ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Nonword Repetition in Children With Developmental Language Disorder: Revisiting the Case of Cantonese

Nga Ching Fu,[a,b] Si Chen,[a,b,c,d] Kamila Polišenská,[e,f] Angel Chan,[a,b,c] Rachel Kan,[a,b] and Shula Chiat[f]

[a]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR [b]Research Centre for Language, Cognition, and Neuroscience, The Hong Kong Polytechnic University, Hong Kong SAR [c]PolyU-PekingU Research Centre on Chinese Linguistics, The Hong Kong Polytechnic University, Hong Kong SAR [d]Research Institute for Smart Ageing, The Hong Kong Polytechnic University, Hong Kong SAR [e]Division of Human Communication, Development and Hearing, The University of Manchester, United Kingdom [f]Department of Language and Communication Science, City, University of London, United Kingdom

ARTICLE INFO

Article History:
Received July 7, 2022
Revision received March 6, 2023
Accepted January 22, 2024

Editor-in-Chief: Stephen M. Camarata
Editor: Cathy Huaqing Qi

https://doi.org/10.1044/2024_JSLHR-22-00397

ABSTRACT

Purpose: Nonword repetition (NWR) has been described as a clinical marker of developmental language disorder (DLD), as NWR tasks consistently discriminate between DLD and typical development (TD) cross-linguistically, with Cantonese as the only reported exception. This study reexamines whether NWR is able to generate TD/DLD group differences in Cantonese-speaking children by reporting on a novel set of NWR stimuli that take into account factors known to affect NWR performance and group differentiation, including lexicality, sublexicality, length, and syllable complexity.
Method: Sixteen Cantonese-speaking children with DLD and 16 age-matched children with TD repeated two sets of high-lexicality nonwords, where all constituent syllables are morphemic in Cantonese but meaningless when combined, and one set of low-lexicality nonwords, where all constituent syllables are nonmorphemic. Low-lexicality nonwords were further classified on sublexicality in terms of consonant–vowel (CV) combination attestedness (whether or not CV combinations in nonword syllables occur in real Cantonese words).
Results: Children with DLD scored significantly below their peers with TD. Effect sizes showed that high-lexicality nonwords and nonword syllables with attested CV combinations offered the greatest TD/DLD group differentiation. Nonword length and syllable complexity did not affect TD/DLD group differentiation.
Conclusions: NWR can capture TD/DLD group differences in Cantonese-speaking children. Lexicality and sublexicality effects must be considered in designing NWR stimuli for TD/DLD group differentiation. Future studies should replicate the present study on a larger sample size and a younger population as well as examine the diagnostic accuracy of this NWR test.
Supplemental Material: https://doi.org/10.23641/asha.25529371

Nonword repetition (NWR) has been advocated as a potential clinical marker for developmental language disorder (DLD; previously known as specific language impairment[1]; Bishop et al., 1996; Conti-Ramsden et al., 2001; Dollaghan & Campbell, 1998), as it has been shown to be able to differentiate between children with typical development (TD) and children with DLD cross-linguistically.

Correspondence to Angel Chan: angel.ws.chan@polyu.edu.hk. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

[1]Older terminology, *specific language impairment* (*SLI*), describes children with significant language difficulties whose nonverbal cognitive abilities fall within the normal range—its diagnostic criteria require a significant mismatch between verbal and nonverbal abilities. Newer terminology, *developmental language disorder* (*DLD*), describes children with significant language difficulties that cause negative functional impact, are associated with poor prognosis, and are not associated with a biomedical condition—its diagnostic criteria no longer require a mismatch between verbal and nonverbal abilities (Bishop et al., 2017).

Significant effects of group (TD vs. DLD) have been found for NWR tests in diverse languages, including Arabic (Shaalan, 2020; Taha et al., 2021), Hebrew (Meir, 2017), Icelandic (Thordardottir, 2008), Italian (Dispaldro et al., 2013), Russian (Meir, 2017), and Swedish (Kalnak et al., 2014), among others, and in two Asian tonal languages, namely, Vietnamese (Pham & Ebert, 2020) and Mandarin (Chi, 2007; Wang & Huang, 2016). NWR also captures TD/DLD group differences across children and adolescents of different ages (Conti-Ramsden et al., 2001; Riches et al., 2011; Schwob et al., 2021).

Cantonese has been a rare exception to these findings. In a study of monolingual Cantonese-speaking children, Stokes et al. (2006) compared NWR performance in children with DLD (aged 4–5 years) with that in age-matched children with TD and younger, language-matched children with TD. Although an age effect was found, whereby younger children with TD achieved significantly lower accuracy in NWR, no significant differences were found between the DLD and TD age-matched groups. These findings suggested that NWR is not a potential indicator of DLD in Cantonese-speaking children, unlike what has been proposed for children speaking other languages.

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 (where "C" is consonant and "V" is vowel), and consonant clusters are not permitted. Cantonese is also a tonal language, with each syllable being marked by one of the 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 such as 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 a possible reason for the discrepant findings is that nonword stimuli designed with reference to Cantonese phonotactic constraints have relatively low levels of segmental and suprasegmental complexity, making Cantonese NWR stimuli less taxing on short-term memory than they appeared to be in other examined languages. However, more recent studies of NWR in Mandarin, which is typologically similar to Cantonese, and Vietnamese, which also has relatively simple phonotactic structures, have found good differentiation between TD and DLD groups (Chi, 2007; Pham & Ebert, 2020; Wang & Huang, 2016).

Nonword Characteristics Affecting NWR Performance and TD/DLD Group Differentiation

If the simple phonotactic properties of Cantonese cannot explain the previous lack of significant findings, the design of the nonword stimuli adopted in Stokes et al. (2006) may be the 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. Stokes et al. intentionally avoided the use of real morphemes and highly predictable consonant–vowel (CV) combinations to reduce the influence of prior lexical knowledge on the grounds that NWR tasks were proposed as strict measures of working memory capacity. However, more recent findings have demonstrated that nonwords with higher resemblance to real words in the ambient language are more effective in capturing TD/DLD group differences (Graf-Estes et al., 2007). In the following, we will discuss nonword-related factors that have been found to affect NWR performance and TD/DLD group differentiation.

Lexicality

We use the term *lexicality* to describe the degree to which nonwords resemble real words in an ambient language in an objective, measurable manner, as opposed to *wordlikeness*, which is often used to describe the subjective measure of a nonword's similarity to real words based on native speaker judgments. Specifically, lexicality can be modified and measured in terms of morphemicity (i.e., whether morphemic elements are incorporated into nonword stimuli). To illustrate with examples in English, compare items in Gathercole and Baddeley's (1996) test (e.g., "defermication," which contains 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 Dollaghan and Campbell can therefore be described as having lower lexicality than those from Gathercole and Baddeley. Meta-analyses (Graf-Estes et al., 2007) and individual studies (Casalini et al., 2007) reported better NWR performance on nonwords that incorporated real morphemes (i.e., had higher lexicality) than those that did not, presumably because high-lexicality nonwords allowed children to draw on their existing lexical–phonological knowledge to support NWR through redintegration. High-lexicality nonwords have also been found to capture greater TD/DLD group differences (Graf-Estes et al., 2007), likely because typically developing children's stronger language skills allow for greater use of redintegration strategies when TD children repeat high-lexicality nonwords. Contrarily, neither group is able to draw on long-term lexical knowledge to support the repetition of

nonwords that do not resemble real words in the ambient language.

Sublexicality

Sublexicality is used to describe the degree to which nonword elements at a sublexical level objectively resemble sublexical elements in real words in the ambient language. One such measure is phonotactic probability, which measures the frequency with which phonemic bigrams or phonemic sequences at other grain sizes within nonwords appear in the ambient language. Higher NWR accuracy has been reported for nonwords containing syllables with high phonotactic probability than those with low phonotactic probability (e.g., English: McKean et al., 2013; Polish: Szewczyk et al., 2018). Another measure of sublexicality is CV combination attestedness, described in Stokes et al. (2006), where syllables within the nonword items were classified into two subtypes labeled "IN" and "OUT." IN syllables contained CV combinations that are attested in Cantonese but do not necessarily constitute real morphemes (e.g., the IN syllable /tɛː/ combines with a final consonant /−ŋ/ to form a real word /tɛːŋ/, meaning "to listen," but /tɛː/ itself is not a real morpheme), whereas OUT syllables consisted of CV combinations that never occur in Cantonese (e.g., /ŋuːt/, where neither itself nor the CV combination /ŋuː/ occurs in Cantonese). Stokes et al. reported significantly higher NWR accuracy on IN syllables than on OUT syllables across both TD and DLD groups, suggesting that children drew on their sublexical representations to facilitate NWR performance. In line with findings on the benefits of sublexical familiarity in other languages, the difference between TD and DLD groups was also larger on IN syllables, albeit not significantly (Stokes et al., 2006).

Length

Unsurprisingly, studies have consistently found children's NWR accuracy to decline as nonwords increased in length, which raises demands on children's limited short-term memory capacities (Schwob et al., 2021). Whereas some studies reported that children with DLD are disproportionately affected by increasing nonword length (Dispaldro et al., 2013; McKean et al., 2013), reflecting more limited phonological short-term memory capacities in children with DLD, others reported no interaction between clinical status and nonword length (Boerma et al., 2015).

Segmental Complexity

Nonwords with more complex segments, such as those containing consonant clusters, have also been reported to be more difficult to repeat than nonwords with less complex segments, such as those without consonant clusters, perhaps because nonwords with higher levels of segmental complexity require more complex phonological analysis, motor planning, and/or phonological memory (Polišenská & Kapalková, 2014; Szewczyk et al., 2018). An increase in segmental complexity is also disproportionally challenging for children with DLD, compared to their peers with TD (Gallon et al., 2007; Jones et al., 2010).

In summary, a number of nonword characteristics may affect children's NWR performance and TD/DLD group differentiation in NWR tests. These include (a) nonword *lexicality*, where nonword stimuli that incorporated real morphemes in a child's language led to better NWR performance and greater TD/DLD group differentiation; (b) nonword *sublexicality*, where nonwords with higher phonotactic probability and those that incorporated attested CV combinations led to better NWR performance; (c) nonword *length*, where NWR performance declines with increasing number of syllables in a nonword, with inconclusive evidence on its effects on TD/DLD group differentiation; and (d) *segmental complexity* of nonwords, where lower NWR performance can be seen on nonwords that contained consonant clusters, and nonwords with consonant clusters may generate greater TD/DLD group differences.

The Present Study

Despite the growing understanding of how nonword characteristics may affect NWR performance and TD/DLD group differentiation, to our knowledge, no study to date has examined the use of any alternative nonword stimuli to those used by Stokes et al. (2006), for capturing TD/DLD group differences in Cantonese-speaking children. We report 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 above. With this novel stimulus set, our objectives are to (a) revisit whether NWR is able to generate TD/DLD group differences in Cantonese-speaking children and (b) identify whether nonwords at particular levels of lexicality, sublexicality, length, and syllable complexity are best suited to capturing TD/DLD group differences in Cantonese.

Specifically, four research questions (RQs) are addressed:

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

RQ3: How does nonword length, a nonword-related variable known to affect NWR performance in children acquiring other languages, affect NWR performance and TD/DLD group differentiation in Cantonese-speaking children?

RQ4: How does syllable complexity in nonwords affect NWR performance and TD/DLD group differentiation in Cantonese-speaking children?

Method

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 were described as "predominantly monolingual," as they acquired Cantonese (the majority community language of Hong Kong) as first language at home and attended local schools where Cantonese was the medium of instruction, while 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, for example, heritage speakers of Cantonese living in an English-speaking country or children attending international schools in Hong Kong, the children included in this study were only exposed to their second languages—English and Mandarin—for less than 20% of their awake time. Therefore, they did not have extensive and intensive exposure to languages other than Cantonese and were described as predominantly monolingual rather than bilingual/multilingual, following common operational definitions of monolingualism/bilingualism in other studies in terms of relative exposure to languages (see, e.g., Paradis, 2023).

Unlike most previous studies looking at younger children (Schwob et al., 2021), we examined older 8- to 11-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 the age of 15;4 (years;months; Riches et al., 2011; Schwob et al., 2021), we believed that examining children from this older age range would still allow our research objectives to be appropriately addressed. Moreover, given that no previous Cantonese NWR studies have yielded significant group effects, the lack of information on effect sizes from previous studies did not allow us to conduct an a priori power analysis. We therefore had a sample size of children with DLD versus age-matched peers with TD that was comparable to the existing Cantonese NWR literature.

Participant Selection Criteria

Sixteen children met the criteria for a DLD diagnosis following the CATALISE (Criteria and Terminology Applied to Language Impairments) criteria (Bishop et al., 2017). Parents and/or school personnel expressed concerns providing evidence of negative functional impact of their language difficulties, affecting daily social interactions or educational progress. Moreover, results from a standardized norm-referenced language assessment, the Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS; T'sou et al., 2006), provided objective evidence for lack of competency even in these children's best language. Specifically, 14 children scored at 1.25 *SD*s below age means in two or more out of six subtests of the HKCOLAS, and two children scored at −1.25 *SD*s in one subtest and −1.0 *SD* in another subtest. Two children had co-occurring attention-deficit/hyperactivity disorder and dyslexia, respectively.[2]

The 16 children with TD were individually matched to each child in the DLD group in age (within 4 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 the 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 were significantly below those of the TD group, $t(30) = 2.27$, $p = .02$ (see Table 1 for group means and standard deviations). None of the participants were suspected to have autism spectrum disorder (ASD) by their parents and school personnel, and none had ever undergone assessments for or received a diagnosis of ASD. Table 1 summarizes the demographic information, language assessment scores, and cognitive assessment scores of the participants.

This study was carried out in accordance with the recommendations of the Human Subjects Ethics Sub-Committee at The Hong Kong Polytechnic University (Reference No. HSEARS20161230004). Written informed consent was given by the parents of each participant.

Materials

HKCOLAS

The HKCOLAS (T'sou et al., 2006) is a norm-referenced language assessment tool designed to examine Cantonese oral language abilities of 5- to 12-year-old

[2]Under the CATALISE diagnostic criteria, both attention-deficit/hyperactivity disorder and dyslexia are considered to be co-occurring conditions with DLD, as opposed to differentiating conditions; thus, the two children were retained in our DLD sample.

Table 1. Demographic information, language assessment scores, and cognitive scores of the developmental language disorder (DLD) and typical development (TD) groups.

Participant information	DLD group (n = 16)	TD group (n = 16)
Demographic information		
Age (years;months)	8;01–11;00 (M = 9;07, SD = 0;11)	8;00–11;03 (M = 9;07, SD = 1;00)
Gender	Male = 11; Female = 5	
Grade in school	Primary 2 = 2; Primary 3 = 2; Primary 4 = 8; Primary 5 = 4	
Language assessment (z scores)[a]		
Word Definition Test	M = –0.85, SD = 0.85	M = 0.63, SD = 1.24
Lexical-Semantic Relations Test	M = –2.10, SD = 0.85	M = –0.11, SD = 0.91
Expressive Nominal Vocabulary Test	M = –1.98, SD = 1.64	M = 0.29, SD = 0.95
Test of Hong Kong Cantonese Grammar	M = –0.96, SD = 1.82	M = 0.74, SD = 0.84
Textual Comprehension Test	M = –1.24, SD = 1.36	M = 0.79, SD = 1.00
Narrative Test	M = –1.66, SD = 1.21	M = 0.04, SD = 0.97
Cognitive assessment (standard scores)		
Raven's Progressive Matrices	M = 102.7, SD = 12.5	M = 113.6, SD = 11.4

[a]Language assessment scores are listed for the six subtests within the standardized norm-referenced Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS). In the HKCOLAS, the diagnostic criterion for language disorder is scoring 1.25 *SD*s below age means in two or more out of six subtests.

children in Hong Kong. The HKCOLAS has six subtests targeting vocabulary (Lexical-Semantic Relations Test, Word Definition Test, and Expressive Nominal Vocabulary Test), morphosyntax (Test of Hong Kong Cantonese Grammar), and narratives (Textual Comprehension Test and Narrative Test). Children who score 1.25 *SD*s below age means in two or more subtests qualify for a diagnosis of language disorder; at this diagnostic cutoff, the HKCOLAS has a sensitivity of 0.95 and a specificity of 0.98. The HKCOLAS also has high test reliabilities based on coefficient alpha (.80–.97 across all subtests) and standardized error of measurement.

Pure-Tone Audiometry Hearing Screening Test

A 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 HL at frequencies of 500, 1000, 2000, and 4000 Hz. To pass the hearing screening, children have to respond to pure tones at all test frequencies at 25 dB HL in both left and right ears.

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 nonverbal intelligence quotient to screen out the possibility of intellectual disability when assessing children suspected of DLD. Raven's Progressive Matrices includes 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.

NWR Stimuli

Three sets of NWR stimuli, which varied in lexicality levels in relation to Cantonese, were used (see Table 2 for a

Table 2. Comparison of nonword sets with examples.

Nonwords	Lexicality	Sublexicality	Vowel range	Examples
High-lexicality	High: 100% syllables are morphemic in Tone 1 but meaningless when combined.	Not examined	Wider	fe* ji* maa*
High-lexicality–vowel-matched	High: 100% syllables are morphemic in Tone 1 but meaningless when combined.	Not examined	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

Note. Morphemic syllables in examples are marked with an asterisk (*). CV = consonant–vowel.

comparison of the nonword sets and examples as well as Supplemental Material S1 for the full list of nonword items).

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"), whereas the other half were constructed solely with relatively complex consonant–vowel–consonant (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). The total number of high-lexicality items was 24.

Low-lexicality nonwords. Low-lexicality nonwords had the lowest lexicality level, where all constituent syllables are nonmorphemic 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 nonmorphemic, although the syllable selection criteria are more relaxed in Stokes et al., where syllables were only nonmorphemic in the tones they were presented in; in contrast, low-lexicality nonwords in this study consist of syllables that are nonmorphemic 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, whereas the other half were created with CVC syllables. The total number of low-lexicality items was 24.

IN versus OUT syllables. Within low-lexicality nonwords, where all syllables were nonmorphemic, the constituent syllables could be further divided into two subtypes—IN and OUT—based on their sublexical 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"), whereas OUT syllables contain unattested CV combinations (e.g., "ngut," where neither itself nor the CV combination "ngu" occurs 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 sublexical feature of NWR stimuli.

High-lexicality–vowel-matched nonwords. To match the smaller vowel range of low-lexicality nonwords, an additional set of high-lexicality nonwords, labeled "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, whereas the other half were created with CVC syllables. The total number of high-lexicality–vowel-matched items was 24.

Other considerations. Consonants and vowels used across all nonword sets were expected to be acquired by the age of 4 years 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 multisyllabic Cantonese words were avoided. To control for prosodic effects, all nonwords were set to be articulated with Cantonese Tone 1, with even length and stress on each syllable. All items were recorded by a female native Cantonese-speaking student speech and language therapist (SLT).

Procedure

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, an NWR task, a standardized language assessment, and a nonverbal intelligence quotient test, lasted for about 2 hr.

The procedures of the NWR task were modeled after those from Polišenská and Kapalková (2014). The computerized NWR task was presented as a picture story through PowerPoint slides. Participants listened to prerecorded instructions and stimuli that were embedded into the slides through noise-canceling 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 randomized. 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.

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 sublexical 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 the initial /ŋ/ consonant and substitutions between the 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 penalized, and such changes were rarely observed.

Interrater 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 (by both whole-nonword and 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 the syllable level. At whole-nonword–level scoring, the average intraclass correlation coefficient (ICC) using a two-way mixed model and absolute agreement was .98 for high-lexicality–vowel-matched nonwords (95% CI [.90, .99]), .90 for high-lexicality nonwords (95% CI [.63, .98]), and .93 for low-lexicality nonwords (95% CI [.70, .98]). At syllable-level scoring, the average ICC using a two-way mixed model and absolute agreement was .89 (95% CI [.88, .91]), indicating good to excellent levels of reliability between raters at both levels of scoring.

Data Analysis

NWR scores were analyzed with logistic mixed-effects models using the R package lme4 (Bates & Maechler, 2010) in R (Version 4.1.3; R Core Development Team, 2021). There were no missing data, and all assumptions under logistic mixed-effects models were met. No data transformation was required. 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 the whole-nonword level (as a categorical variable of correct vs. incorrect for each trial), whereas 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 the 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 the syllable level, as opposed to the whole-nonword level, because IN and OUT syllables co-occurred within items; that is, each nonword consisted of both IN and OUT syllables, so nonwords had to be scored at the syllable level for the effects of CV combination attestedness to be examined. Participants 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 the whole-nonword level (as a categorical variable of correct vs. incorrect for each trial), whereas 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 the whole-nonword level (as a categorical variable of correct vs. incorrect for each trial), whereas 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.

Results

Effects of Lexicality

RQ1 was addressed with Model 1, which examined how lexicality of NWR stimuli affected NWR performance and TD/DLD group differentiation in Cantonese-speaking children. The fixed effects of Model 1 are shown in Table 3. There was a significant main effect of

Table 3. Results of Model 1 on the effects of lexicality on nonword repetition accuracy at the whole-nonword level and typical development (TD)/developmental language disorder 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		

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 × 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*s) 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; high-lexicality–vowel-matched: $OR = 4.04$, medium effect size), compared to low-lexicality nonwords ($OR = 3.10$, small effect size).

Effects of Sublexicality

RQ2 was addressed with Model 2, which examined how sublexicality, in terms of CV combination attestedness, affected NWR performance and TD/DLD group differentiation in Cantonese-speaking children. The fixed effects of Model 2 are shown in Table 4. 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 in the interpretation of this interaction, we used likelihood ratio tests to examine the effects of group on NWR accuracy at the syllable level in IN and OUT syllables separately. When IN syllables were analyzed, the fixed effect of group was highly significant ($p < .001$) with a medium effect size ($OR = 3.08$), and when OUT syllables were analyzed, 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% CIs) for children on IN and OUT syllables, for the DLD and TD groups separately (see Figure 1). Figure 1 shows that in children with DLD, the gap between performance on IN and OUT syllables was small, when the 95% CIs were taken into consideration. On the other hand, in children with TD, the gap was large, indicating a more consistent and prominent improvement in performance on IN syllables compared to OUT syllables, even when the 95% CIs were factored in.

Effects of Nonword Length

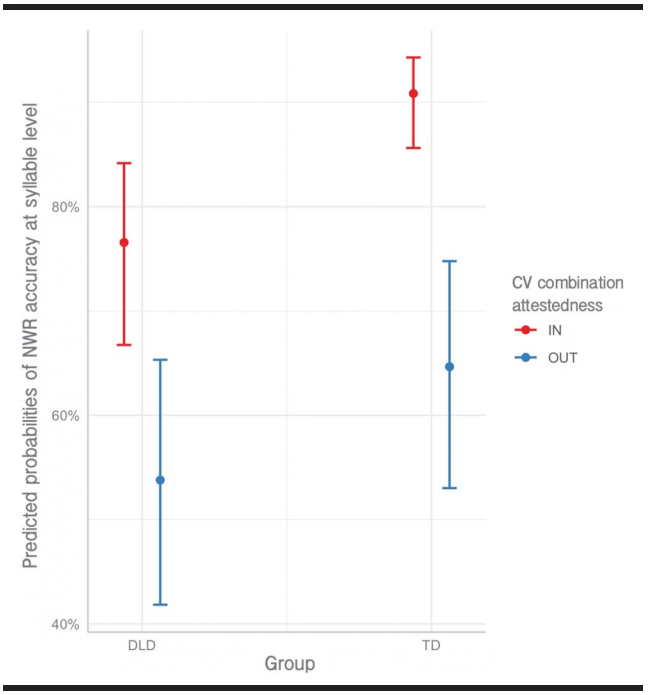
RQ3 was addressed with Model 3, which examined how nonword length affected NWR performance and TD/DLD group differentiation in Cantonese-speaking children. The fixed effects of Model 3 are shown in Table 5. There was a significant main effect of length, where nonword accuracy decreased as nonwords increased in number of syllables ($p < .001$), but the model did not register a significant

Table 4. Results of Model 2 on the effects of sublexicality on nonword repetition accuracy at the syllable level and typical development (TD)/developmental language disorder 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		

Note. CV = consonant–vowel.

Figure 1. Predicted probability of nonword repetition (NWR) accuracy at the syllable level (with 95% confidence intervals) on IN and OUT syllables in the developmental language disorder (DLD) versus typical development (TD) groups. CV = consonant–vowel.



main effect of group (p = .12).[3] The model also did not register a significant interaction between length and group.

Effects of Syllable Complexity

RQ4 was addressed with Model 4, which examined how syllable complexity in nonwords affected NWR performance and TD/DLD group differentiation in Cantonese-speaking children. The fixed effects of Model 4 are shown in Table 6. 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.

Summary of Results

The results showed that these Cantonese-speaking children with DLD performed significantly lower than their age-matched peers with TD in Cantonese NWR tasks. Both children with TD and children with DLD performed better on nonwords with high lexicality compared to those with low lexicality, and high-lexicality nonwords yielded greater TD/DLD group differences, based on effect sizes. Both children with TD and children with DLD also performed better on IN syllables (containing attested CV combinations) compared to OUT syllables (containing unattested CV combinations), and IN syllables captured greater TD/DLD group differentiation than OUT syllables, based on effect sizes. Children in both TD and DLD groups had a drop in NWR performance as nonwords increased in length, but these children's performance was not affected by syllable complexity.

[3]Although 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 3 on the effects of nonword length on nonword repetition accuracy at the whole-nonword level and typical development (TD)/developmental language disorder 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		

Discussion

This study 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, we addressed

Table 6. Results of Model 4 on the effects of syllable complexity on nonword repetition accuracy at the whole-nonword level and typical development (TD)/developmental language disorder 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		

Note. CVC = consonant–vowel–consonant.

four RQs, which asked whether NWR performance and TD/DLD group differentiation in Cantonese-speaking children are affected by (a) lexicality, in terms of morphemicity; (b) sublexicality, in terms of CV combination attestedness; (c) nonword length, in number of syllables; and (d) syllable complexity, by comparing syllables of CV or CVC structure. This study is the first to document that Cantonese-speaking children with TD do perform significantly above their peers with DLD in NWR, and our findings suggest that nonwords with higher levels of lexicality and sublexicality maximize the degree of TD/DLD group differentiation.

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

The present findings suggest that Cantonese may not be 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 short-term/working memory and, therefore, unable to generate TD/DLD group differences in children acquiring Cantonese. Using our newly designed NWR stimuli, 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, we even found significant TD/DLD group differences in NWR performance on the low-lexicality items, which were similar to the stimuli used by Stokes et al. (2006), who previously reported nonsignificant findings. Our results, therefore, point to NWR having potential to be further developed as an assessment tool to aid in 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. (14 children with DLD and 15 age-matched peers with TD), the present findings have established that Cantonese is unlikely to be a true cross-linguistic exception in NWR, supporting the potential utility of NWR in identifying DLD alongside other forms of assessment in Cantonese, as well as cross-linguistically.

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

Our 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.

Lexicality

We report positive findings 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 nonmorphemic 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. We further found that 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 children with TD are expected to have stronger lexical representations (and sublexical representations; see the following Sublexicality section), children with TD can rely on their stronger lexical (and sublexical) 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 for children to draw on their lexical–phonological knowledge to support the repetition of nonwords.

Sublexicality

Our data suggest that even within low-lexicality items, where all constituent syllables are nonmorphemic, children achieve better performance on syllables that contain CV combinations attested in Cantonese (i.e., IN syllables) than on 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 on OUT syllables (as shown by effect size differences), and our plots also indicated that children with DLD showed less benefit on IN syllables

than children with TD. This pattern of findings mirrored those on lexicality, suggesting that sublexicality in nonwords affected NWR performance and TD/DLD group differentiation in an analogous manner to lexicality and that sublexicality 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 sublexical representations—to support NWR and particularly that sublexical representations may be fundamental to successful NWR (Szewczyk et al., 2018).

Length

We also found a significant effect of length 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 short-term memory (Schwob et al., 2021). Our data did not register an interaction between group and length, suggesting that these children with DLD were not disproportionally affected by increasing nonword length.

Syllable Complexity

The present data did not register a significant effect of syllable complexity or an interaction effect between group and syllable complexity. Although 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.

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. In particular, we did not find significant TD/DLD group differences in Model 3, concerning the effects of nonword length—results from a post hoc power analysis indicated that the model was underpowered (power = 28.5%); thus, group differences may be weakened by being underpowered in some analyses. On the other hand, significant TD/DLD group differences were still registered in all other statistical models 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 carry over to younger children, as the lack of positive findings in Stokes et al. (2006) was reported on a younger participant group at preschool 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.

Conclusions

Our results suggest that 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 even in Cantonese. Our results also add to the understanding of optimal NWR stimuli design from a language with very different lexical and sublexical properties from most languages studied. Both lexical and sublexical factors must be considered in the design of NWR stimuli for generating TD/DLD group differences, as children draw on lexical and sublexical representations to support 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.

Data Availability Statement

The original data and the test items presented in the study are included in Supplemental Material S1. Further inquiries can be directed to the corresponding author.

Acknowledgments

This research was supported by a research grant titled "Revisiting Nonword Repetition as a Clinical Marker in Cantonese-speaking Children with Developmental Language Disorder: A Pilot Study" (Grant P0031730 G-UAK5; principal investigator [PI]: Angel Chan) awarded by The

Hong Kong Polytechnic University and a research grant titled "Revisiting Nonword Repetition as a Diagnostic Tool in Cantonese-speaking Children with Developmental Language Disorder" (Grant (PolyU 15609420; PI: Angel Chan) awarded by The Hong Kong Research Grants Council General Research Fund. The authors thank Christy Chu, Sonia Hui, Eva Kwan, Jovial Wong, and Katrina Wong for help as research assistants as well as all children who took part in this study.

References

Bates, D., & Maechler, M. (2010). *lme4: Linear mixed-effects models using S4 classes* (R Package Version 0.999375-33). 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., North, T., & Donlan, C. (1996). Nonword repetition as a behavioural marker for inherited language impairment: Evidence from a twin study. *The 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., & the CATALISE-2 Consortium. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *The Journal of Child Psychology and Psychiatry, 58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721

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

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.

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

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *The Journal of Child Psychology and Psychiatry, 42*(6), 741–748. https://doi.org/10.1111/1469-7610.00770

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., & 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

Gallon, N., Harris, J., & van der Lely, H. (2007). Non-word repetition: An investigation of phonological complexity in children with grammatical SLI. *Clinical Linguistics & Phonetics, 21*(6), 435–455. https://doi.org/10.1080/02699200701299982

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

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)

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, 53*(6), 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), Article e89544. https://doi.org/10.1371/journal.pone.0089544

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

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)

Meir, N. (2017). Effects of specific language impairment (SLI) and bilingualism on verbal short-term memory. *Linguistic Approaches to Bilingualism, 7*(3–4), 301–330. https://doi.org/10.1075/lab.15033.mei

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

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

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

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.

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

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

Shaalan, S. (2020). Nonword repetition skills in Gulf Arabic–speaking children with developmental language disorder. *Journal of Speech, Language, and Hearing Research, 63*(11), 3700–3713. https://doi.org/10.1044/2020_JSLHR-20-00040

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, Language, and Hearing Research, 49*(2), 219–236. https://doi.org/10.1044/1092-4388(2006/019)

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

Taha, J., Stojanovik, V., & Pagnamenta, E. (2021). Nonword repetition performance of Arabic-speaking children with and without developmental language disorder: A study on diagnostic accuracy. *Journal of Speech, Language, and Hearing Research, 64*(7), 2750–2765. https://doi.org/10.1044/2021_JSLHR-20-00556

Thordardottir, E. (2008). Language-specific effects of task demands on the manifestation of specific language impairment: A comparison of English and Icelandic. *Journal of Speech, Language, and Hearing Research, 51*(4), 922–937. https://doi.org/10.1044/1092-4388(2008/068)

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, 56*(1), 103–122. https://doi.org/10.1044/1092-4388(2012/11-0080)

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.

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