	0.56	2.48	.01
Nonword_Set (Low-Lexicality)	-2.32	0.50	-4.66	<.001
Nonword_Set (High-Lexicality)	-0.05	0.49	-0.11	.92
Group (L2-TD) : Nonword_Set (CL-NWR)	-0.06	0.30	-0.19	.85
Group (MonTD) : Nonword_Set (CL-NWR)	-1.05	0.31	-3.37	<.001
Group (L2-TD) : Nonword_Set (Low-Lexicality)	0.06	0.26	0.24	.81
Group (MonTD) : Nonword_Set (Low-Lexicality)	-0.43	0.27	-1.59	.11
Group (L2-TD) : Nonword_Set (High-Lexicality)	0.53	0.25	2.17	.03
Group (MonTD) : Nonword_Set (High-Lexicality)	0.17	0.27	0.61	.54
Random Effect	Variance	*SD*		
Item	2.53	1.59		
Participant	0.74	0.86		

There was also a significant interaction between Group and Nonword Set, when NWR was scored at whole-nonword level. The interpretation of the interaction was assisted by plotting predicted probabilities of NWR accuracy at whole-nonword level, by Nonword Set, for each participant group separately (see Figure 3). Figure 3 shows that within each Nonword Set, MonTD children were predicted to have the highest NWR accuracy, followed by L2-TD children, then the MonDLD group; but the degree of group differentiation varied across the nonword sets. In terms of group differences between MonTD and MonDLD groups, all language-specific nonword sets captured substantial group differences, but CL-NWR nonwords did not, with the MonDLD group predicted to achieve similarly high levels of NWR accuracy as the MonTD group. Focussing on group differences between L2-TD and MonDLD groups, High-Lexicality Nonwords appeared to be the most effective in capturing such differences, while there were large overlaps in predicted probabilities of NWR accuracy in L2-TD and

MonDLD groups on all remaining nonword sets, particularly when the 95% CI were taken into consideration. Regarding group differences between MonTD and L2-TD groups, there was a close to complete overlap on predicted performance of the two groups on CL-NWR nonwords, suggesting that CL-NWR may be the most effective in minimising any disadvantage for L2-TD children, relative to MonTD children. There was some degree of disadvantage for L2-TD children, relative to the MonTD children, on all remaining nonword sets, which was especially prominent on High-Lexicality-Vowel-Matched nonwords.

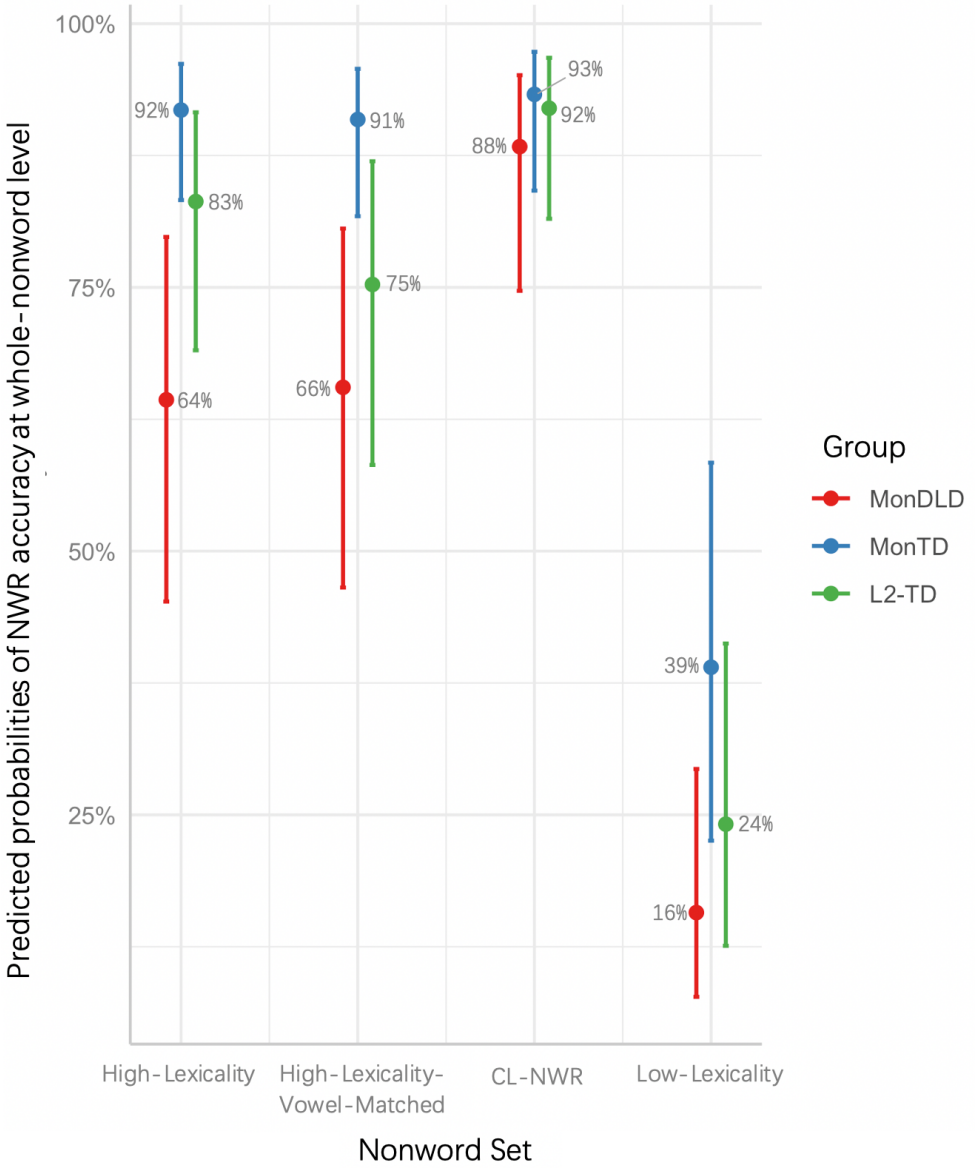

Figure 3. *Predicted Probabilities of NWR accuracy at whole-nonword level (with 95% confidence intervals) by Nonword Set in the MonDLD, MonTD and L2-TD groups*

Post-hoc pairwise comparisons were also conducted to examine the levels of group differentiation within each nonword set separately (see Table 7). Results showed there were significant group differences between the MonDLD and MonTD groups in High-Lexicality, High-Lexicality-Vowel-Matched and Low-Lexicality nonwords ($p \leq .008$) with medium to large effect sizes; but CL-NWR nonwords did not yield significant monolingual TD/DLD group differences. Regarding group differences between L2-TD and MonDLD groups, only High-Lexicality nonwords captured significant group differences ($p = .02$), while other nonword sets did not. Comparing between MonTD and L2-TD groups, while CL-NWR, High-Lexicality and Low-Lexicality nonwords did not capture significant group differences, demonstrating no disadvantage to the L2-TD children, High-Lexicality-Vowel-Matched nonwords yielded significant group differences, suggesting a disadvantage for L2-TD children. Taken together, High-Lexicality nonwords appeared to be the best at capturing TD/DLD group differences between both MonTD and MonDLD, as well as L2-TD and MonDLD groups, without disadvantaging L2-TD children compared to MonTD children, when NWR was scored at whole-nonword level.

Table 7. *P-values (and effect sizes in odds ratios) for pairwise comparisons between predicted group means of NWR accuracy at whole-nonword level in MonDLD, MonTD & L2-TD children, for each nonword set*

Lexicality Level	MonDLD & MonTD	MonDLD & L2-TD	MonTD & L2-TD
High-Lexicality-Vowel-Matched	<.001 (4.27)	.23 (1.56)	.001 (0.37)
CL-NWR	.32 (1.73)	.34 (1.71)	1.00 (0.99)
High-Lexicality	<.001 (5.80)	.02 (2.59)	.08 (0.45)
Low-Lexicality	<.008 (3.42)	.36 (1.75)	.23 (0.51)

3.3.2 Model 6: NWR Accuracy at Syllable Level

The specification of Model 2 was as follows: NWRaccuracy_Syllable ~ Group + Nonword Set + Group*Nonword_Set + (1|Participant) + (1|Nonword_Item).

The results of Model 6 are shown in Table 8. At a syllable-level scoring, NWR

accuracy was significantly better in the MonTD group, compared to the MonDLD group (p <.001), and NWR accuracy also significantly better in L2-TD group, compared to the MonDLD group (p =.003). The main effect of Nonword Set was also significant, with better performance on syllables within CL-NWR items, relative to syllables within High-Lexicality items (p =.008), and lower NWR accuracy on syllables within Low-Lexicality items, compared to syllables within High-Lexicality items (p <.001); there was no difference between performance on syllables within High-Lexicality and High-Lexicality-Vowel-Matched items (p =.92), which shared the same level of lexicality.

Table 8. *Results of Model 6 on analysis of NWR accuracy at syllable level in MonDLD, MonTD and L2-TD groups*

Fixed Effect	β	*SE*	*z*	*p*
(Intercept)	1.99	0.25	7.99	<.001
Group (MonTD)	1.38	0.23	5.91	<.001
Group (L2-TD)	0.66	0.22	2.93	.003
Nonword_Set (High-Lexicality-Vowel-Matched)	-0.03	0.29	-0.11	.92
Nonword_Set (CL-NWR)	0.89	0.34	2.65	.008
Nonword_Set (Low-Lexicality)	-1.17	0.29	-4.10	<.001
Group (MonTD) : Nonword_Set (High-Lexicality-Vowel-Matched)	0.01	0.18	0.08	.94
Group (L2-TD) : Nonword_Set (High-Lexicality-Vowel-Matched)	-0.24	0.15	-1.61	.11
Group (MonTD) : Nonword_Set (CL-NWR)	-0.56	0.22	-2.60	.009
Group (L2-TD) : Nonword_Set (CL-NWR)	0.08	0.21	0.39	.70
Group (MonTD) : Nonword_Set (Low-Lexicality)	-0.66	0.15	-4.32	<.001
Group (L2-TD) : Nonword_Set (Low-Lexicality)	-0.26	0.14	-1.87	.06
Random Effects	Variance	*SD*		
Item	0.86	0.93		
Participant	0.36	0.60		

There was also a significant interaction between Group and Nonword Set, when NWR was scored at syllable level. The interpretation of the interaction was assisted by plotting predicted probabilities of NWR accuracy at syllable level, by Nonword Set, for each participant group separately (see Figure 4). Figure 4 shows that among the four sets of nonwords, CL-NWR nonwords were the only nonword set that were able to effectively minimise group differences between MonTD and L2-TD groups, whist also capturing higher NWR accuracy on the two TD groups, compared to the MonDLD group. On the remaining sets of nonwords, the MonDLD group was consistently predicted to score lower than the MonTD group, and the L2-TD group is predicted to score between the two monolingual groups – on High-Lexicality and Low-Lexicality nonwords, there is substantial overlap in the predicted performance of the L2-TD group and both MonTD and MonDLD groups, but on High-Lexicality-Vowel-Matched nonwords, the predicted performance of the L2-TD group was noticeably below that of the MonTD group, and overlaps greatly with the MonDLD group.

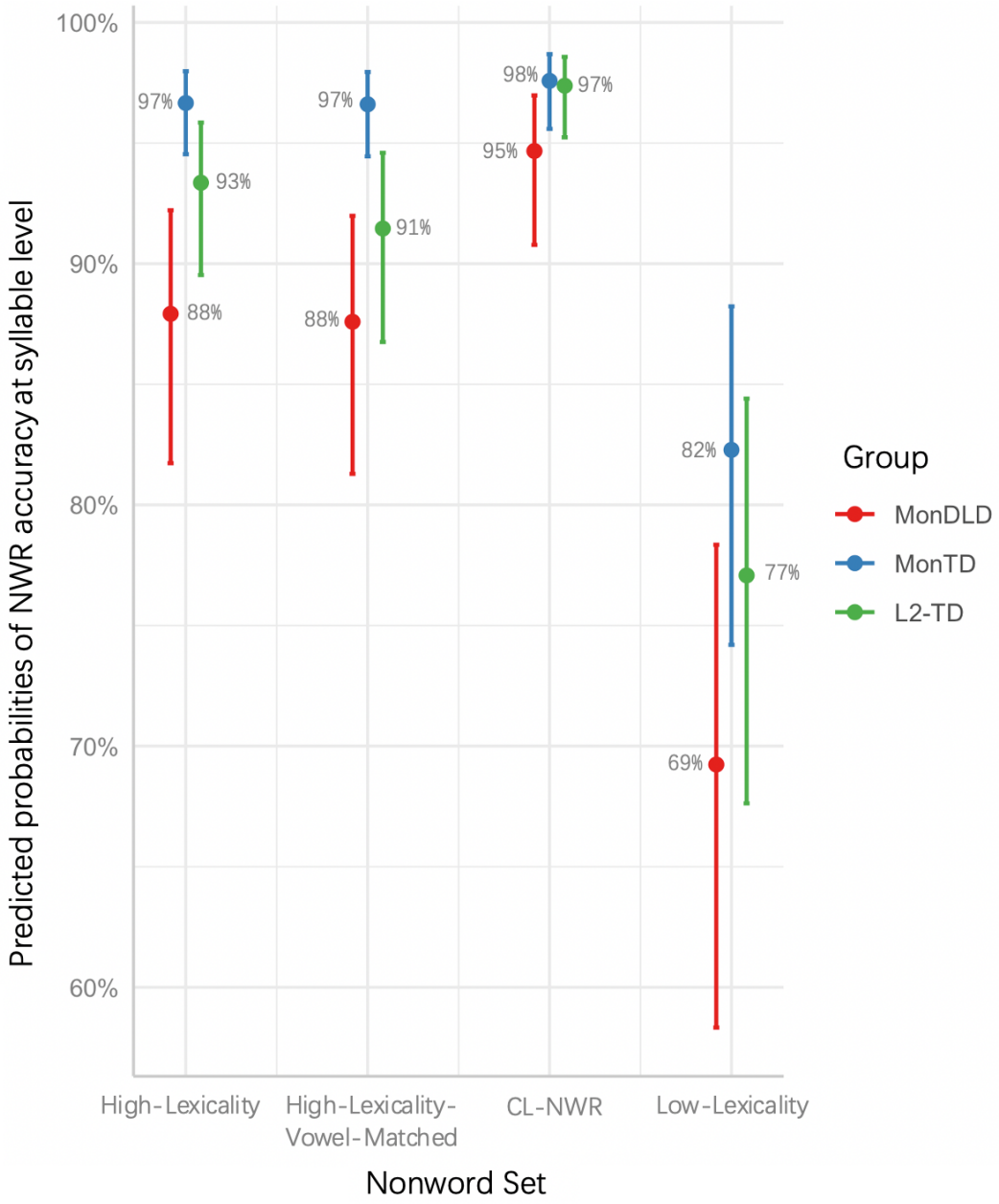

Figure 4. *Predicted Probabilities of NWR accuracy at syllable level (with 95% confidence intervals) by Nonword Set in the MonDLD, MonTD and L2-TD groups*

Post-hoc pairwise comparisons were also conducted to examine the levels of group differentiation within each nonword set (see Table 9). Table 9 shows that consistent with previous findings, all language-specific nonwords (i.e. High-Lexicality, Low-Lexicality and High-Lexicality-Vowel-Matched nonwords) were able to generate statistically significant differences between MonDLD and MonTD groups (p <.001), with MonDLD scoring lower than MonTD. In addition, when scored at syllable level,

Cantonese CL-NWR nonwords also generated significant group differences between MonDLD and MonTD groups (p =.03), in the expected direction. Regarding the group difference between the MonDLD and L2-TD groups, CL-NWR nonwords were the only stimulus set that successfully captured significant differences between the two groups (p =.03), while the other nonword sets did not. Furthermore, focussing on the difference in predicted NWR accuracy between the MonTD and L2-TD groups, no significant group differences were captured on CL-NWR and Low-Lexicality nonwords – in particular, on CL-NWR nonwords, the p-value of 1.00 indicated that the L2-TD children performed as accurately as MonTD children, suggesting that they were not disadvantaged despite their L2 and bilingual status. Overall, when NWR was scored at syllable level, Cantonese CL-NWR nonwords appeared to be the best at capturing TD/DLD group differences between both MonTD and MonDLD, as well as L2-TD and MonDLD groups, without disadvantaging L2-TD children compared to MonTD children.

Table 9. *P-values (and effect sizes in odds ratios) for pairwise comparisons between predicted group means of NWR accuracy at syllable-level in MonDLD, MonTD & L2-TD children, for each nonword set*

Lexicality Level	MonDLD & MonTD	MonDLD & L2-TD	MonTD & L2-TD
High-Lexicality-Vowel-Matched	<.001 (3.57)	0.22 (1.48)	.001 (0.41)
CL-NWR	.03 (2.26)	.03 (2.26)	1.00 (1.00)
High-Lexicality	<.001 (3.76)	.09 (1.78)	.03 (0.47)
Low-Lexicality	<.001 (2.02)	.07 (1.50)	.25 (0.75)

3.4 Discussion

Study II explored the potential of Cantonese language-specific and quasi-universal NWR stimuli to disentangle the effects of language impairment and bilingualism, using two scoring approaches. Specifically, it investigated whether Cantonese language-specific nonwords (including High-Lexicality, Low-Lexicality, High-lexicality-Vowel-Matched nonwords), and cross-linguistic nonwords capture significant group differences between MonTD and MonDLD children, and between L2-TD and MonDLD children, while minimising group differences between MonTD and L2-TD

children, when NWR is scored on (1) whole-nonword accuracy and (2) syllable accuracy.

3.4.1 NWR performance in MonDLD, MonTD and L2-TD groups

This study is the first to document that Cantonese NWR stimuli have potential to disentangle the effects of bilingualism and DLD. As established earlier, the ideal scenario for using NWR to assess bilingual L2 children is whether nonwords maximise differences between TD and DLD groups, regardless of monolingual or bilingual status, while simultaneously minimizing the gap between monolingual and bilingual TD groups, despite bilingual L2 learners having weaker lexical and sub-lexical representations to support NWR than monolingual children due to reduced language exposure and proficiency associated with bilingual language acquisition. The present results showed that at both whole-nonword and syllable levels of scoring, most sets of our Cantonese nonwords yielded significant TD/DLD group differences among monolingual children, and certain sets of NWR stimuli were also able to capture significant differences between L2-TD and MonDLD groups, whilst minimizing differences between MonTD and L2-TD children. Therefore, the findings of this study are the first to provide evidence supporting the development of Cantonese NWR tests into assessment tools for DLD in bilingual L2 learners of Cantonese. Findings also bear on the type of nonwords that may be optimal for this purpose.

3.4.2 Optimal NWR Tasks for Assessing Bilingual Children in their L2

As certain nonword sets were found to be more suited than others for assessing bilingual L2-Cantonese-speaking children, in that they do not disadvantage against L2-TD children, these sets of nonwords will be discussed first in the following sections.

3.4.2.1 CL-NWR Nonwords. Cantonese-adapted, CL-NWR nonwords were one of the stimulus sets that was found to be able to capture significant group differences between MonTD and MonDLD and between L2-TD and MonDLD groups, while avoiding differences between L2-TD and MonDLD children, when scored at syllable-level. This study demonstrates for the first time that CL-NWR nonwords, when adapted to Cantonese, can effectively capture TD/DLD group differences among monolingual-Cantonese-speaking children, as well as between L2-TD and MonDLD children, adding to findings on the Dutch CL-NWR test (using whole-item scoring; Boerma et

al., 2015) with evidence from a typologically different language. Having said so, it is also relevant to note that, similar to findings on the Swedish CL-NWR task (Öberg & Bohnacker, 2022), the present data also indicated substantial overlap in performance between the MonDLD and MonTD groups, and between the MonDLD and L2-TD groups (see Figure 4), despite significant TD/DLD differences yielded at the group level, indicating that the clinical accuracy of the Cantonese CL-NWR task may be lower at individual level. This may be related to the older age of participants within this study, as evidenced by a near-ceiling effect on Cantonese CL-NWR test for all groups of children – even MonDLD children were predicted to have, on average, a 95% probability of correct repetition on syllables within CL-NWR items, suggesting that even children with a DLD diagnosis did not find the task challenging. These findings are also in line with previous data suggesting that at above six years of age, the magnitude of TD/DLD group differences and clinical accuracy of the CL-NWR task begins to drop, even though TD/DLD group differences remain significant and large (Boerma & Blom, 2021). Based on the current findings at group level, our data suggest that quasi-universal, CL-NWR nonwords, have potential to be further developed into informative assessment tools for L2 Cantonese-speaking children, especially for children below eight years of age.

While significant L2-TD/MonDLD group differences were only observed with syllable-level scoring (see later Section 3.4.3 on scoring approaches for a more thorough discussion), it nevertheless provides further evidence to conclusions drawn from Study I, that Cantonese is not a true cross-linguistic exception in NWR, and for the potential utility of CL-NWR nonwords in assessing monolingual and bilingual children for DLD in a Cantonese context. Moreover, the finding that MonTD and L2-TD groups achieved close to the exact same levels of NWR accuracy on CL-NWR nonwords in this study, demonstrating that CL-NWR nonwords do not disadvantage L2-TD children, even though they may have less experience with the testing language. This is presumably because CL-NWR nonwords only utilise cross-linguistically frequent consonant and vowels arranged into simple syllable structures with limited prosodic structure, which allow bilingual TD children to use their linguistic knowledge from any and all languages they are acquiring to support redintegration during NWR. In fact, all consonant and vowels in the Cantonese-adapted version of CL-NWR test are also present in the Urdu phonemic inventory (Ambreen & To, under review),

meaning that L2-TD children could also draw on their lexical and sub-lexical knowledge of Urdu (i.e. their L1) when repeating Cantonese CL-NWR nonwords.

3.4.2.2. Language-Specific, High-Lexicality Nonwords. Cantonese language-specific, High-Lexicality nonwords were the second stimulus set found not to disadvantage L2-TD children, when NWR was scored at whole-nonword level. High-Lexicality nonwords were also able to capture significant group differences between MonTD and MonDLD groups, as previously demonstrated in Study I, and between L2-TD and MonDLD groups. Furthermore, High-Lexicality nonwords captured larger group differences between MonTD and MonDLD groups, and between L2-TD and MonDLD groups, compared to CL-NWR nonwords, suggesting that they are optimal for maximising differences between TD children and those with DLD, regardless of the monolingual or bilingual status of the TD children. This is presumably because nonwords with higher levels of lexicality and sub-lexicality allowed monolingual and bilingual L2 TD children, who have better access to lexical and sub-lexical representations than those with DLD, to draw on their long-term linguistic knowledge in the redintegration process during NWR. L2-TD children were also not disadvantaged by High-Lexicality nonwords, relative to MonTD children, suggesting that this set of High-Lexicality nonwords allowed L2-TD children to benefit from their lexical and sub-lexical representations from one or both languages (Urdu and Cantonese in this case) to support NWR. Together, the findings suggest that a combination of High-Lexicality nonwords and Cantonese CL-NWR nonwords might be most effective in capturing group differences between TD children, monolingual or bilingual, and monolingual children with DLD.

3.4.2.3 Language-Specific, High-Lexicality-Vowel-Matched Nonwords. Interestingly, High-Lexicality-Vowel-Matched nonwords, which were of the same level of lexicality as High-Lexicality nonwords by design, did not generate the same pattern of group differences – at both whole-item and syllable levels of scoring, there was substantial overlap between performance by L2-TD and MonDLD groups, yielding non-significant differences between these two groups, while L2-TD children scored significantly lower than MonTD children. These findings suggested that some nonwords with high lexicality levels, like High-Lexicality-Vowel-Matched nonwords, may still disadvantage L2-TD learners, presumably because there are factors other than

lexicality that affect NWR performance, especially in L2-TD children. Particularly, studies on NWR in monolingual children, including Study I of this thesis, have demonstrated that sub-lexical representations also support NWR, where children repeat nonwords more accurately when nonwords had higher levels of phonotactic probability (McKean et al., 2013; Szewczyk et al., 2018) or attested CV combinations (Stokes et al., 2006). If L2 Cantonese-speaking TD children also draw on their sub-lexical representations to support NWR, the difference in their performance on High-Lexicality and High-Lexicality-Vowel-Matched nonwords could perhaps be explained by differences in sub-lexical factors, such as phonotactic probability, between the two sets of nonwords, despite them sharing equally high levels of lexicality in terms of morphemicity; future studies could examine how these sub-lexical factors affect NWR in L2-TD learners of Cantonese.

3.4.2.4 Language-Specific, Low-Lexicality Nonwords. The data also showed that Low-Lexicality nonwords are not optimal for disentangling language impairment and language experience, given that they did not capture significant group differences between MonDLD and L2-TD groups at both whole-item and syllable levels of scoring. Although the L2-TD group also performed similarly to the MonTD group (as evidenced by the lack of significant difference between MonTD and L2-TD groups), the lack of significant group difference between L2-TD and MonDLD groups suggested that at least some children from the L2-TD group were disproportionally challenged by Low-Lexicality nonwords, compared to MonTD children.

Even though Low-Lexicality nonwords should theoretically be equally challenging to both monolingual and bilingual TD children, in terms of them sharing non-morphemic status, L2-TD children may be further taxed by (Cantonese) language specific elements in the present stimulus set. For example, Low-Lexicality nonwords included the Cantonese initial velar consonant /ŋ-/, initial rounded labial-velar approximant /w/, and final unreleased stop consonants /-p/, /-t/ and /-k/, all of which occur in Cantonese but not occur in Urdu (Ambreen et al., under review). Therefore, at least some L2 learners could still be disproportionally disadvantaged due to reduced experience of Cantonese as a L2 and lack of support from L1.

Additionally, the plotted predicted probabilities of NWR accuracy also indicated wide 95% confidence intervals on Low-Lexicality nonwords, demonstrating great within-group variability and substantial overlap in NWR performance across groups.

Thus, Low-Lexicality nonwords are less optimal for assessing bilingual L2 learners of Cantonese.

3.4.3 Scoring of NWR Accuracy at Whole-Nonword Level vs. Syllable Level

The present data suggested that when using Cantonese NWR stimuli to assess Cantonese-speaking MonTD, MonDLD and L2-TD groups, scoring both at whole-item and syllable levels were able to maximise TD/DLD group differences while minimizing monolingual/bilingual TD group differences depending on the set of nonword stimuli. As NWR accuracy at whole-nonword level is already commonly adopted in NWR studies and has been demonstrated to differentiate between TD and DLD groups in both monolingual and bilingual children (Schwob et al., 2021), the following discussion will be focused on the less-used, syllable-level scoring approach.

The different patterns of findings on the two scoring approaches may be related to the level of detail captured by each of the two scoring methods. NWR accuracy at syllable level could be seen as a more lenient level of scoring, as children are still able to score when they correctly repeat only certain components of a nonword, instead of being penalised as soon as one mistake has been made within a nonword, which would be the case when NWR is scored at whole-nonword level. Our finding that Cantonese CL-NWR nonwords only captured significant group differences between MonTD and MonDLD groups and between MonDLD and L2-TD groups when syllable-level scoring is adopted demonstrated the phenomenon that while MonDLD children were repeating whole nonwords at a similar level of accuracy to the MonTD group, MonTD children accurately repeated more components within each nonword compared to MonDLD children. Similarly, L2-TD children had more difficulties with accurately repeating whole nonwords than MonTD children, but they were also making fewer mistakes within each nonword when compared to MonDLD children, and such nuanced differences across the groups could only be captured by a more detailed level of scoring, such as syllable-level accuracy.

In addition to its benefits of being a more fine-grained measure of NWR, scoring NWR performance at syllable-level is also relatively quick and straight-forward compared to even more fine-grained measures documented in the literature, such as scoring NWR in percentage of phonemes correct (PPC), suggesting that future studies might also explore this method of NWR scoring in other language versions of NWR tests.

3.4.4 Limitations and Future Directions

Whilst being the first study to report that Cantonese NWR stimuli are capable of generating group differences between MonDLD and L2-TD children, and simultaneously minimising group differences between monolingual and bilingual L2 TD children in a Cantonese context, this study represents only the first steps in research on developing NWR as an assessment tool helping to identify DLD in both monolingual and bilingual Cantonese-speaking children.

First, this study did not include a bilingual DLD group, thus it is yet to be determined how Cantonese-speaking children with both reduced language experience (L2) and impaired language learning capacity (DLD) perform in NWR compared to other groups of children. Future studies will be in a better position to examine NWR performance in an L2-DLD group, when assessment tools targeting DLD in L2 Cantonese become available, and when guidelines and methods for identifying DLD in L2 Cantonese children are better established.

Second, it would be beneficial to further increase the sample size, particularly that of the bilingual L2-TD group, given the substantial heterogeneity in children acquiring multiple languages.

Third, the current findings may be specific to L1-Urdu-L2-Cantonese-speaking children residing in Hong Kong. Whether these Cantonese NWR stimuli have potential to be developed into assessment tools for assessing bilingual Cantonese-speaking children acquiring languages other than Urdu awaits further investigation. Having said so, the findings on CL-NWR nonwords are expected to be more generalizable to children acquiring other L1s, as CL-NWR nonwords are designed to minimise the potential influence of language-specific knowledge. Apart from evaluating this expectation, future studies could examine whether the present findings on language-specific High-Lexicality nonwords generalise to L2-TD Cantonese-speaking children with other L1s.

Moreover, while this study focused on bilingual L2 Cantonese children, future research could also extend the scope to examine bilingual L1 Cantonese children with and without DLD who are developing their first language under heavy influence from another language (e.g. children who acquire Cantonese as their first, heritage and minority language in countries having another language as the majority community language), to further examine the diagnostic potential of our Cantonese NWR stimuli.

Finally, the present study only addressed the ability of Cantonese NWR stimuli to capture differences between MonTD and MonDLD, and between L2-TD and MonDLD at a group level. For Cantonese NWR to be developed into an assessment tool with diagnostic value, further research is needed to determine how accurate Cantonese NWR classifies individual children into TD and DLD groups, regardless of monolingual or bilingual status, by investigating sensitivity and specificity of NWR stimuli in a Cantonese context.

3.5 Conclusion

The results of Study II suggest that certain Cantonese nonword stimuli is able to maximize the differences between MonTD and MonDLD groups, and between L2-TD and MonDLD groups, while minimizing the differences between MonTD and L2-TD children. In particular, when language-specific, High-Lexicality nonwords are scored at whole-nonword level, and when quasi-universal, Cantonese CL-NWR nonwords are scored at syllable level, L2-TD children were not disadvantaged relative to monolingual TD children, despite them having reduced language knowledge and experience in Cantonese, associated with bilingual language acquisition. The findings demonstrated potential for these Cantonese NWR stimuli to be developed into clinically informative assessment tools for identifying DLD in bilingual Cantonese-speaking children.

3.6 Summary of Chapter III

To extend the line of research on Cantonese NWR, following positive findings on monolingual Cantonese-speaking children reported in Study I, Study II investigated whether Cantonese NWR stimuli are able to minimize disadvantage for bilingual L2-TD children, compared to MonTD children. NWR performance in 19 MonTD, 19 MonDLD and 19 L1-Urdu-L2-Cantonese TD children on three sets of language-specific nonwords and one set of quasi-universal nonwords was examined. When NWR accuracy was scored at whole-nonword level, language-specific, High-Lexicality nonwords captured group differences between DLD and TD groups (both monolingual and L2), while not disadvantaging the L2-TD group, compared to MonTD children. When NWR accuracy was scored at syllable level, quasi-universal, CL-NWR was the only set of nonwords that did not disadvantage L2-TD children relative to monolingual TD children, while still being able to generate significant group differences between the monolingual DLD and L2-TD groups. These findings provide supporting evidence

from an understudied language like Cantonese, that NWR can disentangle the effects of language impairment (in MonDLD) and bilingualism (in L2-TD). Future studies can explore how an L2-DLD group performs relative to the examined groups, explore alternative bilingual groups of children acquiring other L1-L2 combinations, and explore sensitivity and specificity of Cantonese NWR in a bilingual context.

CHAPTER IV – Study III
Sub-lexical Effects on Nonword Repetition in Cantonese-speaking Children

4.1 Introduction

Recent studies on NWR have found that in addition to lexicality of nonwords, sub-lexicality in nonword stimuli, most notably in terms of phonotactic probability, also plays a significant role in NWR performance (e.g. Szewczyk et al., 2018). This suggests that children's ability to develop and access sub-lexical, phonological representations contributes substantially to their success in the repetition of novel phonetic information.

As Study I and Study II of this thesis have already thoroughly examined the effects of lexicality on NWR performance and TD/DLD group differentiation in Cantonese-speaking children, this chapter presents an additional study, Study III, which dives deeper into sub-lexical effects of Cantonese NWR stimuli on NWR performance in Cantonese-speaking children. Given that Cantonese is typologically distinct from most previously examined languages, this study takes into consideration the unique phonotactic properties of Cantonese, and proposes a novel measure of sub-lexicality, Neighbourhood Density (ND) of Syllables, that is particularly suited for addressing whether sub-lexicality in Cantonese nonword stimuli affects children's NWR performance.

4.1.1 Effects of CV Combination Attestedness on NWR Performance in Cantonese-Speaking Children

In the context of Cantonese, the only sub-lexical factor in NWR stimuli examined previously is CV combination attestedness. As already established in Study I (see Chapter II, Section 2.2.2.1), Stokes et al. (2006) examined pre-schooler's NWR performance on two types of syllables, labelled IN and OUT, where IN syllables consisted of CV (consonant, vowel) combinations that are attested in Cantonese, and OUT syllables consisted of CV combinations that are unattested in Cantonese. Their results indicated that both children with DLD and their age-matched TD peers performed better on IN syllables than OUT syllables, suggesting that sub-lexical factors in NWR stimuli, specifically CV combination attestedness in this case, do affect NWR performance in Cantonese-speaking children, similarly to children acquiring other languages. Study I of this thesis also conducted a follow-up analysis on the effects of

CV combination attestedness in Cantonese NWR, on a subset of NWR stimuli (specifically, Low-Lexicality nonwords) that were comparable in design to the stimuli used in Stokes et al. (2006), and reported findings with the same pattern of results, where IN syllables were repeated more accurately than OUT syllables.

Furthermore, the two studies also found that IN syllables captured greater TD/DLD group differences in Cantonese-speaking children – although this finding was non-significant in Stokes et al. (2006), Study I reported significant findings with effect sizes confirming an interaction effect between CV combination attestedness and participant group. These results indicated that sub-lexicality of nonword stimuli, particularly CV combination attestedness, can affect Cantonese-speaking children with and without DLD differently, where TD children were found to have greater benefits from attested CV combinations in nonwords than children with DLD, presumably because they have better access to stronger sub-lexical representations to support NWR.

4.1.2 Effects of Phonotactic Probability on NWR Performance

While Cantonese-focused studies have only examined the sub-lexical property of CV combination attestedness and its effects on NWR accuracy, most studies on other languages that have argued for the relevance of sub-lexicality in NWR have found supporting evidence from findings that children repeat nonwords with higher phonotactic probability (PP) at a higher accuracy than nonwords with lower PP (Coady & Aslin, 2004; McKean et al., 2013; Munson et al., 2005; Rispens et al., 2015; Thorn & Frankish, 2005; Zamuner et al., 2004). PP quantifies the degree to which phoneme combinations within a word are common within a given language, with the most common measure of PP being phonemic bigram frequency, which indicates the mean frequency of all adjacent phoneme pairs within a nonword. A more recent study by Szewczyk et al. (2018) also examined an alternative measure of PP – mean ngram frequency of all phonemes – where the calculation of PP takes into account the frequency of occurrence of phonemic chunks at all different grain sizes within a word (from bigrams, to trigrams, and so on), and found it to be an even stronger predictor of NWR accuracy than mean bigram frequencies.

The authors of these studies argue that because sub-lexicality in nonwords influence children's NWR performance, children likely rely on sub-lexical representations, including phonemic bigrams and or phonemic chunks at all grain sizes (as in the case for ngrams), to support the repetition of nonwords. Such support is

analogous to the manner in which lexical representations assist NWR, where children use their long-term knowledge of phonotactic and other lexical and sub-lexical information to restore the rapidly decaying trace of a nonword within working memory – this process is known as redintegration (Gathercole, 1999; Schweickert, 1993).

A study has even argued that sub-lexical representations have a facilitative effect on NWR that is above and beyond that of lexical representations, based on the findings that PP is a stronger predictor of NWR performance, compared to neighbourhood density (ND; Szewczyk et al., 2018), which will be discussed in the following section.

4.1.3 Effects of Neighbourhood Density on NWR Performance

Neighbourhood density, commonly considered as a lexical (rather than sub-lexical) measure, measures the degree to which a word or nonword is phonologically similar to other words in a language; words or nonwords with a dense phonological neighbourhood indicate those that have many similar sounding words in a given language, whereas words or nonwords with a sparse neighbourhood indicate that they are phonologically dissimilar to other real words in a language. ND is most commonly measured as the number of words in a given language with a Levenshtein distance of 1 from the target, that is, the number of real lexical items with a one phoneme difference from the target – such "differences" can be in forms of addition, substitution, or deletion of a phoneme.

In the processing of real words, such as in receptive word recognition tasks, high ND has a negative effect on processing speed and accuracy, due to a competition effect arising from the activation of more neighbours for items with dense neighbourhoods (Rispens et al., 2015). On the other hand, high ND is often found to have a facilitative effect on the production of nonwords, such as in nonword repetition, as denser phonological neighbourhoods provide stronger phonological representations of word forms for the repairment of a decaying memory trace of the newly-heard nonword (Janse & Newman, 2013; Roodenrys & Hinton, 2002).

Between the sub-lexical measure, PP, and the lexical measure, ND, a recent study found ND to be a weaker predictor of NWR accuracy compared to PP (Szewczyk et al., 2018); although other studies have found the opposite to be true (Janse & Newman, 2013; Roodenrys & Hinton, 2002), and some studies suggested that both predictors explained unique variance in NWR accuracy (Bailey & Hahn, 2001; Thorn & Frankish, 2005). Thus, there is inconclusive evidence about whether sub-lexical representations

are more important than lexical representations for the successful repetition of nonwords, but it is widely acknowledged that both types of knowledge play an important role in NWR alongside each other.

4.1.4 A Discussion on Phonotactic Probability and Neighbourhood Density as Lexical vs. Sub-Lexical Measures

Traditionally, PP and ND have been pitted against each other as predictors of NWR performance to examine whether sub-lexical or lexical representations contributed more heavily towards successful NWR. PP was taken as a measure that concerned sub-lexical representations, because it measures the frequency of occurrence of a phonemic sequence (typically bigrams) *within* a nonword, while ND has been considered a measure that concerned lexical representations, because it measures the degree to which the nonword item as a whole is similar to real words in a lexicon.

However, this study proposes that PP and ND can each be either measures of lexicality or sub-lexicality. Consider monosyllabic words with a CV structure, such as *sea* (/si/), or *bar* (/bɑː/), measuring their PP in terms of the frequency in which the phonemic bigrams occur in the English language would render PP a lexical measure, given that these bigrams form the entirety of the lexical items. Equally, similar to how PP can be calculated for phonemic chunks of all grain sizes, ND may also be computed for smaller phonemic chunks within nonwords, rendering it a sub-lexical measure. Therefore, PP and ND should not be taken as absolute measures of lexicality or sub-lexicality, but rather, they simply describe two different features of word-likeness at a sub-lexical level – PP quantifies the degree to which a lexical or sub-lexical element is *frequently occurring* in a language, whereas ND quantifies the degree to which a lexical or sub-lexical element is *similar* to others in a language.

4.1.5 A Novel Measure of Sub-Lexicality: Neighbourhood Density of Syllables

Based on the discussions above, this study presents a new approach to calculating ND – rather than computing ND for entire nonwords, NDs can be computed for each independent syllable within a nonword item, to measure *ND of Syllables*. For example, if we take a nonword, *bamudi*, and calculate its ND using a traditional, lexical approach, its ND would be 0 in English, as there are no English words that are within a one-phoneme difference from the target; the new, sub-lexical approach of calculating ND

instead identifies neighbours for each of the constituent syllables, *ba*, *mu*, and *di*, searching for and counting neighbours such as /baɪ/ (as in *buy*) for *ba*, /hu/ (as in *who*) for *mu*, and /si/ (as in *see*) for *di*.

Using this newly proposed method of calculating ND of Syllables, rather than for entire nonwords, is both conceptually and methodologically interesting in the context of NWR. Conceptually, most studies that have examined the effects of sub-lexicality on NWR have focused solely on PP, which typically targets sub-lexical representations at the unit of phonemic bigrams, but studies have rarely explored alternative sub-lexical measures that target sub-lexical representations at other levels, such as syllabic level, or sub-syllabic level (e.g. onsets, nuclei and codas).

Methodologically, due to the nature that ND for entire words decreases drastically with an increasing number of syllables (given that longer nonwords have to be matched with potential neighbours in more phonemes than short, monosyllabic nonwords) and drops to zero quickly, many studies that have examined ND in nonwords only investigated monosyllabic nonwords (e.g. Janse & Newman, 2013; Thorn & Frankish, 2005), even though it is extremely uncommon for NWR tasks to test only monosyllabic nonwords in a clinical context. Other studies (e.g. Szewczyk et al., 2018) that have examined ND in longer, 2- to 4- syllable nonwords may not have been able to create good-enough variability in ND values to capture the variance in children's NWR accuracy. Thus, previous findings may have underestimated the role played by ND in NWR.

4.1.6 Phonotactic Probability and Neighbourhood Density in Cantonese

To examine whether sub-lexical factors in NWR stimuli affect NWR performance in Cantonese-speaking children, ND of Syllables could be a better candidate measure over PP of phonemic bigrams, due to the unique phonotactic properties of Cantonese.

Cantonese phonologists have argued that relative to English, PP plays a weaker role in Cantonese word-likeness, and ND provides a stronger cue to Cantonese word-likeness than it does in English (Kirby & Yu, 2007). This has been demonstrated in a Cantonese subjective word-likeness judgment experiment, where a stronger correlation between ND and word-likeness judgments for both Cantonese words and nonwords was reported, while there was a less pronounced correlation between PP and word-likeness ratings.

The authors attributed this finding to the phonotactic nature of the two languages.

One apparent phonotactic difference between Cantonese and English is that complex onsets and codas, such as consonant clusters in the syllable initial and final positions, are only permitted in English, and not in Cantonese. This results in a larger number of possible monosyllables in English (approximately 158,000 logically possible syllables in English; Jespersen, n.d., as cited by Kirby & Yu, 2007), relative to Cantonese (5,130 logically possible syllables; Kirby & Yu, 2007). Despite the larger phonotactic space of English, English only has around 10,000 non-homophonous monosyllables, taking up about 6% of the phonotactic space, whereas Cantonese has around 1,900 non-homophonous monosyllables, occupying a larger proportion (approximately 36%) of the phonotactic space. Such discrepancy means that for Cantonese nonwords and monosyllables, there is a much higher chance to have real-word neighbours within a one-phoneme difference, resulting in ND being a relatively strong cue to word-likeness ratings in Cantonese. In English, not only is it rarer for nonword syllables to have phonological neighbours due to lexical items occupying a much smaller proportion of the potential phonotactic space, the possibility of combining phonemes to form complex onsets and codas means that there is naturally a greater variation in bigram frequencies in English, making PP a stronger predictor of English word-likeness, compared to ND.

While both PP and ND are sub-lexical factors of interest in the context of NWR, to delimit the scope of this study, ND of Syllables was chosen as the target measure to examine sub-lexical effects on NWR performance in Cantonese-speaking children, given that ND has been suggested to be a stronger cue to word-likeness in Cantonese.

4.1.7 The Present Study

As recent studies (such as Szewczyk et al., 2018) suggest that sub-lexical representations are fundamental to the accurate repetition of nonwords, this study aims to examine whether NWR performance in Cantonese-speaking children is also affected by sub-lexical features of nonword stimuli.

Furthermore, this study also examines whether sub-lexicality in Cantonese nonword stimuli affects the degree of TD/DLD group differentiation, similarly to the effects of nonword lexicality on TD/DLD group differentiation, as demonstrated in Study I and Study II of this thesis. To the extent that sub-lexical knowledge supports NWR, TD children are expected to have a larger advantage on an increase in sub-lexicality of nonword stimuli (such as nonwords containing attested CV combinations,

relative to those that do not; and nonwords with higher ND of Syllables), relative to children with DLD, given their stronger sub-lexical representations that could be used to support the use of redintegration strategies in NWR.

Taking into consideration both previous findings on sub-lexical factors in Cantonese NWR, and the unique features of Cantonese lexical phonology, where ND has been suggested to be important in Cantonese word-likeness judgements (Kirby & Yu, 2007), this study will examine both sub-lexical factors of CV combination attestedness and ND of Syllables.

4.1.7.1 Research Questions. Three research questions (RQs) are addressed in Study III:

RQ1: Extending the analysis of CV combination attestedness to all sets of Cantonese nonword stimuli (as Study I only examined CV combination attestedness in Low-Lexicality nonwords), does CV combination attestedness predict NWR performance in Cantonese-speaking children and affect TD/DLD group differentiation?

RQ2: Does ND of Syllables in Cantonese nonword stimuli predict NWR performance and affect TD/DLD group differentiation in Cantonese-speaking children?

RQ3: Compared to CV attestedness, is ND of Syllables a stronger sub-lexical predictor of NWR performance in Cantonese-speaking children?

4.2 Methods

4.2.1 Participants

Data from the same thirty-eight predominantly monolingual Cantonese-speaking children (19 MonDLD, 19 MonTD), who were included in the participant sample of Study II in this thesis, were analysed in this study. Information on demographics, language background and diagnostic criteria for the participants have already been described in Chapter III, Section 3.2.1. This study was carried out in accordance with the recommendations of the Human Subjects Ethics Sub-committee at the Hong Kong Polytechnic University (reference number: HSEARS20161230004). All parents gave written consent for their children to participate in the study.

4.2.2 Materials

Screening/diagnostic tests used in this study for confirming children's clinical

status included the Hong Kong Cantonese Oral Language Assessment Scale (described in Chapter II, Section 2.2.2.1), Pure Tone Audiometry Hearing Screening Test (described in Chapter II, Section 2.2.2.2), and Raven's Progressive Matrices (described in Chapter II, Section 2.2.2.3).

4.2.2.1 Nonword Repetition Stimuli. Four sets of Cantonese nonword stimuli, also included in Study I and Study II of this thesis, were used in this study. For details on High-Lexicality, High-Lexicality-Vowel-Matched, and Low-Lexicality nonwords, see Chapter II, Section 2.2.2.4. For details on Cantonese CL-NWR nonwords, see Chapter III, Section 3.2.2.2.1). The following section will elaborate on the calculation of ND of Syllables for the nonword stimuli.

4.2.2.1.1 Neighbourhood Density Calculations. A computer program was developed on Google Colaboratory to compute the ND values for each constituent syllable in all sets of nonwords. ND of syllable was computed as the number of lexical items with a one phoneme difference (i.e. neighbours) from the target syllable (i.e. each constituent syllable in the NWR stimuli), weighted by the frequency of occurrence per million tokens for each neighbouring lexical item.

A one phoneme difference can take forms of addition, substitution or deletion of a phoneme – for instance, the target syllable /ka:3/ (meaning price) would have neighbours including /ka:k3/ (addition; meaning grid), /p^ha:3/ (substitution; meaning afraid), and /a:3/ (deletion; meaning Asian). Because Cantonese uses contrastive lexical tones, a word could also constitute a neighbour if it only has a tonal difference with the target (e.g. /ka:1/, meaning home). Given that each Cantonese syllable carries one tone, it is impossible to apply an addition or deletion to the tone of the target syllable. All Cantonese syllables have obligatory elements of a vowel and a tone, while optional elements include an initial consonant, an additional vowel (i.e. having a diphthong rather than a vowel), or a final consonant – consonant clusters do not exist in Cantonese (Matthews & Yip, 2011). Therefore, Cantonese syllables take the following structures: (C)V(V) or (C)V(C). For CVC syllables, Cantonese does not permit the addition of any other consonants or vowels, hence only the substitution or deletion of a phoneme, or tone substitutions, would be considered when identifying neighbours for target syllables with a CVC structure.

Neighbours of each target are identified from Cifu (Lai & Winterstein, 2020),

which is a lexical database of Hong Kong Cantonese with phonological, orthographical, frequency of occurrence, and lexical neighbourhood information for each entry. Lexical entries and frequency information in Cifu were drawn from several existing corpora on Hong Kong Cantonese, which target four different genres of discourse. The first genre is adult-spoken Cantonese, which incorporates items from HKCanCor (Luke & Wong, 2015), HKCAC (Leung & Law, 2001), CantoMap (Lai & Winterstein, 2019). The second and third genres are child-spoken Cantonese and Cantonese child-directed speech respectively, which draw items from HKU-70 (Fletcher et al., 2000) and the Lee/Wong/Leung Corpus (Lee et al., 1994). The fourth genre targets written discourse. As there are no published corpora on written Cantonese, lexical items were identified from an online platform for digital literature (website: https://www.shikoto.com/), which hosts amateur novels on various themes, where stories were written in a style that is representative of the written text that Hong Kong Cantonese speakers are typically exposed to. For the purpose of this study, however, as the digital literature on this online platform targets young adults, the frequency information computed from this digital literature source was deemed to be unrepresentative of the written language that young Cantonese-speaking children would be exposed to on a daily basis. For the written genre, frequency information was instead obtained from the Hong Kong Corpus of Primary School Chinese (M.-T. Leung & Lee, 2002), which documents the frequency of occurrence of lexical items in Hong Kong primary school level textbooks.

Taken together, when computing the ND for each constituent nonword syllable, the computer program first identified all lexical neighbours that were within one phoneme difference to the target syllable, using the phonetic transcriptions of lexical entries within Cifu.

Then, a weighting was applied to each identified lexical neighbour, based on its frequency of occurrence information across the four different genres. The weighting was calculated as follows: for each lexical neighbour, its frequency of occurrence per million tokens was first averaged across the three spoken genres (i.e. adult-spoken Cantonese, child-spoken Cantonese, and Cantonese child-directed speech), and then added to the frequency of occurrence per million tokens in the written genre (i.e. the primary school textbook corpus). The reason behind averaging the frequency of occurrence across all three spoken genres is that all child-spoken and child-directed speech corpora included in Cifu featured only productions by preschool aged children and their caregivers. Given that the present study targeted 8 to 11-year-old children,

using either child-to-adult or adult-to-adult corpora could heavily underestimate or overestimate the linguistic repertoire of upper elementary grade students. Without alternative available corpora, the average frequency of occurrence across the three spoken genres was taken as an estimate of the linguistic experience in the auditory modality of an 8 to 11-year-old child participating in this study. The frequencies of occurrence in the spoken and written modalities were then summed up, to represent the linguistic experience of an upper elementary school grade student in both the spoken and written modalities.

Finally, the computer program summed up the calculated weightings (i.e. the averaged and summed frequency of occurrence per million tokens across the four genres) for every identified lexical neighbour to derive the ND values for each target syllable. Under this method of calculating ND, the NDs of the Syllables used in the current study ranged from 0 to 207266.17.

4.2.3 Procedures

The procedures for administering the NWR task have been described in Chapter II, Section 2.2.3.

4.2.3.1 Scoring. Responses were audio recorded and transcribed. Children's responses were scored on accuracy at syllable level, as the present study mainly targeted syllable-level predictors for NWR performance (i.e. ND and CV combination attestedness).

4.2.4 Inter-Rater Reliability

Information on inter-rater reliability of NWR scores at syllable-level has been reported in Chapter III, Section 3.2.4.1.

4.2.5 Data Analysis

NWR scores at syllable level were analysed with Logistics Mixed Effect Models, using the lme4 package (Bates & Maechler, 2010) in R (version 4.1.3, R Core Development Team, 2021).

Three RQs were asked in this study, concerning whether CV combination attestedness and ND of Syllables in each of the four sets of Cantonese nonword stimuli

predict NWR performance and affect TD/DLD group differentiation in Cantonese-speaking children, as well as which sub-lexical factor, between the two, is a stronger predictor of NWR performance. All RQs would be addressed with one statistical model, Model 7. The dependent variable of the model was NWR Accuracy at syllable level (scored as a categorical variable; correct vs. incorrect), while Participants and Nonword Items were added as random effects. For fixed effects, Group (DLD vs. TD), CV combination attestedness (IN vs. OUT), ND of Syllables (as a continuous numeric variable) and the interaction between CV combination attestedness and Group were added to the model (the interaction between ND of Syllables and Group was found to be not significant when factors were added to the model in a stepwise, forward selection approach, thus it was not included in the final model). In addition, fixed effects of Nonword Set (High-Lexicality vs. High-Lexicality-Vowel-Matched vs. Low-Lexicality vs. CL-NWR), Length (in number of syllables, as a continuous numeric variable), and Syllable Complexity (CV vs. CVC) were also added to the model, in order to examine whether the effects of the two sub-lexical factors, CV combination attestedness and ND of Syllables, predict NWR performance above and beyond the other nonword-related factors known to affect NWR performance. As these effects of these factors are not of primary interest in this study, only results on Group, CV combination attestedness, and ND of Syllables will be presented in the following.

4.3 Results

4.3.1 Model 7: Effects of Sub-lexical factors on NWR Accuracy and Group Differentiation

The specification of Model 7 was as follows: NWRaccuracy_syllable ~ Group + CV_attestedness + ND of Syllables + CV_attestedness*Group + Nonword Set + Length + Syllable_Structure + (1|participant) + (1|Nonword_Item).

The results of Model 7 are shown in Table 10. At syllable level of NWR accuracy, there was a significant main effect of Group (p <.001), where TD children repeated syllables within nonwords more accurately than the DLD group. Regarding the two sub-lexical factors of interest, the main effects of both CV combination attestedness (p =.02) and ND of Syllables (p <.001) were statistically significant; a comparison of the two p-values also indicated that ND of Syllables was a stronger predictor of NWR accuracy at syllable level, relative to CV attestedness. The interaction between CV combination attestedness and Group was also statistically significant (p <.001).

Table 10. *Result of Model 7 on sub-lexical effects on NWR accuracy at syllable level*

Fixed Effect	β	*SE*	*z*	*p*
(Intercept)	0.32	1.17	0.27	.79
Group (TD)	1.23	0.22	5.51	<.001
CV Combination Attestedness (OUT)	1.48	0.66	2.25	.02
ND of Syllables	0.04	0.008	5.36	<.001
Group : CV Combination Attestedness (OUT)	-0.68	0.13	-5.16	<.001
Nonword Set (CL-NWR)	0.64	0.26	2.45	.01
Lexicality (Low-Lexicality)	-0.60	0.21	-2.92	.003
Lexicality (High-Lexicality)	-0.02	0.21	-0.10	.91
Length	-0.53	0.07	-7.58	<.001
Syllable Structure (CVC)	-0.36	0.17	-2.09	.04
Random Effects	Variance	*SD*		
Item	0.38	0.62		
Participant	0.42	0.65		

To assist the interpretation of this interaction between CV combination attestedness and Group, predicted probabilities of NWR accuracy at syllable level (with 95% confidence intervals) were plotted for IN and OUT syllables, for each group of children separately (see Figure 5). Post-hoc pairwise comparisons were also conducted, revealing that while both DLD and TD groups performed significantly better on IN syllables compared to OUT syllables (both *Ps* <.0001), only IN syllables differentiated between the two groups (p <.0001), whereas OUT syllables did not differentiate between the two groups (p =.10). Figure 5 shows that while within-group variability in NWR accuracy was relatively small for both groups on IN syllables, there was a much wider 95% CI on NWR accuracy of OUT syllables, for both TD and DLD groups. As a result, there was substantial overlap of the performance on OUT syllables between the two groups of children, thus OUT syllables were not able to differentiate between TD and DLD groups.

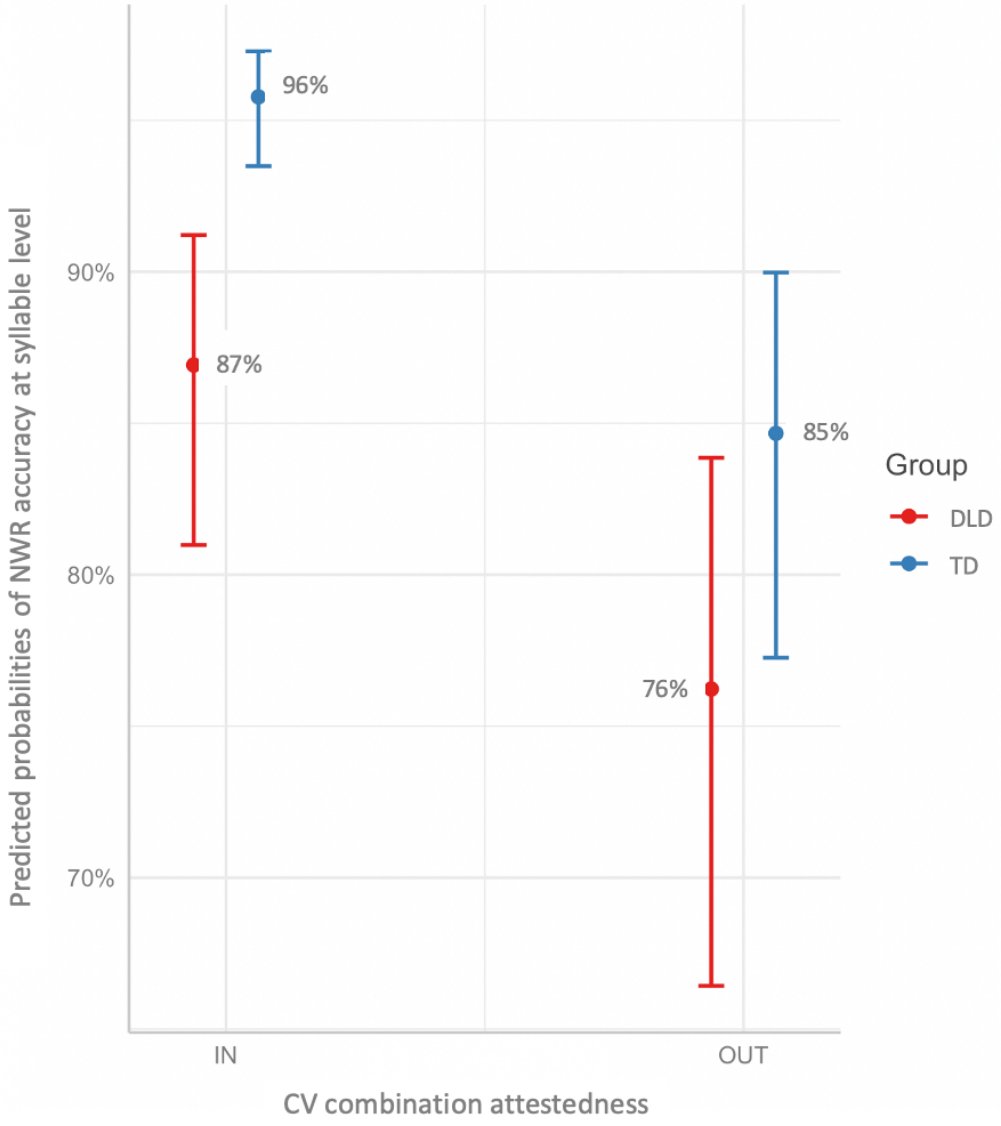

Figure 5. *Predicted probability of NWR accuracy at syllable level (with 95% confidence intervals) on IN and OUT syllables within all nonword sets, in the DLD vs. TD groups.*

4.4 Discussion

Study III examined the effects of sub-lexical factors in NWR performance and TD/DLD group differentiation in Cantonese-speaking children. Specifically, this study addressed three research questions: 1) Whether CV combination attestedness predicts NWR performance and affect TD/DLD group differentiation, when the analysis of IN vs. OUT syllables are extended to all four sets of Cantonese nonword stimuli; 2)

whether ND of Syllables predicted NWR accuracy at a syllable level and affected TD/DLD group differentiation in Cantonese-speaking children; 3) whether CV combination attestedness or ND was a stronger predictor of NWR performance in Cantonese-speaking children.

This study provided further evidence that sub-lexical factors do contribute to children's NWR performance, even when lexicality effects were considered. This study is also the first to document that ND of Syllables is predictive of NWR accuracy, at least in Cantonese-speaking children.

4.4.1 Sub-lexical Predictors of NWR Accuracy at Syllable Level

Like findings reported on other languages (e.g., Szewczyk et al., 2018), this study found that NWR performance in Cantonese-speaking children is also affected by sub-lexical factors of nonwords, on multiple measures.

4.4.1.1 CV Combination Attestedness. Consistent with the results from Stokes et al. (2006) and the first findings from the Study I of this thesis, CV combination attestedness was found to be a significant factor in NWR accuracy, where better NWR performance was seen on IN syllables (i.e. those with attested CV combinations) compared to OUT syllables (i.e. those with unattested CV combinations), when the analyses were extended to High-Lexicality, High-Lexicality-Vowel-Matched, and CL-NWR nonwords.

Moreover, when all nonword stimuli were considered, CV combination attestedness affected TD/DLD group differentiation – only IN syllables captured significant TD/DLD group differences in NWR performance, but not OUT syllables, suggesting that the TD group of children may be better able to draw on their long-term sub-lexical representations to support the accurate repetition of nonwords, when the nonwords had sub-parts (e.g. CV combinations) that resembled elements in the ambient language, as in the case of IN syllables. Contrastively, OUT syllables consisted of unattested CV combinations, which do not allow either groups of children to draw on their sub-lexical knowledge to support NWR, and as such the TD group may not be able to benefit from their stronger sub-lexical representations to support the repetition of OUT syllables.

Furthermore, the results indicated that there was a large variability in performance on OUT syllables by both groups of children, indicating that within TD and within DLD

groups, children responded differently to the challenge of repeating unattested syllables, with some being more heavily impacted by the lack of existing phonological/sub-lexical representations to support redintegration processes during repetition of syllables with unattested CV combinations, while others being able to rely on stronger phonological processing skills and/or larger working memory capacities to compensate for the challenge of repeating syllables with unattested CV combinations. Large variability in NWR performance by both groups of children suggest that OUT syllables, i.e. nonwords with syllables containing only unattested CV combinations, are not ideal for capturing TD/DLD group differences.

Overall, the present findings clearly demonstrate that sub-lexicality of nonwords, in terms of CV combination attestedness, does affect NWR performance in children acquiring Cantonese, reflecting that children draw on sub-lexical representations during NWR. Moreover, such effect is present even when nonword lexicality is considered, based on the findings that both lexicality and CV combination attestedness were found to be significant factors in the statistical model, suggesting that children draw on both streams of information to support the repetition of nonwords.

4.4.1.2 ND of Syllables. In addition to CV combination attestedness, this study is also first to report findings that the novel sub-lexical measure, ND of Syllables, affects NWR performance in Cantonese-speaking children, where syllables with higher NDs were repeated at a higher level of accuracy compared to syllables with lower NDs.

In Cantonese, as ND has been suggested to be a stronger cue to word-likeness than other measures (such as PP) that quantify phonological similarity of nonwords to real words (Kirby & Yu, 2007), this study opted to examine sub-lexical effects on NWR through ND of Syllables, to test whether sub-lexical features that are phonologically similar to those in real words of a language support more successful repetition of nonwords. The findings confirm this to be true, and found that ND of Syllables significantly contributed to the model, even when other factors including lexicality, and CV combination attestedness were considered. Furthermore, ND of Syllables was also found to be a stronger predictor of NWR performance in Cantonese-speaking children, relative to CV combination attestedness. This is unsurprising, as ND of Syllables, being a continuous rather than categorical measure, is a more detailed measure of sub-lexicality than CV combination attestedness, and could capture more nuanced differences in the levels of sub-lexicality in nonword stimuli, and how that affects

children's NWR performance.

This study did not find a significant interaction between ND and participant group, thus it appears that both Cantonese-speaking TD children and those with DLD responded similarly to this sub-lexical measure in nonword stimuli – the denser the phonological neighbourhood of a target syllable (i.e. the more phonological neighbours it has), the easier it is for children to draw on sub-lexical representations to support NWR, and the higher scores they are able to achieve. Although, it is also possible that a significant interaction may emerge with a larger sample size.

Finally, it is important to acknowledge that it remains unknown, whether the effects of ND of Syllables will also be found in other languages such as English, where ND may not play such a strong role in the word-likeness of syllables and nonwords. Based on the current positive findings, ND of nonword syllables do appear to contribute to NWR performance, and together with the findings on CV combination attestedness, this study provides further cross-linguistic evidence from an understudied language like Cantonese, that sub-lexical representations are fundamental to the successful repetition of nonwords.

4.4.2 Limitations and Future Directions

To test whether the present findings on the relationship between ND of Syllables and NWR performance could be generalized to other languages, future studies can re-examine ND of Syllables as a factor in NWR performance in children acquiring other languages. Testing so could generate further insights on whether the novel finding that ND of Syllables, as a sub-lexical factor, influences NWR performance is specific to Cantonese and typologically similar languages, where ND is a strong cue to word-likeness; or whether ND truly plays a stronger role in NWR than previously suggested, due to limitations in previous methods used to compute the ND of stimuli, which typically focused on entire nonword items.

Furthermore, while it was not within this study's objectives to test PP as a sub-lexical factor in Cantonese NWR, given that PP is also an important sub-lexical factor of interest, future studies could examine the effects of PP on NWR performance in Cantonese-speaking children, and compare PP and ND of Syllables as predictors of NWR performance.

4.5 Conclusion

The results of Study III demonstrated that both CV combination attestedness and ND of Syllables affect Cantonese-speaking children's NWR performance, in addition to lexicality effects, providing further supporting evidence to cross-linguistic findings that sub-lexical representations are fundamental to the successful repetition of nonwords. While CV combination attestedness was found to affect TD/DLD group differentiation, this study found no evidence that ND of Syllables influenced the size of TD/DLD group differences in Cantonese-speaking children. Overall, Study III demonstrated that children draw on multiple-streams of information, including both lexical and sub-lexical representations, to support the use of redintegration strategies during NWR.

4.6 Summary of Chapter IV

With a recent study suggesting that sub-lexical representations may have a facilitative effect on NWR, above and beyond that of lexical representations (Szewczyk et al., 2018), Study III examined whether sub-lexicality in Cantonese NWR stimuli affected Cantonese-speaking children's NWR performance and TD/DLD group differentiation. Data on NWR accuracy of 19 DLD-TD pairs of monolingual speaking children were analysed with two measures of sub-lexicality – CV combination attestedness, as some positive first findings have previously been reported on this sub-lexical measure (Stokes et al., 2006); and ND of Syllables, as ND has been suggested as a strong cue to Cantonese word-likeness. Both measures significantly predicted NWR performance in Cantonese-speaking children, with ND of Syllables being a stronger predictor. CV combination attestedness interacted with participant group, where only syllables with attested CV combinations (i.e. IN syllables) captured significant TD/DLD group differences. There was no evidence that ND of Syllables affected the degree of TD/DLD group differentiation in these Cantonese-speaking children. These findings provide further support to the growing evidence suggesting that sub-lexical representations play a crucial role in NWR, either alongside or above and beyond the influence of lexical representations. Future studies can examine whether ND of Syllables predict NWR performance in children acquiring other languages, to confirm whether the findings on the novel sub-lexical measure, ND of Syllables, can be generalised cross-linguistically; and also examine PP as a predictor of NWR performance in Cantonese-speaking children, in addition to the two examined measures.

CHAPTER V – Discussion

While consistent evidence from cross-linguistic studies has demonstrated that NWR has the potential to identify children with DLD in both monolingual and bilingual children, Cantonese has been the sole exception, where previous findings diverged from cross-linguistic findings. This thesis is first to revisit this topic of NWR in Cantonese since Stokes et al. (2006) and Leung (2010), and is first to present evidence that NWR can capture significant group differences between Cantonese-speaking children with and DLD and their TD peers, in both monolingual TD vs. DLD and bilingual L2-TD vs. monolingual DLD comparisons. NWR performance in Cantonese-speaking children was also found to be affected by nonword characteristics known to affect NWR accuracy in children acquiring other languages, particularly nonword lexicality and sub-lexicality. The present findings, therefore, suggest that Cantonese is not a true cross-linguistic exception in NWR, despite it being typologically distinct from previously examined languages. Overall, findings arising from this thesis add to the narrative that NWR can be an informative assessment tool for DLD, for both monolingual and bilingual children, and that NWR deserves to be included in language assessment test batteries alongside complementary tools for diagnostic purposes.

In the following sections, main findings from the three studies will be summarized. The empirical and clinical significance of the findings and future directions will also be highlighted.

5.1 Summary of Study I

Study I introduced three sets of novel Cantonese NWR stimuli – High-Lexicality, High-Lexicality-Vowel-Matched, and Low-Lexicality nonwords – and evaluated their ability to capture significant TD/DLD group differences between Cantonese-speaking monolingual children with DLD and their TD peers. Low-lexicality nonwords were further analysed on sub-lexicality, in terms of CV combination attestedness (IN syllables vs. OUT syllables), to examine whether sub-lexicality in nonwords also affected the degree of TD/DLD group differentiation.

Unlike the nonword stimuli used by Stokes et al. (2006), which were constructed with strictly non-morphemic syllables and had low lexicality levels; and stimuli used by Leung (2010), which were overly long in terms of nonword length (nonwords reached nine syllables in length), the newly designed NWR stimuli took into account

factors known to affect NWR performance and group differentiation from more recent, cross-linguistic studies. The three sets of nonwords differed in lexicality levels, with High-Lexicality and High-Lexicality-Vowel-Matched nonwords sharing the same high levels of lexicality, where 100% of their constituent syllables were morphemic with no meaning when combined; and Low-Lexicality nonwords having low lexicality, where 0% of their constituent syllables were morphemic. Sub-lexicality was manipulated within Low-lexicality nonwords, where 50% of their constituent syllables consisted of CV combinations attested in Cantonese, and the remaining syllables consisted of unattested CV combinations.

While all three sets of nonwords captured significant TD/DLD group differences, effect sizes suggested that both sets of High-Lexicality nonwords captured greater TD/DLD group differences than Low-Lexicality nonwords. Concerning sub-lexicality, while both types of syllables (those containing attested CV combinations and those that did not) captured significant TD/DLD group differences, the magnitude of group difference was also larger on syllables with attested CV combinations, than those with unattested CV combinations.

There are two implications of these findings. First, the findings suggest that Cantonese is not truly a cross-linguistic exception, despite speculations from Stokes et al. (2006) that Cantonese nonword stimuli may not be as taxing on working memory, as nonwords created under the phonotactic properties of Cantonese have much lower segmental and suprasegmental complexity, compared to nonwords created on other languages that allow more complex structures, such as consonant clusters and variable stress patterns. The present findings that Cantonese-speaking children with DLD still performed worse than their TD peers even in repeating nonwords with simple CV structures suggest that the previous lack of significant findings may be more heavily influenced by methodological factors rather than typological differences between Cantonese and other examined languages.

Second, the findings that larger TD/DLD group differences are captured by nonwords with higher lexicality levels suggest that long-term vocabulary knowledge is crucial to the accurate repetition of nonwords. TD children are better able to draw on stronger lexical representations to support redintegration during NWR, compared to children with DLD, allowing them to perform at a higher level of NWR accuracy when nonwords had sufficient resemblance to real lexical items in the ambient language – this also seems to be the mechanism behind how NWR differentiates between children

with and without DLD. Similarly, the findings that CV combination attestedness affected NWR accuracy and nonword syllables with attested CV combinations captured larger differences between TD and DLD groups provide first evidence that sub-lexical knowledge also supports the repetition of nonwords in Cantonese-speaking children, alongside lexical knowledge. Therefore, it is crucial that both lexical and sub-lexical factors are considered in the design of NWR stimuli for generating TD/DLD group differences, with reference to the ambient language that children are exposed to.

5.2 Summary of Study II

Study II explored the potential of Cantonese NWR stimuli to capture TD/DLD group differences in NWR performance of Cantonese-speaking children, even when bilingual L2 learners of Cantonese are taken into account. Specifically, it examined whether L1-Urdu-L2-Cantonese-speaking TD children were disadvantaged by Cantonese NWR tests, on the three sets of Cantonese-language-specific stimuli examined in Study I, and on an additional set of quasi-universal, Cantonese-adapted CL-NWR nonwords, because L2-TD children are at risk of being misclassified as having language disorders in NWR, due to their reduced language exposure and knowledge in the testing language, associated with bilingual language acquisition.

Study II is the first to present evidence suggesting that at least certain sets of Cantonese NWR stimuli have potential to differentiate between language impairment (in children with DLD) and bilingualism (in L2-TD children), in terms of yielding significant group differences between L2-TD children and MonDLD children, while avoiding group differences between L2-TD and MonTD children. These findings provide yet another piece of evidence that Cantonese is not a true cross-linguistic exception in NWR, and NWR has potential to be developed into a clinically informative assessment tool for DLD, even in the context of multilingualism, where diagnosis of DLD is complicated by overlapping presentations of language difficulties in language impaired-children and language-unimpaired children with reduced language exposure.

Furthermore, the findings also suggest that Cantonese-language-specific, High-Lexicality nonwords, when scored at whole-nonword level; and quasi-universal, Cantonese-adapted CL-NWR nonwords, when scored at syllable level, did not disadvantage bilingual Urdu-Cantonese-speaking TD children – no significant differences were captured between MonTD and L2-TD groups. This implies that the two sets of nonwords allowed L2-TD children to draw on their lexical and sub-lexical

knowledge across one or both of their languages, Urdu and Cantonese in this case, to support the use of redintegration strategies during NWR.

In particular, there was a close to complete overlap in NWR performance by MonTD and bilingual L2-TD children on CL-NWR nonwords, which demonstrated that the CL-NWR test (Chiat, 2015) can achieve its intended goals of minimising the performance gap between TD children from monolingual and bilingual groups, through including only simple and cross-linguistically frequent consonants and vowels in its stimuli set – this is true even when the test is adapted to a language, Cantonese, that has previously been demonstrated to be an exception in NWR. Therefore, quasi-universal NWR tests appeared to be advantageous for assessing bilingual children. As for Cantonese-language-specific, High-Lexicality nonwords, L1-Urdu-L2-Cantonese-speaking TD children were also not disadvantaged relative to monolingual Cantonese-speaking TD children, likely because the majority of consonants and vowels included in this stimulus set were coincidentally within the Urdu phonemic inventory, with the exception of one unreleased final consonant, /-k/, that is only present in Cantonese. Whether this set of Cantonese-specific, High-Lexicality nonwords remain capable of minimising group differences in monolingual and L2-TD children who acquire first languages other than Urdu will likely depend on whether their constituent vowels and consonants are also present in the inventory of the other languages in question. Together, these findings imply that, language-specificity, or language-(non)-specificity of nonword stimuli, are a particularly important factor to be considered in the design of nonword stimuli for assessing bilingual children.

Furthermore, Study II demonstrated that different scoring methods of NWR result in different levels and patterns of group differentiation – a more detailed level of scoring (scoring at syllable level, rather than whole-nonword level) appears to be better suited for Cantonese adapted CL-NWR nonwords, as only syllable-level scoring, but not whole-nonword scoring, captured significant differences between monolingual DLD and monolingual TD / L2-TD groups. This is because scoring at syllable level allows for more subtle differences in NWR performance between groups to be captured, such as whether monolingual TD and L2-TD groups repeat more nonword components correctly than the monolingual DLD group, despite both groups performing at similar levels of accuracy on entire nonwords. The present findings are in line with previous suggestions that scoring methods do influence group differentiation in NWR, and therefore, the suitability of different scoring methods on a particular stimulus set should

be considered when the NWR test is being developed as an assessment tool for DLD.

5.3 Summary of Study III

As Study I and Study II have thoroughly examined the effects of lexicality on NWR performance and TD/DLD group differentiation in Cantonese-speaking children, Study III examines whether sub-lexicality of nonword stimuli also affected NWR performance and TD/DLD group differentiation in Cantonese-speaking children, given that both lexical and sub-lexical representations have been suggested to be fundamental to NWR. Two measures of sub-lexicality were chosen, which were CV combination attestedness, as Study I provided positive first findings on this measure with Low-Lexicality nonwords; and ND of Syllables, a newly proposed measure of sub-lexicality, based on suggestions that ND is a stronger cue to word-likeness in Cantonese, compared to the more frequently examined sub-lexical factor, phonotactic probability (Kirby & Yu, 2007).

The data indicated that both CV combination attestedness and ND of Syllables significantly predicted the accuracy of NWR at a syllable level, even when the effects of nonword lexicality are also taken into account in the statistical model. When the analysis of CV combination attestedness was extended beyond Low-Lexicality nonwords to all sets of Cantonese NWR stimuli, it was also found to interact with participant group, where only syllables containing attested CV combinations captured significant group differences between TD and DLD groups, but not syllables containing unattested CV combinations. These findings demonstrated that, similar to how lexicality affects TD and DLD groups differently, TD children may have stronger sub-lexical representations to draw on when repeating newly heard nonwords, thus they have a larger advantage when nonwords increase in levels of sub-lexicality, compared to children with DLD, who may have weaker sub-lexical representations that are less effective in providing support for successful NWR, limiting the advantage offered by nonwords with increased levels of sub-lexicality. While ND of Syllables was not found to affect the magnitude of TD/DLD group differences in these children, this may be due to the relatively small sample size of the study. Overall, these findings provide further support to the growing evidence pointing towards the primacy of sub-lexical representations in nonword repetition, either alongside or above and beyond effects of lexicality.

As this study was the first to examine the novel sub-lexical measure of ND of

Syllables, it is currently unknown whether this factor will also predict NWR performance in children acquiring other languages, or if this effect remains true only in languages that are typologically similar to Cantonese, with similar phonotactic properties. ND of Syllables will need to be examined in NWR by children acquiring diverse languages in future studies, and future Cantonese NWR studies should also examine the effects of phonotactic probability, to further the understanding of whether different types of sub-lexical information in nonword stimuli affected NWR performance and TD/DLD group differentiation in similar or different ways, and whether such effect also depends on the unique phonotactic properties of a language being examined. Based on the findings of Study III, ND, when computed as a measure of sub-lexicality, appears to be more relevant to children's NWR performance than previously demonstrated.

5.4 Clinical Significance

The overall findings presented in this thesis clearly indicate that Cantonese is not a true cross-linguistic exception in NWR. The present group-level findings, that Cantonese-speaking MonTD and bilingual L2-TD children perform similarly in NWR, and both groups perform above MonDLD children, support the development of Cantonese NWR into clinically informative tools as part of assessment toolkits used for assessing Cantonese-speaking children's speech and language profiles, like recommendations made for other linguistic populations (Schwob et al., 2021).

This is especially true given the ease and quickness of task administration for NWR tasks, making it ideal as an early screening tool when used together with other screening tools like parental questionnaires (Ortiz, 2021), and the present findings encourage continuous development of NWR tasks for the assessment of Cantonese-speaking children.

Here, a note of caution has to be made that the current findings on NWR do not support its use as the sole assessment tool for diagnosing DLD, given that children with DLD are a heterogeneous population where not all children will show impairments in NWR – this is also reflected by the variability in NWR performance demonstrated within each group of children included in this study. As future studies continue to assess the suitability of NWR for assessing DLD in Cantonese-speaking children, it is expected that such heterogeneity will remain, or even increase with larger samples of children. With that said, studies in this thesis have laid important groundwork for future

studies to examine Cantonese NWR with larger sample sizes, in order to investigate diagnostic accuracy of Cantonese NWR. Moreover, regardless of the diagnostic accuracy of NWR for assessing DLD, NWR tasks can also be informative for clinicians, in terms of uncovering children's skills and deficits in areas such as vocabulary development, phonological awareness and working memory, which can inform clinical decisions on target areas of intervention.

Furthermore, the present findings also suggest that it is possible to assess bilingual ethnic minority children in Hong Kong, using NWR tasks in Cantonese, despite Cantonese typically being the weaker language of these children. Particularly, the quasi-universal, CL-NWR task and Cantonese language-specific, High-Lexicality nonwords have demonstrated potential in assessing bilingual children, although this does not replace the need for more comprehensive assessments in the child's stronger language(s). The data presented in this thesis is still encouraging as it takes the first steps towards developing much needed language assessment tools for bilingual children.

5.5 Limitations and Directions for Future Research

Although the findings presented in this thesis are very encouraging, the studies of this thesis have only taken the very first steps towards developing NWR tests into assessment tools for DLD in Cantonese-speaking children. In the following, suggestions for future directions in further studies on Cantonese NWR will be discussed.

First, one limitation across all three studies of this thesis is its relatively small sample size, with sixteen to nineteen children being included in each of the monolingual DLD, monolingual TD, and bilingual L2-TD groups. While such sample size is comparable to previous Cantonese NWR studies, and most effects that were expected to be significant were still registered (i.e. there is only limited risk of committing type II errors), it is still important for the results to be replicated in larger scale studies to confirm that the findings remain consistent when a wider population is considered, especially given the heterogeneity in children with DLD, as well as bilingual children.

Second, future work is needed to confirm whether significant group differences in NWR accuracy between Cantonese-speaking children with and without DLD carries over to younger children, as the lack of positive findings reported in Stokes et al. (2006) and Leung (2010) were reported on younger participant groups at pre-school age. Based on findings that significant TD/DLD group differences have been reported on children and adolescents across different ages, at least up to 15 years, it is expected that the

present pattern of results, particularly significant differences between TD and DLD groups, would remain generally consistent when younger Cantonese-speaking children are tested. With that said, replicating the present studies on a younger population is still warranted, given the eventual goal of developing Cantonese NWR into a clinically accurate assessment tool, as the development of clinical screening tools should always strive to allow for early identification of developmental disorders.

Third, while this thesis has also reported findings pointing to the potential for NWR in Cantonese to be developed into accurate assessment tools for bilingual Cantonese-speaking children, the present studies have only taken the first steps in this direction, by demonstrating that a particular group of L2-TD children were not disadvantaged relative to monolingual TD children in Cantonese NWR. The suitability of Cantonese NWR tasks for assessing bilingual children will need to be tested on other bilingual populations, such as bilingual L2 Cantonese-speaking TD children acquiring first languages other than Urdu, or bilingual L1-TD children acquiring Cantonese as a heritage and minority language, under heavy influence of another language that is considered a majority language in locations outside of Hong Kong. Furthermore, it was not possible for this study to include a bilingual L2-DLD group, due to the lack of available assessment tools for confirming the presence of language disorder in L1-Urdu-L2-Cantonese-speaking children. As assessment tools for L2-Cantonese-speaking children are currently under development, future studies will be in a better position to examine NWR performance in an L2-DLD group, and investigate group differentiation among monolingual TD, monolingual DLD, L2-TD and L2-DLD children in Cantonese NWR.

Finally, this thesis only addressed the differentiation of children with DLD and TD at a group level, but for Cantonese NWR to be developed into a clinically informative assessment tool, it must also be clinically accurate when classifying children at an individual level. As future studies aim to increase the number and groups (e.g. younger age groups and other bilingual Cantonese groups) of children examined in NWR, they will be in a better position to examine sensitivity and specificity of Cantonese NWR tasks. These future directions will further verify whether NWR tasks can be developed into clinically informative tools for DLD in Cantonese-speaking children.

CONCLUSIONS

Despite previous suggestions that Cantonese may be a cross-linguistic exception in NWR, due to its typological differences from other previously studied languages, this thesis presents evidence suggesting otherwise – that NWR is able to capture significant group differences between TD children and those with DLD, even in Cantonese, and therefore has the potential to be a cross-linguistically informative assessment tool for DLD. In addition, despite evidence suggesting that lexical and sub-lexical knowledge is crucial to the successful repetition of nonwords, bilingual L2-TD children with weaker knowledge of the testing language (but without a true language impairment) can still perform at a level of accuracy similar to that of monolingual TD children, at least on certain sets of nonword stimuli, suggesting that NWR has potential to be an appropriate and informative assessment tool for DLD, even in the context of multilingualism. The findings in this thesis also document that factors affecting NWR accuracy in children acquiring other languages affect children acquiring Cantonese in similar ways, particularly in terms of nonword lexicality and sub-lexicality. Overall, studies presented in this thesis bring Cantonese NWR research in line with cross-linguistic and international trends, and have established important groundwork for future studies to develop NWR into clinically informative assessment tools for DLD in Cantonese.

REFERENCES

Ambreen, S., & To, K. S. C. (2023). Review of the contemporary Urdu phonological system. Faculty of Education, The University of Hong Kong.

Archibald, L., M. (2008). The promise of nonword repetition as a clinical tool. *Canadian Journal of Speech-Language Pathology and Audiology*, *32*(1), 21–28.

Archibald, L. M. D., & Gathercole, S. E. (2006). Short-term and working memory in specific language impairment. *International Journal of Language & Communication Disorders*, *41*(6), 675–693. https://doi.org/10.1080/13682820500442602

Armon-Lotem, S. (2012). Introduction: Bilingual children with SLI – the nature of the problem. *Bilingualism: Language and Cognition*, *15*(1), 1–4. https://doi.org/10.1017/S1366728911000599

Armon-Lotem, S. (2018). SLI in bilingual development: How do we approach assessment? In A. Bar-On, D. D. Diskin Ravid, & E. Dattner (Eds.), *Handbook of communication disorders* (pp. 617–641). De Gruyter Mouton.

Armon-Lotem, S., & Chiat, S. (2012). How do sequential bilingual children perform on non-word repetition tasks? In A. K. Biller, E. Y. Chung, & A. E. Kimball (Eds.), *Proceedings of the 36th Annual Boston University Conference on Language Development* (pp. 53–62). Cascadilla Press.

Bailey, T. M., & Hahn, U. (2001). Determinants of Wordlikeness: Phonotactics or Lexical Neighborhoods? *Journal of Memory and Language*, *44*(4), 568–591. https://doi.org/10.1006/jmla.2000.2756

Bates, D., & Maechler, M. (2010). *lme4: Linear mixed-effects models using S4 classes* (R package version 0.999375-33) [Computer software]. http://CRAN.R-project.org/package=lme4

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology:* DE GRUYTER MOUTON. https://doi.org/10.1515/9783110823707

Bishop, D. V. M. (2014). Ten questions about terminology for children with unexplained language problems. *International Journal of Language &*

Communication Disorders, *49*(4), 381–415. https://doi.org/10.1111/1460-6984.12101

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*, *37*(4), 391–403. https://doi.org/10.1111/j.1469-7610.1996.tb01420.x

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., & Greenhalgh, T. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry*, *58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., & CATALISE consortium. (2016). CATALISE: A Multinational and Multidisciplinary Delphi Consensus Study. Identifying Language Impairments in Children. *PLOS ONE*, *11*(7), e0158753. https://doi.org/10.1371/journal.pone.0158753

Boerma, T., & Blom, E. (2021). Crosslinguistic nonword repetition and narrative performance over time: A longitudinal study on 5- to 8-year-old children with diverse language skills. In S. Armon-Lotem & K. K. Grohmann (Eds.), *Trends in Language Acquisition Research* (Vol. 29, pp. 302–328). John Benjamins Publishing Company. https://doi.org/10.1075/tilar.29.10boe

Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A Quasi-Universal Nonword Repetition Task as a Diagnostic Tool for Bilingual Children Learning Dutch as a Second Language. *Journal of Speech, Language, and Hearing Research*, *58*(6), 1747–1760. https://doi.org/10.1044/2015_JSLHR-L-15-0058

Bortolini, U., Arfé, B., Caselli, C. M., Degasperi, L., Deevy, P., & Leonard, L. B. (2006). Clinical markers for specific language impairment in Italian: The contribution of clitics and non-word repetition. *International Journal of Language & Communication Disorders*, *41*(6), 695–712. https://doi.org/10.1080/13682820600570831

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. https://doi.org/10.1006/jecp.1996.0042

Bowey, J. A. (1997). WhatDoesNonword Repetition Measure? A Reply to Gathercole and Baddeley. *Journal of Experimental Child Psychology*, *67*(2), 295–301. https://doi.org/10.1006/jecp.1997.2408

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. https://doi.org/10.1111/1469-7610.00726

Camilleri, B., & Law, J. (2007). Assessing children referred to speech and language therapy: Static and dynamic assessment of receptive vocabulary. *Advances in Speech Language Pathology*, *9*(4), 312–322. https://doi.org/10.1080/14417040701624474

Casalini, C., Brizzolara, D., Chilosi, A., Cipriani, P., Marcolini, S., Pecini, C., Roncoli, S., & Burani, C. (2007). Non-Word Repetition in Children with Specific Language Impairment: A Deficit in Phonological Working Memory or in Long-Term Verbal Knowledge? *Cortex*, *43*(6), 769–776. https://doi.org/10.1016/S0010-9452(08)70505-7

Chan, J. (1984). Raven's progressive matrices test in Hong Kong. *New Horizons: The Journal of Education, Hong Kong Teachers' Association*, *25*, 43–49.

Cheng, H.-C., Chen, H.-Y., Tsai, C.-L., Chen, Y.-J., & Cherng, R.-J. (2009). Comorbidity of motor and language impairments in preschool children of Taiwan. *Research in Developmental Disabilities*, *30*(5), 1054–1061. https://doi.org/10.1016/j.ridd.2009.02.008

Chi P. H. (2007). Te Ding Xing Yu Yan Zhang Ai Er Tong Yin Yun Duan Qi Ji Yi Neng Li Zhi Chu Tan [Phonological Short-Term Memory in Children with Specific Language Impairment]. *Bulletin of Special Education*, *32*(4), 19–45. https://doi.org/10.6172/BSE200712.3204002

Chiat, S. (2015). Non-Word Repetition. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing Multilingual Children—Disentangling Bilingualism from*

Language Impairment (pp. 125–150). Multilingual Matters. https://www.degruyter.com/document/doi/10.21832/9781783093137-008/html

Chiat, S., & Polišenská, K. (2016). A Framework for Crosslinguistic Nonword Repetition Tests: Effects of Bilingualism and Socioeconomic Status on Children's Performance. *Journal of Speech, Language, and Hearing Research*, *59*(5), 1179–1189. https://doi.org/10.1044/2016_JSLHR-L-15-0293

Chiat, S., & Roy, P. (2007). The Preschool Repetition Test: An Evaluation of Performance in Typically Developing and Clinically Referred Children. *Journal of Speech, Language, and Hearing Research : JSLHR*, *50*, 429–443. https://doi.org/10.1044/1092-4388(2007/030)

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. https://doi.org/10.1111/j.1469-7610.2004.00342.x

Coady, J. A., & Aslin, R. N. (2004). Young children's sensitivity to probabilistic phonotactics in the developing lexicon. *Journal of Experimental Child Psychology*, *89*(3), 183–213. https://doi.org/10.1016/j.jecp.2004.07.004

Coady, J., & Evans, J. (2008). Uses and interpretations of non-word repetition tasks in children with and without specific language impairment (SLI). *International Journal of Language & Communication Disorders / Royal College of Speech & Language Therapists*, *43*, 1–40. https://doi.org/10.1080/13682820601116485

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic Markers for Specific Language Impairment (SLI). *Journal of Child Psychology & Psychiatry & Allied Disciplines*, *42*(6), 741. https://doi.org/10.1111/1469-7610.00770

Conti-Ramsden, G., Durkin, K., Toseeb, U., Botting, N., & Pickles, A. (2018). Education and employment outcomes of young adults with a history of developmental language disorder. *International Journal of Language & Communication Disorders*, *53*(2), 237–255. https://doi.org/10.1111/1460-6984.12338

Curtis, P. R., Frey, J. R., Watson, C. D., Hampton, L. H., & Roberts, M. Y. (2018). Language Disorders and Problem Behaviors: A Meta-analysis. *Pediatrics*, *142*(2), e20173551. https://doi.org/10.1542/peds.2017-3551

de Bree, E., Wijnen, F., & Gerrits, E. (2010). Non-word repetition and literacy in Dutch children at-risk of dyslexia and children with SLI: Results of the follow-up study. *Dyslexia*, *16*(1), 36–44. https://doi.org/10.1002/dys.395

Dispaldro, M., Leonard, L. B., & Deevy, P. (2013). Real-Word and Nonword Repetition in Italian-Speaking Children With Specific Language Impairment: A Study of Diagnostic Accuracy. *Journal of Speech, Language, and Hearing Research*, *56*(1), 323–336. https://doi.org/10.1044/1092-4388(2012/11-0304)

Dollaghan, C. A., Biber, M. E., & Campbell, T. F. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics*, *16*(2), 211–222. https://doi.org/10.1017/S0142716400007098

Dollaghan, C., Biber, M., & Campbell, T. (1993). Constituent Syllable Effects in a Nonsense-Word Repetition Task. *Journal of Speech, Language, and Hearing Research*, *36*(5), 1051–1054. https://doi.org/10.1044/jshr.3605.1051

Dollaghan, C., & Campbell, T. F. (1998). Nonword Repetition and Child Language Impairment. *Journal of Speech, Language, and Hearing Research*, *41*(5), 1136–1146. https://doi.org/10.1044/jslhr.4105.1136

Ebbels, S. (2014a). Introducing the SLI debate: Editorial. *International Journal of Language & Communication Disorders*, *49*(4), 377–380. https://doi.org/10.1111/1460-6984.12119

Ebbels, S. (2014b). The SLI debate: Diagnostic criteria and terminology [Special issue]. *International Journal of Language & Communication Disorders*, *49*(4), 377–509.

Ebbels, S., Marić, N., Murphy, A., & Turner, G. (2014). Improving comprehension in adolescents with severe receptive language impairments: A randomized control trial of intervention for coordinating conjunctions: Improving comprehension in adolescents with severe receptive language impairments. *International Journal of Language & Communication Disorders*, *49*(1), 30–48. https://doi.org/10.1111/1460-6984.12047

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. https://doi.org/10.1017/S0142716400010079

Ferré, S., & Santos, C. dos. (2015). Comment évaluer la phonologie des enfants bilingues ? *Revue de Linguistique et de Didactique Des Langues*, *51*, 11–34. https://doi.org/10.4000/lidil.3678

Fletcher, P., Leung, S. C.-S., Stokes, S. F., & Weizman, Z. O. (2000). *HKU-70. 1.0.*

Forrest, C. L., Gibson, J. L., Halligan, S. L., & St Clair, M. C. (2020). A Cross-Lagged Analysis of Emotion Regulation, Peer Problems, and Emotional Problems in Children With and Without Early Language Difficulties: Evidence From the Millennium Cohort Study. *Journal of Speech, Language, and Hearing Research*, *63*(4), 1227–1239. https://doi.org/10.1044/2020_JSLHR-19-00188

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, *3*(11), 410–419. https://doi.org/10.1016/S1364-6613(99)01388-1

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. https://doi.org/10.1016/0749-596X(90)90004-J

Gathercole, S. E., & Baddeley, A. D. (1996). *The children's test of non-word repetition*. London 1996.

Graf-Estes, K., Evans, J. L., & Else-Quest, N. M. (2007). Differences in the Nonword Repetition Performance of Children With and Without Specific Language Impairment: A Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *50*(1), 177–195. https://doi.org/10.1044/1092-4388(2007/015)

Gray, S., Plante, E., Vance, R., & Henrichsen, M. (1999). The Diagnostic Accuracy of Four Vocabulary Tests Administered to Preschool-Age Children. *Language, Speech, and Hearing Services in Schools*, *30*(2), 196–206. https://doi.org/10.1044/0161-1461.3002.196

Guiberson, M., & Rodríguez, B. L. (2013). Classification Accuracy of Nonword Repetition When Used With Preschool-Age Spanish-Speaking Children. *Language, Speech, and Hearing Services in Schools*, *44*(2), 121–132. https://doi.org/10.1044/0161-1461(2012/12-0009)

Janse, E., & Newman, R. S. (2013). Identifying Nonwords: Effects of Lexical Neighborhoods, Phonotactic Probability, and Listener Characteristics. *Language and Speech*, *56*(4), 421–441. https://doi.org/10.1177/0023830912447914

Jespersen, O. (n.d.). Monosyllabism in English. *Proc. British Academy*, *14*, 341–368.

Jones, G., Tamburelli, M., Watson, S., Gobet, F., & Pine, J. (2010). Lexicality and Frequency in Specific Language Impairment: Accuracy and Error Data from Two Nonword Repetition Tests. *Journal of Speech, Language, and Hearing Research : JSLHR*, *53*, 1642–1655. https://doi.org/10.1044/1092-4388(2010/09-0222)

Kalnak, N., Peyrard-Janvid, M., Forssberg, H., & Sahlén, B. (2014). Nonword Repetition – A Clinical Marker for Specific Language Impairment in Swedish Associated with Parents' Language-Related Problems. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0089544

Kamhi, A. G., & Catts, H. W. (1986). Toward an Understanding of Developmental Language and Reading Disorders. *Journal of Speech and Hearing Disorders*, *51*(4), 337–347. https://doi.org/10.1044/jshd.5104.337

Kamhi, A. G., Catts, H. W., Mauer, D., Apel, K., & Gentry, B. F. (1988). Phonological and Spatial Processing Abilities in Language- and Reading-Impaired Children. *Journal of Speech and Hearing Disorders*, *53*(3), 316–327. https://doi.org/10.1044/jshd.5303.316

Kirby, J. P., & Yu, A. C. (2007). Lexical and phonotactic effects of wordlikeness judgements in Cantonese. *Proceedings of the 16th International Congress of Phonetic Sciences*. The 16th International Congress of Phonetic Sciences, Saarbrücken.

Kohnert, K., Windsor, J., & Yim, D. (2006). Do Language-Based Processing Tasks Separate Children with Language Impairment from Typical Bilinguals?

Learning Disabilities Research and Practice, *21*(1), 19–29. https://doi.org/10.1111/j.1540-5826.2006.00204.x

Lai, R., & Winterstein, G. (2019). *CantoMap. 1.0*.

Lai, R., & Winterstein, G. (2020). Cifu: A frequency lexicon of Hong Kong Cantonese. *Proceedings of the 12th Conference on Language Resources and Evaluation*, 3069–3077.

le Clercq, C. M. P., van der Schroeff, M. P., Rispens, J. E., Ruytjens, L., Goedegebure, A., van Ingen, G., & Franken, M.-C. (2017). Shortened Nonword Repetition Task (NWR-S): A Simple, Quick, and Less Expensive Outcome to Identify Children With Combined Specific Language and Reading Impairment. *Journal of Speech, Language, and Hearing Research*, *60*(8), 2241–2248. https://doi.org/10.1044/2017_JSLHR-L-16-0060

Lee, S. A. S., & Gorman, B. K. (2013). Nonword repetition performance and related factors in children representing four linguistic groups. *International Journal of Bilingualism*, *17*(4), 479–495. https://doi.org/10.1177/1367006912438303

Lee, T. H. T., Wong, C. H., Leung, S., Man, P., Cheung, A., Szeto, K., & Wong, C. S. P. (1994). *Lee/Wong/Leung Corpus. 1.0.*

Leonard, L. B. (1989). Language learnability and specific language impairment in children. *Applied Psycholinguistics*, *10*(2), 179–202. https://doi.org/10.1017/S0142716400008511

Leonard, L. B. (2020). A 200-Year History of the Study of Childhood Language Disorders of Unknown Origin: Changes in Terminology. *Perspectives of the ASHA Special Interest Groups*, *5*(1), 6–11. https://doi.org/10.1044/2019_PERS-SIG1-2019-0007

Leung, M.-T., & Law, S.-P. (2001). *HKCAC: The Hong Kong Cantonese Adult Language Corpus*.

Leung, M.-T., & Lee, A. (2002, May). *Hong Kong Corpus of Primary School Chinese*. 9th meeting of the International Clinical Phonetics and Linguistics Association, Hong Kong.

Leung, Y. N. (2010). *Diagnostic accuracy of nonword repetition and sentence repetition in Cantonese-speaking children with specific language impairment*

(SLI) [Unpublished undergraduate dissertation]. https://core.ac.uk/download/pdf/38002947.pdf

Lin, X., & Wang, L. J. (1992). *Yu Yan Xue Jiao Cheng [Linguistics Tutorial]*. Peking University Press.

Luke, K. K., & Wong, M. L. Y. (2015). *HKCan-Cor: The Hong Kong Cantonese Corpus. 1.0.*

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 : JSLHR*, *46*, 1138–1153. https://doi.org/10.1044/1092-4388(2003/089)

Matthews, S., & Yip, V. (2011). *Cantonese: A comprehensive grammar*. Routledge. http://public.ebookcentral.proquest.com/choice/publicfullrecord.aspx?p=668254

McGregor, K. K. (2009). Semantics in child language disorders. In *Handbook of child language disorders* (pp. 365–387). Psychology Press.

McKean, C., Letts, C., & Howard, D. (2013). Developmental Change Is Key to Understanding Primary Language Impairment: The Case of Phonotactic Probability and Nonword Repetition. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1579–1594. https://doi.org/10.1044/1092-4388(2013/12-0066)

Messer, M. H., Leseman, P. P. M., Boom, J., & Mayo, A. Y. (2010). Phonotactic probability effect in nonword recall and its relationship with vocabulary in monolingual and bilingual preschoolers. *Journal of Experimental Child Psychology*, *105*(4), 306–323. https://doi.org/10.1016/j.jecp.2009.12.006

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology*, *91*(1), 3–19. https://doi.org/10.1037/0022-0663.91.1.3

Mok, P. (2009). On the syllable-timing of Cantonese and Beijing Mandarin. *Chinese Journal of Phonetics*, *2*, 148–154.

Montgomery, J. W. (2002). Understanding the Language Difficulties of Children With Specific Language Impairments: Does Verbal Working Memory Matter?

American Journal of Speech-Language Pathology, *11*(1), 77. https://doi.org/10.1044/1058-0360(2002/009)

Munson, B. (2001). Phonological Pattern Frequency and Speech Production in Adults and Children. *Journal of Speech, Language, and Hearing Research*, *44*(4), 778–792. https://doi.org/10.1044/1092-4388(2001/061)

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. https://doi.org/10.1044/1092-4388(2005/072)

Norbury, C. F., Gooch, D., Wray, C., Baird, G., Charman, T., Simonoff, E., Vamvakas, G., & Pickles, A. (2016). The impact of nonverbal ability on prevalence and clinical presentation of language disorder: Evidence from a population study. *Journal of Child Psychology and Psychiatry*, *57*(11), 1247–1257. https://doi.org/10.1111/jcpp.12573

Öberg, L., & Bohnacker, U. (2022). Non-Word Repetition and Vocabulary in Arabic-Swedish-Speaking 4–7-Year-Olds with and without Developmental Language Disorder. *Languages*, *7*(3), 204. https://doi.org/10.3390/languages7030204

Ortiz, J. A. (2021). Using Nonword Repetition to Identify Language Impairment in Bilingual Children: A Meta-Analysis of Diagnostic Accuracy. *American Journal of Speech-Language Pathology*, *30*(5), 2275–2295. https://doi.org/10.1044/2021_AJSLP-20-00237

Paradis, J. (2023). Sources of individual differences in the dual language development of heritage bilinguals. *Journal of Child Language*, *50*(4), 793–817. https://doi.org/10.1017/S0305000922000708

Pham, G., & Ebert, K. D. (2020). Diagnostic Accuracy of Sentence Repetition and Nonword Repetition for Developmental Language Disorder in Vietnamese. *Journal of Speech, Language, and Hearing Research*, *63*(5), 1521–1536. https://doi.org/10.1044/2020_JSLHR-19-00366

Plante, E., & Vance, R. (1994). Selection of Preschool Language Tests: A Data-Based Approach. *Language, Speech, and Hearing Services in Schools*, *25*(1), 15–24. https://doi.org/10.1044/0161-1461.2501.15

Polišenská, K., & Kapalková, S. (2014). Improving child compliance on a computer administered nonword repetition task. *Journal of Speech, Language, and Hearing Research: JSLHR*, *57*(3), 1060–1068. https://doi.org/10.1044/1092-4388(2013/13-0014)

Poon, A. Y. K. (2010). Language use, and language policy and planning in Hong Kong. *Current Issues in Language Planning*, *11*(1), 1–66. https://doi.org/10.1080/14664201003682327

R Core Development Team. (2021). *R: A Language and Environment for Statistical Computing* [Computer software]. R Foundation for Statistical Computing. http://www.R-project.org/

Raven, J. C., Court, J. H., & Raven, J. (1996). *Raven's Standard Progressive Matrices*. Oxford Psychologists Press.

Reilly, S., Tomblin, B., Law, J., McKean, C., Mensah, F. K., Morgan, A., Goldfeld, S., Nicholson, J. M., & Wake, M. (2014). Specific language impairment: A convenient label for whom? *International Journal of Language & Communication Disorders*, *49*(4), 416–451. https://doi.org/10.1111/1460-6984.12102

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., Conway, L., Eadie, P., & Bretherton, L. (2010). Predicting Language Outcomes at 4 Years of Age: Findings From Early Language in Victoria Study. *Pediatrics*, *126*(6), e1530–e1537. https://doi.org/10.1542/peds.2010-0254

Riches, N. G., Loucas, T., Baird, G., Charman, T., & Simonoff, E. (2011). Non-word repetition in adolescents with Specific Language Impairment and Autism plus Language Impairments: A qualitative analysis. *Journal of Communication Disorders*, *44*(1), 23–36. https://doi.org/10.1016/j.jcomdis.2010.06.003

Rispens, J., Baker, A., & Duinmeijer, I. (2015). Word Recognition and Nonword Repetition in Children With Language Disorders: The Effects of Neighborhood Density, Lexical Frequency, and Phonotactic Probability. *Journal of Speech, Language, and Hearing Research*, *58*(1), 78–92. https://doi.org/10.1044/2014_JSLHR-L-12-0393

Roodenrys, S., & Hinton, M. (2002). Sublexical or lexical effects on serial recall of nonwords? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 29–33. https://doi.org/10.1037/0278-7393.28.1.29

Ruben, R. J. (2000). Redefining the Survival of the Fittest: Communication Disorders in the 21st Century: Redefining the Survival of the Fittest: Communication Disorders in the 21st Century. *The Laryngoscope*, *110*(2), 241–241. https://doi.org/10.1097/00005537-200002010-00010

Sahlén, B., Reuterskiöld-Wagner, C., Nettelbladt, U., & Radeborg, K. (1999). Non-word repetition in children with language impairment—Pitfalls and possibilities. *International Journal of Language & Communication Disorders*, *34*(3), 337–352. https://doi.org/10.1080/136828299247441

Santos, C. dos, & Ferré, S. (2018). A Nonword Repetition Task to Assess Bilingual Children's Phonology. *Language Acquisition*, *25*(1), 58–71. https://doi.org/10.1080/10489223.2016.1243692

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, *21*(2), 168–175. https://doi.org/10.3758/BF03202729

Schwob, S., Eddé, L., Jacquin, L., Leboulanger, M., Picard, M., Oliveira, P. R., & Skoruppa, K. (2021). Using Nonword Repetition to Identify Developmental Language Disorder in Monolingual and Bilingual Children: A Systematic Review and Meta-Analysis. *Journal of Speech, Language, and Hearing Research*, *64*(9), 3578–3593. https://doi.org/10.1044/2021_JSLHR-20-00552

Sharp, K. M., & Gathercole, V. C. M. (2013). Can a novel word repetition task be a language-neutral assessment tool? Evidence from Welsh–English bilingual children. *Child Language Teaching and Therapy*, *29*(1), 77–89. https://doi.org/10.1177/0265659012465208

Snowling, M., Chiat, S., & Hulme, C. (1991). Words, nonwords, and phonological processes: Some comments on Gathercole, Willis, Emslie, and Baddeley. *Applied Psycholinguistics*, *12*(3), 369–373. https://doi.org/10.1017/S0142716400009279

Stark, R. E., & Blackwell, P. B. (1997). Oral volitional movements in children with language impairments. *Child Neuropsychology*, *3*(2), 81–97. https://doi.org/10.1080/09297049708401370

Stokes, S. F., Wong, A. M.-Y., Fletcher, P., & Leonard, L. B. (2006). Nonword Repetition and Sentence Repetition as Clinical Markers of Specific Language Impairment: The case of Cantonese. *Journal of Speech and Hearing Research*, *49*(2), 219–236.

Szewczyk, J. M., Marecka, M., Chiat, S., & Wodniecka, Z. (2018). Nonword repetition depends on the frequency of sublexical representations at different grain sizes: Evidence from a multi-factorial analysis. *Cognition*, *179*, 23–36. https://doi.org/10.1016/j.cognition.2018.06.002

Thal, D. J., Miller, S., Carlson, J., & Vega, M. M. (2005). Nonword Repetition and Language Development in 4-Year-Old Children With and Without a History of Early Language Delay. *Journal of Speech, Language, and Hearing Research*, *48*(6), 1481–1495. https://doi.org/10.1044/1092-4388(2005/103)

Thordardottir, E., & Brandeker, M. (2013). The effect of bilingual exposure versus language impairment on nonword repetition and sentence imitation scores. *Journal of Communication Disorders*, *46*(1), 1–16. https://doi.org/10.1016/j.jcomdis.2012.08.002

Thorn, A. S. C., & Frankish, C. R. (2005). Long-Term Knowledge Effects on Serial Recall of Nonwords Are Not Exclusively Lexical. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(4), 729–735. https://doi.org/10.1037/0278-7393.31.4.729

To, C. K. S., Cheung, P. S. P., & McLeod, S. (2013). A population study of children's acquisition of Hong Kong Cantonese consonants, vowels, and tones. *Journal of Speech, Language, and Hearing Research: JSLHR*, *56*(1), 103–122. https://doi.org/10.1044/1092-4388(2012/11-0080)

Tomblin, J. B., Records, N. L., Buckwalter, P., Zhang, X., 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. https://doi.org/10.1044/jslhr.4006.1245

Tomblin, J. B., Records, N. L., & Zhang, X. (1996). A System for the Diagnosis of Specific Language Impairment in Kindergarten Children. *Journal of Speech, Language, and Hearing Research*, *39*(6), 1284–1294. https://doi.org/10.1044/jshr.3906.1284

T'sou, B., Lee, T., Tung, P., Man, Y., Chan, A., To, C., & Chan, Y. (2006). *Hong Kong Cantonese Oral Language Assessment Scale*. City University of Hong Kong.

Tuller, L. (2015). Clinical use of parental questionnaires in multilingual contexts. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing Multilingual Children—Disentangling Bilingualism from Language Impairment* (pp. 301–330). Multilingual Matters.

van den Bedem, N. P., Dockrell, J. E., van Alphen, P. M., Kalicharan, S. V., & Rieffe, C. (2018). Victimization, Bullying, and Emotional Competence: Longitudinal Associations in (Pre)Adolescents With and Without Developmental Language Disorder. *Journal of Speech, Language, and Hearing Research*, *61*(8), 2028–2044. https://doi.org/10.1044/2018_JSLHR-L-17-0429

Wang, X., & Huang, J. (2016). Non-word repetition performance in Mandarin-speaking preschool children with and without specific language impairment (SLI). *Chinese Journal of Applied Linguistics*, *39*(3), 337–353. https://doi.org/10.1515/cjal-2016-0022

Washington, J. A., & Craig, H. K. (2004). A Language Screening Protocol for Use With Young African American Children in Urban Settings. *American Journal of Speech-Language Pathology*, *13*(4), 329–340. https://doi.org/10.1044/1058-0360(2004/033)

Watkins, R. V., Kelly, D. J., Harbers, H. M., & Hollis, W. (1995). Measuring Children's Lexical Diversity: Differentiating Typical and Impaired Language Learners. *Journal of Speech, Language, and Hearing Research*, *38*(6), 1349–1355. https://doi.org/10.1044/jshr.3806.1349

Young, A. R., Beitchman, J. H., Johnson, C., Douglas, L., Atkinson, L., Escobar, M., & Wilson, B. (2002). Young adult academic outcomes in a longitudinal sample of early identified language impaired and control children. *Journal of Child Psychology and Psychiatry*, *43*(5), 635–645. https://doi.org/10.1111/1469-7610.00052

Zamuner, T. S., Gerken, L., & Hammond, M. (2004). Phonotactic probabilities in young children's speech production. *Journal of Child Language*, *31*(3), 515–536. https://doi.org/10.1017/S0305000904006233

Appendix

Full List of NWR Task Stimuli

Number of Syllables	Stimuli (CV)	Stimuli (CVC)
High-Lexicality		
2	Fe Maa	Mak Lim
	Ji Haa	Bam Jik
	Maa Lo	Duk Bam
3	Fe Ji Maa	Lim Duk Mak
	Haa Ji Lo	Bam Duk Jik
	Lo Fe Haa	Mak Duk Lim
4	Fe Lo Haa Maa	Mak Bam Lim Jik
	Maa Haa Ji Lo	Bam Jik Mak Duk
	Ji Fe Haa Lo	Jik Lim Bam Mak
5	Lo Ji Fe Maa Haa	Duk Lim Mak Jik Bam
	Haa Fe Lo Maa Ji	Bam Duk Jik Mak Lim
	Ji Maa Lo Haa Fe	Jik Bam Lim Duk Mak
High-Lexicality-Vowel-Matched		
2	Mi Lo	Duk Gap
	Bo Ngo	Hap Bik
	Lo Fo	Bik Juk
3	Bo Mi Ngo	Juk Bik Hap
	Fo Ngo Bo	Gap Duk Bik
	Ngo Mi Fo	Hap Juk Gap
4	Fo Mi Lo Bo	Duk Gap Juk Bik
	Lo Mi Fo Ngo	Gap Duk Hap Juk
	Ngo Bo Mi Lo	Duk Bik Juk Hap
5	Lo Bo Mi Fo Ngo	Bik Gap Duk Hap Juk
	Ngo Mi Bo Lo Fo	Gap Bik Juk Duk Hap
	Mi Ngo Lo Fo Bo	Juk Gap Bik Hap Duk

Low-Lexicality		
2	Fi Lu	Lut Wek
	Bu Ngu	Jek Hik
	Lu Hu	Hik Ngut
3	Bu Fi Ngu	Ngut Hik Jek
	Hu Ngu Bu	Wek Lut Hik
	Ngu Fi Hu	Jek Ngut Wek
4	Hu Fi Lu Bu	Lut Wek Ngut Hik
	Lu Fi Hu Ngu	Wek Luk Jek Ngut
	Ngu Bu Fi Lu	Lut Hik Ngut Jek
5	Lu Bu Fi Hu Ngu	Hik Wek Lut Jek Ngut
	Ngu Fi Bu Lu Hu	Wek Hik Ngut Lut Jek
	Fi Ngu Lu Hu Bu	Ngut Wek Hik Jek Lut
CL-NWR		
2	Sibu	N/A
	Dula	
	Magi	
	Lumi	
3	Sipula	
	Bamudi	
	Malidu	
	Lumiga	
4	Sipalida	
	Mugidala	
	Gasulumi	
	Litisaku	
5	Sipumagila	
	Duligasumu	
	Malusikuba	
	Lidabimudi	