



<b>Title</b>	<b>Diagnostic accuracy of nonword repetition and sentence repetition in Cantonese-speaking children with specific language impairment (SLI)</b>
<b>Author(s)</b>	<b>Leung, Yick-nam, Norberta; 梁翊嵐</b>
<b>Citation</b>	
<b>Issued Date</b>	<b>2010</b>
<b>URL</b>	<b><a href="http://hdl.handle.net/10722/173720">http://hdl.handle.net/10722/173720</a></b>
<b>Rights</b>	<b>Creative Commons: Attribution 3.0 Hong Kong License</b>

**Diagnostic accuracy of Nonword Repetition and Sentence Repetition in Cantonese-speaking  
Children with Specific Language Impairment (SLI)**

Student number: 2006640723

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2010.

**Diagnostic accuracy of Nonword Repetition and Sentence Repetition in Cantonese-speaking  
Children with Specific Language Impairment (SLI)**

Leung Norberta Yick Nam

A dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2010.

**Diagnostic accuracy of Nonword Repetition and Sentence Repetition in Cantonese-speaking  
Children with Specific Language Impairment (SLI)**

Leung Norberta Yick Nam

Abstract

The study investigated the diagnostic accuracy of nonword repetition (NWR) and sentence repetition (SR) tasks in Cantonese-speaking children with Specific Language Impairment (SLI). Thirteen children with language impairment (LI) and 17 of their typically developing (TD) peers were recruited from local kindergartens with integrated programs. The SLI group was a subsample of the LI group, which contained 9 children. All the children were compared on NWR and SR tasks. The use of i) clinical status, ii) the children's Hong Kong Cantonese Grammar (HKCG) subtest scores as the grouping variable was explored. The TD group scored significantly higher than the SLI group on NWR and SR when (ii) was used as the grouping variable. In terms of diagnostic accuracy, both NWR and SR failed to achieve a value of above 80% for both sensitivity and specificity. Theoretical and clinical implications of the results obtained are discussed.

Children with specific language impairment (SLI) often exhibit a significant and longstanding deficit in language development despite the fact that they do not have an accompanying problem such as hearing impairment, neurological damage, socioemotional disturbances or mental retardation (Leonard, 1998). There has been a great deal of speculation over the past decades regarding the cognitive mechanisms that underlie these children's language impairments. Researchers have reported that children with SLI perform less well than their typically developing age-matched (TDAM) peers across a range of working memory (WM) tasks (Dollaghan & Campbell, 1998; Vance, 2008). Before we discuss these findings, it is helpful to review the working memory framework on which these work are based.

According to Baddeley (2000), working memory is a multi-component system which comprises of several mental processes. These processes allow an individual to store and process information simultaneously. All sensory input, auditory or visual, must first be held momentarily in the short-term memory store, which is a component in the working memory, while further processing, such as accessing relevant information from long-term memory, takes place. This temporary storage eventually leads to the establishment and storage of permanent representations in the long-term memory, and this is how acquisition of knowledge occurs. However, the working memory capacity available for storage and processing is limited and has to be shared between these two functions (Just & Carpenter, 1992). There is a trade-off between the storage and processing functions when the task demands exceed one's processing capacity. It results in either (1) decay of some of the input material or (2) a reduction in the speed of computational processing. Children with SLI were found to have a reduced working memory capacity when compared to their TDAM peers, and this limitation in working memory has been shown to associate with their poor performance in vocabulary acquisition (Montgomery, 2002) and sentence

comprehension (Marton & Schwartz, 2003; Montgomery & Evans, 2009).

Nonword repetition (NWR) and sentence repetition (SR) tasks are measures of phonological working memory as they both involve temporary storage and simultaneous processing of the incoming material (Vance, 2008). NWR and SR tasks involve an exact repetition of nonsense words and sentences respectively. For English-speaking children, research has consistently shown that children with SLI performed significantly poorer than their TDAM peers in both NWR and SR tasks (Conti-Ramsden, 2003; Vance, 2008). The consistency of their difficulties in these tasks has led to investigations of NWR and SR as potential clinical markers for discriminating children with SLI from their TDAM peers.

Conti-Ramsden, Botting & Faragher (2001) examined the contributions of four potential psycholinguistic markers (i.e. a third person singular task, a past tense task, a NWR task and a SR task) for identifying eleven-year-old children with SLI. Results indicated that children with SLI scored significantly lower than their TDAM peers in both the NWR and SR tasks, and SR was the best clinical marker of SLI as it yielded high sensitivity and specificity values of 90% and 85% (at the 16<sup>th</sup> percentile cut-point) respectively, followed by NWR, with sensitivity and specificity values of 78% and 87% respectively.

In addition to the involvement of working memory, it has been suggested that performance in both NWR and SR are enhanced by redintegration, a process in which existing vocabulary and language knowledge in the long-term memory supports the successful recall of nonsense words and sentences respectively (Vance, 2008). For NWR, it was found that children were more able to repeat nonsense words that correspond to real words in terms of lexical and sublexical features (Gathercole, 1995; Dollaghan, Biber & Campbell, 1995). Regarding to lexical feature, nonwords that are more word-like

(e.g. 'prindle') are more accurately repeated than those that are less word-like (e.g. 'woogalamic'). In terms of sublexical feature, nonwords that contain syllables that resemble a real word (e.g. /'bleiməfət/) are more easily recalled than those that do not (e.g. /'bliməfət/). During NWR, part of the input material might have been forgotten due to the limited working memory capacity to support storage. The process of redintegration allows children to re-create the original stimuli or its close approximation based on their existing word and language knowledge. Although researchers working with English-speaking children stated that the NWR stimuli should be free of lexicality effects to avoid the contribution of redintegration, it was found that many of the stimuli used in their tests resemble English real words, such as /tei/ and /nai/ (Dollaghan & Campbell, 1998; Marton & Schwartz, 2003). Stokes, Wong, Fletcher & Leonard (2006) suggested that due to the nature of English, it is difficult to construct CV combinations that are unattested (i.e. CV structures that do not appear in the language), and hence unavoidably redintegration could have been involved in the repetition of nonword stimuli in the research reported. This may be one of the reasons why English-speaking children with SLI performed more poorly than their TDAM peers, as the former have poorer language knowledge to use redintegration and possibly also a reduced working memory capacity that limit their performance in the NWR task. For SR, research confirmed that children can recall more words if the sentences they are to repeat are meaningful rather than nonsense (Vance, 2008). When there are errors in repetition, the gist of the sentence is often preserved and the incomplete sentence tends to be substituted with synonyms rather than unrelated words (Alloway & Gathercole, 2005a). The authors concluded that the reasons for poor SR performance may be a result of reduced working memory, language deficit, or a combination of both. Children with SLI, who have a reduced working memory capacity and fewer linguistic resources to

support the process of redintegration, will be ‘doubly disadvantaged’ (Vance, 2008, p.31). To summarize, studies on the English-speaking children with SLI suggested that both NWR and SR require working memory and prior language knowledge to succeed.

Do Cantonese-speaking children with SLI show similar performance in the NWR and SR tasks as compared to the results obtained in English-speaking children? Stokes et al. (2006) compared the scores of 14 Cantonese-speaking children with SLI (aged 4;2 – 5;7), 15 of their TDAM peers (aged 4;1 – 6;9) and 15 typically developing younger (TDY) children (aged 2;11 – 3;6) in the NWR and SR tasks. Results indicated that SR but not NWR discriminates children with SLI from their TDAM peers, with sensitivity and specificity at 77% and 97% respectively. These results did not resemble the patterns in the English-speaking population as both tasks yielded high sensitivity and specificity in several English studies (e.g. Conti-Ramsden et al., 2001). Since NWR is where the English and Cantonese results differ, here we focus on those factors that might contribute to this variation. The NWR stimuli used in the Stokes et al. (2006) study consist of both IN and OUT nonwords: IN nonwords contain CV structures that appear in Cantonese (i.e. attested) whereas OUT nonwords do not (i.e. unattested). While it is possible to use the reintegration strategy (i.e. the stored sublexical knowledge) to reconstruct the incomplete IN nonwords, the authors suggested that successful recall of the OUT nonwords can only be achieved by ‘true repetition without calling on the long-term store’ as there is no existing linguistic knowledge to support their recall (Stokes et al., 2006, p.231). It was hypothesized that children with SLI, being less competent language users, should perform poorer than their TDAM peers in the repetition of IN nonwords. Results in IN nonwords showed that the TDAM group (74%) scored higher than the SLI group (67%), although the difference failed to reach statistical significance. The two groups however



performed at the same level for the OUT nonwords (45%). These findings suggested that children with SLI perform comparably to their TDAM peers in NWR when only phonological working memory skills are assessed (OUT nonwords), yet they are disadvantaged by their limitations in language skills when asked to repeat 'word-like' nonwords. However, the authors stressed that the results obtained were not statistically significant and this might be due to a relatively small sample size. Another possibility is about the nature of the NWR task in this study. Perhaps the NWR task was too difficult even for normal five-year-old children, as they showed an overall accuracy of only 68.35%. Task difficulty could have minimized a possible gap between the performance of children with SLI and their TDAM peers. Therefore, further investigation of the nonword repetition ability of children with SLI using another NWR task is warranted.

Hong Kong Cantonese Oral Language Assessment Scale (HKCOLAS, T'sou et al., 2006) is a standardized language assessment tool developed for children between 5;0 – 12;0. There are two NWR tasks in the test: one of them requires children to repeat pseudomorphemes whereas the other one asks them to repeat pseudosyllables. Pseudomorphemes are real Cantonese morphemes which are combined to form non-existing multisyllabic Cantonese word. Pseudosyllables are syllables that are constructed according to the phonotactic rule but are unattested in Cantonese. Comparing to the NWR stimuli used in Stokes et al. (2006), the pseudomorphemes are similar to the IN nonwords as both of them contain CV combinations that are attested in Cantonese, while the pseudosyllables are comparable to the OUT nonwords. Due to time constraint, only the pseudomorpheme subtest will be implemented in the present study. Moreover, it was found that the NWR task using pseudomorphemes has higher correlations with other language subtests in HKCOLAS (T'sou et al., 2006), which might therefore expose the underlying

impairments in children with SLI and facilitate identification. Similar to the NWR task used in the English-speaking population, it is possible for children to use redintegration in the repetition of pseudomorphemes. Therefore, if children with SLI performed poorly in the pseudomorpheme task, this study will add another piece of evidence to suggest that poor NWR for children with SLI is likely to be attributable to their poor word and language knowledge which in turn hinder their use of redintegration strategy in NWR. For the SR task, Stokes et al. (2006) reported that sentences with passives and aspects successfully identified children with SLI among their 5-year-old subjects using the 0-3 error scoring scheme. To examine the validity of their findings and the diagnostic accuracy of their SR task, the task must be carried out on a different sample of children.

The two research questions for the present study are:

1. Do Cantonese-speaking children with SLI perform differently from the children with typical language skills (TD) in the pseudomorpheme NWR task (T'sou et al., 2006) and the SR task (Stokes et al., 2006)?
2. What are the sensitivity and specificity of the pseudomorpheme NWR task (T'sou et al., 2006) and the SR task (Stokes et al., 2006) in differentiating children with SLI and their TD peers?

## **Method**

### *Participants*

Cantonese-speaking children with typical language (TD) skills and children with language impairment (LI) were recruited from local kindergartens with integrated programs and 30 children participated in this study. Seventeen of the children, aged 5;2 to 6;1 ( $M = 5;5$ ,  $SD = 3.04$  months), who were reported to have age-appropriate language development by their parents and teachers, were in the TD group. The other 13 children, aged 5;5 to 7;2 ( $M = 6;0$ ,  $SD = 5.75$  months), who had been previously diagnosed as

having language deficits by a speech-language pathologist and were receiving speech therapy services at their kindergartens by the time this study was carried out, were in the LI group. Among these 13 children in the LI group, 9 demonstrated language impairments with no report of oral-motor, hearing, cognitive, neurological or psychosocial dysfunction. They were loosely defined as SLI in this study. Typically a research classification of SLI requires formal assessment of the children's cognitive and hearing abilities, but such testing was not possible due to time and resource constraints. The remaining four children in the LI group had either a mild delay in cognitive development or autistic features. To summarize, the participants were divided into three groups for data analysis: i) typically developing children with age appropriate language development (TD,  $N = 17$ ), ii) children with language impairment (LI,  $N = 13$ ), iii) children with SLI (SLI,  $N = 9$ ).

#### *NWR Stimuli*

The Test of Nonword Repetition with Pseudomorphemes in HKCOLAS (T'sou et al., 2006) was administered. The test items consisted of 45 real Cantonese morphemes which had been combined to form non-existing Cantonese words with different syllable lengths, for example, 'kin1 set9' and 'ley5 tig1 phai3'. Each child must be able to repeat the whole syllable (including the phonemes and tones) correctly in order to score one point, and the maximum score for all items combined was 45. The syllable length ranged from one to nine.

#### *SR Stimuli*

The SR stimuli in the Stokes et al. (2006) study were used in the present study. Sentences with passives and aspect markers were included in the task. There were 8 sentences for each sentence type, so altogether there were 16 trials in the task. For the sentences with aspect markers, elements of subject,

verb, aspect marker, and object were included and an example was ‘go4go1 m4 siu2sum1 dou2 se2 zo2 bui1 sei2’ (Big brother has spilled the glass of water by accident). For the passive sentences, they contained the passive elements bei2, agent noun phrase and verb, and an example was ‘jau5 go3 nei2zai2 bei2 go3 naam4zai2 daa2’ (A girl was hit by the boy). In terms of scoring, the 0-3 point method used in the Stokes et al. (2006) study was adopted as their findings suggested that this method was the best for successful discrimination of children in the TDAM and SLI groups. A score of three points was given for complete sentence accuracy, two points for one error, one point for two or three errors and no points was awarded for four or more errors.

### *Procedure*

The study was carried out in two phases. In phase one, the Hong Kong Cantonese Grammar (HKCG) subtest in HKCOLAS (T’sou et al., 2006) was administered to all participants in order to assess their language abilities. The task instructions and the test items were presented via free-field speakers. The subtest consisted of four parts: the participants were required to point to the target picture for the first part, and they had to give verbal responses in the remaining three parts. All the responses were recorded in real time. In phase two, the participants were asked to perform the NWR and SR tasks. These two tasks were randomized so that some children completed the NWR task first and the others received the SR task first. All the stimuli were delivered via free-field speakers, and the participants were asked to listen and repeat exactly what they had heard. Trial items were presented first: there were two of them in the NWR task, and six (three for each sentence type) in the SR task. One repeat was allowed for the trial items if the participant’s response was incorrect. No feedback was given for all other trials. Each experimental item was presented once only. All the responses were audiorecorded for later transcription.

To avoid bias, the task administrators in phase two were blinded to the language status of the participants.

## **Results**

Results from the Kolmogorov-Smirnov and the Shapiro-Wilks tests ( $p < 0.05$ ) suggested that none of the variables showed a normal distribution. Non-parametric statistics were therefore used to explore the differences among the groups and the correlations between variables.

### **Grouping of participants**

The participants were classified into the TD and LI groups according to their clinical status. For the SLI group, they were the children in the LI group who had no report of oral-motor, hearing, cognitive, neurological or psychosocial dysfunction. We also explored the use of the children's performance in the HKCG subtest as an alternative in the classification of TD and SLI/LI groups. Children who scored below -1 SD in the subtest were classified into the LI group. The SLI group composed of children who scored below -1 SD in the subtest but had no accompanying problems. Since some of the LI children scored above -1 SD in the subtest while some of the TD children scored below the cut-off point, the groups were composed of different children when compared to those grouped according to the clinical status (although the number of participants in each group was similar). The number of participants in each group was summarized in Table 1.

Table 1.

*The number of participants in the TD, LI and SLI groups according to (i) the participants' clinical status; (ii) the participants' test scores in the HKCG subtest*

Gold standards			Number of participants		
(i) Clinical status	TD	17	LI	13	Clinical status group 1
					TD + LI = 30
(ii) HKCOLAS – HKCG	TD	18	LI	12	Clinical status group 2
					TD + LI = 26
(ii) HKCOLAS – HKCG	TD	17	SLI	9	HKCG group 1
					TD + LI = 30
(ii) HKCOLAS – HKCG	TD	17	SLI	9	HKCG group 2
					TD + SLI = 26

**NWR**

Means and standard deviations for the NWR task were derived for the TD group, the LI group as well as the SLI group. Table 2 summarizes these descriptive statistics below.

Table 2.

*Means (in percentages) of the three participant groups on the nonword repetition test.*

	Clinical status			HKCOLAS - HKCG		
	TD <sup>a</sup>	LI <sup>a</sup>	SLI	TD <sup>a</sup>	LI <sup>a</sup>	SLI <sup>b</sup>
Total % accuracy	38.17	26.50	29.38	39.75	23.15	25.93
	<i>12.77</i>	<i>14.41</i>	<i>14.43</i>	<i>12.77</i>	<i>11.07</i>	<i>10.06</i>

*Note.* Numbers in italics are standard deviations

<sup>a</sup>A significant difference between the TD group and LI group at  $p < .05$ . <sup>b</sup>A significant difference between the TD group and the SLI group at  $p < .05$

The Mann-Whitney U test was conducted on the NWR test scores to evaluate whether the means in the i) TD and LI groups and ii) TD and SLI groups were significantly different from the others. When clinical status was used as the grouping variable, the TD group scored significantly higher than the LI group (38.17% and 26.50% accuracy,  $U = 54$ ,  $p = 0.018$  (2-tailed). The effect size of 0.87 was indicative of a large effect (Cohen, 1988). There was no significant difference between the TD and the SLI groups. When the test score in HKCG was used as the grouping variable, significant differences were found in both the i) TD and LI groups ( $U = 34.5$ ,  $p = 0.02$  (2-tailed) and ii) TD and SLI groups ( $U = 31$ ,  $p = 0.014$  (2-tailed). The effect size of 1.37 and 1.16 for the HKCG group 1 and HKCG group 2 respectively suggested a large effect (Cohen, 1988).

## SR

Means and standard deviations for the SR task were derived for the TD group, the LI group as well as the SLI group. Table 3 summarizes these descriptive statistics below.

Table 3.

*Means (in percentages) of the three participant groups on the sentence repetition task.*

	Clinical status			HKCOLAS - HKCG		
	TD <sup>a</sup>	LI <sup>a</sup>	SLI	TD <sup>a</sup>	LI <sup>a</sup>	SLI <sup>b</sup>
Total % accuracy	77.33	53.20	57.64	81.60	44.79	51.16
	<i>18.40</i>	<i>33.18</i>	<i>25.73</i>	<i>16.03</i>	<i>28.27</i>	<i>21.45</i>

*Note.* Numbers in italics are standard deviations

<sup>a</sup>A significant difference between the TD group and LI group at  $p < .05$ . <sup>b</sup>A significant difference between the TD group and the SLI group at  $p < .05$

When clinical status was used as the grouping variable, it was revealed that the TD group outscored both the LI and SLI groups in the SR task, however, only the group difference between the TD and LI groups was significant ( $U = 61.5, p = 0.40$  (2-tailed)). The effect size was 0.98 which indicated a large effect (Cohen, 1988). When HKCG was used as the grouping variable, significant differences were found in both the i) TD and LI groups ( $U = 21, p = 0.00$  (2-tailed) and ii) TD and SLI groups ( $U = 18, p = 0.02$  (2-tailed)). The effect size of 1.75 and 1.65 for the HKCG group 1 and HKCG group 2 respectively suggested a large effect (Cohen, 1988).

### **Identifying LI and SLI**

One of the purposes of the present study was to evaluate the diagnostic accuracy of the NWR and the SR tasks in differentiating children with LI (and SLI) and their TD peers. To examine the diagnostic accuracy of the index measures (i.e. NWR and SR), we were interested in knowing how accurate these measures were in classifying individuals who have, and those who do not have the target behavior (i.e.



language impairment) with reference to the gold standard (Dollaghan, 2007). Gold standard is defined as the “best available method for establishing presence or absence of the target condition” (Bossuyt et al., 2003, p.8). In this study, we examined two different gold standards available to us: (i) the participant’s current clinical status as given to us by the referring agency and (ii) the participant’s test scores in the HKCG subtest. For the HKCG subtest, a standard score of below -1.0 was used as a cut-off criterion for identifying children who were positive (+) for language impairment.

To examine the extent to which the index measures provide accurate diagnostic information, metrics such as sensitivity, specificity, positive likelihood (LR+) as well as negative likelihood ratios (LR-) were used. Sensitivity concerns the proportion of participants who have language impairment according to the gold standard and who are also correctly identified by the index measures (i.e. true positive/[true positive + false negative]). Specificity is the proportion of participants who do not have language impairment based on the gold standard and who are also identified as negative by the index measures (i.e. true negative/[false negative + true negative]). Although there were no generally accepted thresholds for sensitivity and specificity, researchers generally accepted that values above 80% for both indices should be achieved if the findings were said to have diagnostic value (Plante & Vance, 1994). However, both sensitivity and specificity were highly susceptible to variations in the base rate of the diagnostic condition in the sample. Base rate is defined as the percentage of participants in the sample who are positive for the diagnostic condition of interest according to the gold standard (Dollaghan, 2007). Due to the limitation of sensitivity and specificity, metrics that were less affected by the prevalence of the disorder in the population, such as LR+ and LR-, were more preferred. LR+ reflects the probability that a positive (disordered) score on a test came from a person who has the disorder

whereas LR- indicates the probability that a score in the normal range came from a person who truly does not have the target disorder. The formulae for calculating LR+ and LR- were  $\text{sensitivity}/(1 - \text{specificity})$  and  $(1 - \text{sensitivity})/\text{specificity}$  respectively. According to Dollaghan (2007), a LR+ of  $\geq 10$  indicates that a positive (disordered) test score is very likely to have come from a person with the disorder and a LR- of  $\leq 0.1$  reflects that a negative (normal) test score is highly unlikely to have come from a person with the disorder. The diagnostic accuracy of the NWR and SR tasks were explored using these metrics and the results were summarized as below.

#### *NWR*

A standard score of -1.0 or below in the test of Nonword Repetition with Pseudomorphemes in the HKCOLAS (HKCOLAS, T'sou et al., 2006) was taken as a cut-off criterion to distinguish between children with and without language impairment in this study.

Table 4 summarizes the sensitivity, specificity, LR+ and LR- for the NWR task when (i) the participants' clinical status and (ii) the participants' test scores in the HKCG subtest were used as the gold standard.

Table 4.

*Sensitivity, specificity, LR+ and LR- of the NWR task*

	Gold standard			
	Clinical status		HKCG	
	TD and LI	TD and SLI	TD and LI	TD and SLI
	groups	groups	groups	groups
Sensitivity	53.8% (7/13)	55.6% (5/9)	50% (6/12)	44.4% (4/9)
Specificity	94.1% (16/17)	94.1% (16/17)	83.3% (15/18)	82.3% (14/17)
<i>Overall accuracy</i>	<i>76.7% (23/30)</i>	<i>80.8% (21/26)</i>	<i>70% (21/30)</i>	<i>69.2% (18/26)</i>
LR+	9.12	9.49	2.99	2.51
LR-	0.491	0.468	0.6	0.68

The specificity of the NWR task was high (> 80%) when either i) clinical status or ii) HKCG was used as the gold standard. The corresponding LR- values were moderately negative, which indicated that a negative test score in the NWR task was suggestive but not sufficient to rule out the disorder. The sensitivity of the NWR task was low (< 60%) when the abovementioned gold standards were used. The corresponding LR+ values were moderately positive, which suggested that a child who scored positive in the NWR task was likely to have language impairment.

### SR

Since there was no normative data for the SR task in the Cantonese-speaking population at the time this study was carried out, the means and the standard deviations ( $M = 73.47$ ,  $SD = 8.28$ ) obtained from the Stokes et al. (2006) study were used as a standard to classify children as TD or LI in the present

study. Children who scored -1 SD of the mean (i.e. 65.19) in the SR task were classified as LI.

Table 5 summarizes the sensitivity, specificity, LR+ and LR- for the SR task when (i) the participants' clinical status and (ii) the participants' test scores in the HKCG subtest were used as the gold standard.

Table 5.

*Sensitivity, specificity, LR+ and LR- of the SR task*

	Gold standard			
	Clinical status		HKCOLAS- HKCG	
	TD and LI	TD and SLI	TD and LI	TD and SLI
	groups	groups	groups	groups
Sensitivity	61.5% (8/13)	66.7% (6/9)	75% (9/12)	77.8% (7/9)
Specificity	76.5% (13/17)	76.5% (13/17)	77.8% (14/18)	76.5% (13/17)
<i>Overall accuracy</i>	<i>70% (21/30)</i>	<i>73.1% (19/26)</i>	<i>76.7% (23/30)</i>	<i>76.9% (20/26)</i>
LR+	2.62	2.84	3.38	3.30
LR-	0.503	0.44	0.32	0.29

As shown in the table above, the sensitivity and specificity of the SR task were slightly higher when the participants' test scores in the HKCG subtest were used as the gold standard, although both indices failed to achieve a value of 80% or above. In terms of the likelihood ratios, the LR+ ratios in the TD and LI groups as well as the TD and SLI groups were in the range of 2 - 3 which indicated that a positive test score was merely suggestive but insufficient to diagnose a disorder (Dollaghan, 2007). For the LR- ratios, the results ranged from 0.29 to 0.503 which reflected that a negative test score was

inadequate to rule out the disorder.

### Relationships among variables

Relationships among the test scores in HKCOLAS, NWR and SR in the different groups of participants were explored using the Spearman's rank correlation. Table 6 summarizes the details below.

Table 6.

#### *Relationship between HKCOLAS, NWR and SR across groups*

	TD	LI	SLI
NWR – SR	0.782*	0.687*	0.795*
HKCOLAS – NWR	0.645*	0.548*	0.374
HKCOLAS – SR	0.774*	0.727*	0.619*

\* $p < 0.05$  (2-tailed)

There were significant correlations between NWR and SR across groups, with the strongest correlation found in the SLI group ( $r = 0.795, p = 0.01$ ). Relationships among the index measures (NWR and SR) and one of the gold standards (HKCOLAS) were examined separately in each group. For the TD group, both NWR ( $r = 0.645, p = 0.005$ ) and SR ( $r = 0.774, p = 0.001$ ) were significantly correlated with the results in HKCOLAS. For the LI group, SR showed high correlations with HKCOLAS ( $r = 0.727, p = 0.005$ ) while the correlation between NWR and HKCOLAS was moderate ( $r = 0.548, p = 0.043$ ). For the SLI group, there was significant correlation between SR and HKCOLAS ( $r = 0.619, p = 0.007$ ) but the correlation between NWR and HKCOLAS was statistically significant.

### Overall summary

When the participants were classified into the TD, LI and SLI groups according to their clinical

status, significant differences were only found between the TD and LI groups in both NWR and SR. For the TD and SLI groups, although the former scored higher than the latter group in both tasks, the differences were not statistically significant. When they were grouped according to their HKCG subtest scores, significant differences were found between the TD and the LI group as well as between the TD and the SLI group in NWR and SR.

In terms of diagnostic accuracy, both NWR and SR failed to achieve a value of above 80% for both sensitivity and specificity. Therefore, using these tasks to distinguish between LI (and SLI) and their TD peers must be further evaluated.

### **Discussion**

The present study aimed to compare the performance of Cantonese-speaking children with SLI and their TD peers on the NWR and SR tasks, and to examine the diagnostic accuracy of these tasks. In this study, the SLI group was a subsample of the LI group. The LI group was more inclusive, whereas the SLI group included only those children with LI who had no parent or teacher report of hearing and cognitive deficits. We will also report our findings on the LI group as a whole where appropriate.

#### **Nonword repetition (NWR) and Sentence repetition (SR)**

Stokes et al. (2006) compared 14 children with SLI with 15 typically-developing (TD) peers on the NWR and SR tasks. For NWR, their TD group scored higher than the SLI group in the IN nonwords but the difference was not statistically significant. The IN nonwords used in the Stokes et al. (2006) study were similar to the pseudomorphemes used in this study's NWR task in which it was possible for children to use the strategy of redintegration to re-create the original stimuli or its close approximation based on their existing word and language knowledge. Similar results were obtained in the present study:

the TD group outscored the SLI group but the group difference was not statistically significant. Unlike the Stokes et al. (2006) study, however, there were no significant differences between the TD and the SLI groups in the SR task. At first glance, the results of the NWR and SR tasks did not suggest that a working memory capacity limitation underpinned language impairment in Cantonese-speaking children with SLI (at least in this cohort) as they did not perform significantly lower than their TD peers.

However, there are alternative interpretations to this finding.

#### *Participant characteristics*

All the participants were recruited from local kindergartens with an integrated program. Children in the LI (included those with SLI) group were those who had been diagnosed by their speech therapists as having language impairments and were receiving speech therapy services. However, it was found that some of these children scored within the normal range in the HKCG subtest, and the NWR and SR tasks when -1 SD was used as the cut-off criterion. Due to time constraint and limited resources, we were not able to complete the testing required and collect enough evidence for a more definite diagnosis on their language status. Moreover, no information about their latest language test scores was available from the agency who referred them to us. Therefore, it was possible that some of the children with LI (and SLI) might have already attained age-appropriate language skills at the time of this study. As a result, the mean differences between the TD group and the LI (and SLI) groups in both the NWR and SR tasks might have been minimized. For children in the SLI subgroup, the full battery of assessment, including hearing and cognitive, was not carried out to verify they were truly SLI as in Stokes et al. (2006) and therefore these children might not be representative of the SLI condition. This sampling difference might be one reason why the results in the present study differed from those reported in

Stokes et al. (2006).

The mean age of the LI and SLI groups was significantly higher than the TD group. Despite the effort in trying to restrict the age of participants to 5;0 - 6;0 during recruitment, it was reported that some of the children in the LI and SLI groups repeated the last year of their study at kindergarten and therefore their age was higher than the TD group. If these children were excluded, an extremely small sample size would have been resulted. Therefore, they were still included in the study. Stokes et al. (2006) reported that both NWR and SR were developmentally sensitive for Cantonese-speaking children and older children should perform significantly better than younger children. With the LI (SLI children included) group being older than the TD group in the present study, the possible group differences might have minimized and thus no significant results were obtained.

Due to the possible effect of the time gap between the clinical status and the index measures, the use of the children's HKCG subtest scores as the grouping variable was also explored. Interestingly, significant differences were found between the TD and the SLI group as well as the TD and the LI group in both NWR and SR with this new grouping criterion. However, the results must be interpreted with caution as only one subtest in HKCOLAS was administered and it required that the child failed at least 2 out of 6 subtests in order to be diagnosed to have language impairment. The subtest alone could not be used as a diagnostic tool. Yet this piece of information leads us to question about how valid the clinical statuses of the children with LI (and SLI) were. When the participants were grouped according to their performance in the HKCG subtest, unlike the Stokes et al. (2006) study, significant differences were found between the TD and the SLI group in NWR when the pseudomorpheme NWR test was used. Although it was possible for children to deploy the strategy of redintegration in both IN nonwords



and pseudomorphemes repetition tasks, they differed in several aspects which might contribute to the differences in the results obtained. In terms of syllable length, the IN nonword test contains trials with a maximum of four syllables while the pseudomorpheme NWR test has longer items ranging from five to nine syllables. Marton & Schwartz (2003) reported that the children with SLI in their study showed a greater decline in the NWR accuracy than their TDAM peers as the syllable length increased. The authors suggested that this was probably a result of an increase in working memory demands for repeating longer syllables. With the use of the pseudomorpheme NWR test in the present study, the deficit of the SLI group in NWR might have become more evident when compared to the TD group as longer syllables were included in this test. In future studies, it would be interesting to perform the pseudomorpheme NWR test on a group of more representative, homogeneous sample of children with SLI and their TD peers to see if any significant group differences were found.

Results from the present study lead us to think about the theoretical issues behind: What exactly do these repetition tasks measure? For a number of years, researchers working with English-speaking children suggested that both NWR and SR are measures of phonological working memory (Gathercole & Baddeley, 1990). Moreover, the process of redintegration, in which existing word and language knowledge support temporary storage, is likely to be involved in both tasks (Vance, 2008). Although both tasks depend on phonological working memory and language abilities, some researchers argued that the degrees of involvement of these skills vary according to the nature of task. Conti-Ramsden et al. (2001) suggested that NWR has a greater involvement of phonological working memory while SR draws on more prior linguistic knowledge. Their findings were supported by significant correlation found between the SR task and the linguistic measures in their SLI group. For Cantonese-speaking

children, Stokes et al. (2006) reported that the NWR scores were not related to any language scores in their SLI group while the SR scores were moderately correlated with receptive grammar scores.

Moreover, there was no significant correlation between NWR and SR for the TD, SLI as well as TDY (typically developing younger) groups. The results obtained in the present study were slightly different from Stokes et al. (2006). First of all, moderate to strong correlations were found between NWR and SR across the TD, SLI and LI groups. Involvement of similar skills in both tasks could have contributed to these significantly high correlations. Apart from phonological working memory, it was possible for children to deploy the strategy of redintegration in both the pseudomorpheme NWR and the SR tasks. However, redintegration could be used for only half of the NWR items (i.e. IN nonwords) in Stokes et al. (2006). The difference in the skills required might have reduced the correlation between NWR and SR in their study. Secondly, NWR was found to be moderately correlated with the score of HKCG subtest in the TD and LI groups, and mildly correlated with the SLI group. Conti-Ramsden et al. (2001) suggested that NWR might tap on single-word processing skills. The correlation between the pseudomorpheme NWR task and the HKCG subtest might imply that single-word processing skills used in NWR, such as decoding individual words, are required in comprehension of grammar. The weaker correlation in the SLI group might be resulted from a small sample size. Thirdly, SR showed moderate to strong correlations with the HKCG subtest across the groups. This finding might suggest that SR taps on the language knowledge similar to those required for comprehension of grammar. Although the data obtained in the present study cannot provide a firm answer to the question of the underlying deficits tapped by SR and NWR, these tasks might not be as language independent as Stokes et al. (2006) suggested.

### **Diagnostic accuracy of the NWR and SR tasks**

It was no surprise that the sensitivity and specificity of the NWR as well as the SR tasks were low in the present study as no significant group difference was found when clinical status was used as the grouping variable. Stokes et al. (2006) reported that the sensitivity and specificity of the SR task were 77% and 97% respectively. No information was available for the sensitivity and specificity of their NWR task. There are two possible factors that might contribute to the low sensitivity and specificity of both tasks in the present study: (i) the choice of the gold standard and (ii) sample size. The use of clinical status and the score in the HKCG subtest as the gold standard was explored. Both gold standards were bounded by some limitations, and if the children were classified wrongly in the first place, the sensitivity and specificity of the index measures would not be valid. As mentioned above, the clinical status of the participants was not fully and formally assessed for this study. It was found that many of the children in the LI (and SLI) group scored within the normal range in all the tests if  $-1.25$  SD was used as the cut-off point. In view of the mismatch between the participants' clinical status and their scores obtained in the tests, a more conservative cut-off point (i.e.  $-1$  SD) was used for each test to minimize the gap. For HKCOLAS, as only the HKCG subtest was administered in the present study, children might not have language impairment even if they failed in the subtest. Therefore, the results must be interpreted with caution. Consistent with the Stokes et al. (2006) study, the SR task was found to have better diagnostic accuracy than the NWR task in the present study as both sensitivity and specificity were approaching 80% regardless of the gold standard adopted. The consistency of high sensitivity and specificity across studies leads us to think about the use of this SR task as a clinical measure. The SR stimuli involved sentences with passive constructions and aspect markers. Cantonese-speaking children with SLI have

been found to be less capable than their TDAM peers in the production of passives (Leonard, Wong, Deevy, Stokes & Fletcher, 2006). For aspectual forms, although the Fletcher, Leonard, Stokes & Wong (2005) study revealed that their children with SLI were less able to produce the targeted aspect markers, Stoke & Fletcher (2003) found that the SLI group and their TDAM peers were similar in the ability to encode aspect in their repetition task. When choosing the SR stimuli, structures which unveil the difficulties that children with SLI have would be more appropriate as they can distinguish this group of children from their TDAM peers. In this regard, sentences with aspect markers might not be a suitable choice. Wong, Leonard, Fletcher & Stokes (2004) found that Cantonese-speaking children with SLI performed poorer in the production of who-object questions than their TDAM peers and younger peers. Future study might explore the use of both passives and who-object questions as the SR stimuli in the repetition task. If this is the case, researchers have to ensure that the SR stimuli are sufficiently complex to avoid ceiling effect of the typically developing children, and the stimuli must be constructed with appropriate syllable length so that children with SLI will not be disadvantaged.

Apart from the choice of the gold standard, a small sample size might also affect the degrees of sensitivity and specificity of the index measures. The effect is more apparent on the SLI group as it included only nine participants. As we can see, a small change in the number of participant correctly identified will lead to a great change in the indices, for example, from four out of nine participants to seven out of nine participants will result in an increase of sensitivity from 44.4% to 77.8%. Therefore, a larger sample is needed in order to measure the diagnostic accuracy of the tests.

## **Conclusion**

Although no concrete conclusions can be made from the results obtained in the present study due

to the limitations mentioned before, it is suggested that the NWR and SR tasks might not be as language independent as Stokes et al. (2006) suggested. Moreover, future work might explore the performance of Cantonese-speaking children with SLI and their TDAM peers with a different set of NWR and SR tasks. For NWR, both pseudomorphemes and pseudosyllables in HKCOLAS can be used so as to examine the degree of involvement of working memory and redintegration in NWR. For SR, sentences with passives and who-object questions can be included.

## References

- Alloway, T. P., & Gathercole, S. E. (2005a). Working memory and short-term sentence recall in young children. *European Journal of Cognitive Psychology, 17*, 207-220.
- Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417-422.
- Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P.P., Irwig, L.M., Lijmer, J.G., Moher, D., Rennie, D., de Vet H.C.W. (2003). Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *American Journal of Roentgenology, 181*, 51-56.
- Conti-Ramsden, G (2003). Processing and linguistic markers in young children with specific language impairment (SLI). *Journal of Speech, Language, and Hearing Research, 46*, 1029-1037.
- Conti-Ramsden, G, Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry, 42*, 741-748.
- Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore: Paul H. Brookes Publishing Co.
- Dollaghan, C., Biber, M. E., & Campbell, T. E. (1995). Lexical influences on nonword repetition. *Applied Psycholinguistics, 16*, 211-222.
- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and children language impairment. *Journal of Speech, Language, and Hearing Research, 41*, 1136-1146.
- Fletcher, P, Leonard, L. B., Stokes, S. F., & Wong, A. M. Y. (2005). The Expression of Aspect in Cantonese-Speaking Children With Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 48*, 621-634.

Gathercole, S. E. (1995). Is nonword repetition a test of phonological memory or long-term knowledge?

It all depends on the nonwords. *Memory and Cognition*, 23, 83-94.

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, 336-360.

Gaulin, C., & Campbell, T. (1994). Procedure for assessing verbal working memory in normal

school-age children: Some preliminary data. *Perceptual and Motor Skills*, 79, 55-64.

Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working

memory. *Psychological Review*, 99, 122-149.

Leonard, L. B. (1998). *Children with specific language impairment*. Cambridge, MA: MIT Press.

Leonard, L. B., Wong, A. M. Y., Deevy, P., Stokes, S. F., & Fletcher, P. (2006). The production of

passives by children with specific language impairment: Acquiring English or Cantonese. *Applied Psycholinguistics*, 27, 267-299.

Marton, K., & Schwartz, R. G. (2003). Working memory capacity and language processes in children

with specific language impairment. *Journal of Speech, Language, and Hearing Research*, 46, 1138-1153.

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, 77-91.

Montgomery, J. W., & Evans, J. L. (2009). Complex sentence comprehension and working memory in

children with specific language impairment. *American Journal of Speech, Language, and Hearing Research*, 52, 269-288.

- Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25*, 15-24.
- Stokes, S. F., & Fletcher, P. (2003). Aspectual forms in Cantonese children with specific language impairment. *Linguistics, 41*(2), 381-405.
- 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*, 219-236.
- T'sou, B., Lee, T., Tung, P., Man, Y., Chan, A. To, C., Cheung, P., Ng, A., & Chan, Y. (2006). Hong Kong Cantonese Oral Language Assessment Scale. Hong Kong: City University of Hong Kong.
- Wong, A. M. Y., Leonard, L. B., Fletcher, P., & Stokes, S. F. (2004). Questions Without Movement: A Study of Cantonese-Speaking Children With and Without Specific Language Impairment. *Journal of Speech, Language, and Hearing Research, 47*, 1440-1453.
- Vance, M. (2008). Short-term memory in children with developmental language disorder. In C. F. Norbury, J. B. Tomblin, & D. V. M. Bishop (eds). *Understanding developmental language disorders*. New York: Psychology Press.



### **Acknowledgments**

I would like to express my sincere gratitude to my dissertation supervisor, Dr. Anita Wong, for her guidance and support throughout this dissertation project. I very much appreciate all the suggestions and comments I have received from her.

I also want to show my appreciation to the St. Vincent De Paul Nursery School, Sisters of the Immaculate Heart of Mary Gospel Sau Mau Ping Child Care Centre Cum Kindergarten and Shun Tin Rhenish Nursery for their support. Special thanks are extended to the children and their families.

Lastly, I wish to thank my family and friends for their continuous help and friendship. I am especially grateful to my parents for their unconditional support and encouragement.

## Appendix

## Sentence Repetition Task (Stokes, Wong, Fletcher &amp; Leonard, 2006)

## A. Aspect marker

## Trial items

## 1. 我哋去公園

Ngo5	Dei6	Heoi3	Gung1	Jyun2
------	------	-------	-------	-------

## 2. 細佬喺房度讀書

Sai3	Lo2	Hai2	Fong2	Dou6	Duk6	Syu1
------	-----	------	-------	------	------	------

## 3. 黃老師今日好靚

Wong4	Lou5	Sil	Gam1	Jat6	Hou2	Leng3
-------	------	-----	------	------	------	-------

## Experimental items

## 1. 個男仔好大力擦緊對鞋

Go3	Naam4	Zai2	Hou2	Daai6	Lik6	Caat3	Gan2	Deoi3	Haai4
-----	-------	------	------	-------	------	-------	------	-------	-------

## 2. 紮辮個女仔整跌咗本書

Zaat3	Bin1	Go3	Neoi5	Zai2	Zing2	Dit3	Zo2	Byun2	Syu1
-------	------	-----	-------	------	-------	------	-----	-------	------

## 3. 哥哥唔小心倒瀉咗杯水

Go4	Go1	M4	Siu2	Sum1	Dou2	Se2	Zo2	Bui1	Seoi2
-----	-----	----	------	------	------	-----	-----	------	-------

## 4. 戴眼鏡個男仔打緊鞭鞦

Daai3	Ngaan5	Geng2	Go3	Naam4	Zai2	Daa2	Gan2	Cin1	Cau1
-------	--------	-------	-----	-------	------	------	------	------	------

## 5. 個女仔好大聲講緊故事

Go3	Neoi5	Zai2	Hou2	Daai6	Seng1	Gong2	Gan2	Gu3	Si6
-----	-------	------	------	-------	-------	-------	------	-----	-----

## 6. 著褲個男仔打爛咗個碗

Zeok3	Fu3	Go3	Naam4	Zai2	Daa2	Laan6	Zo2	Ho3	Wun2
-------	-----	-----	-------	------	------	-------	-----	-----	------

## 7. 個女仔好快咁洗咗件衫

Go3	Neoi5	Zai2	Hou2	Faai3	Gam2	Sai2	Zo2	Gin6	Saam1
-----	-------	------	------	-------	------	------	-----	------	-------

## 8. 著短裙個女仔砌緊砌圖

Zeok3	Dyun2	Kwan4	Go3	Neoi5	Zai2	Cai3	Gan2	Cai3	Tou4
-------	-------	-------	-----	-------	------	------	------	------	------

## B. Passive

## Trial items

1. 佢鍾意食飯

Keoi5	Zung1	Ji3	Sik6	Fan6
-------	-------	-----	------	------

2. 我想同媽媽行街

Ngo5	Soeng2	Tung4	Maa1	Mi4	Haang4	Gaai1
------	--------	-------	------	-----	--------	-------

3. 學校有好多嘢玩

Hok6	Haau6	Jau5	Hou2	Do1	Je5	Waan2
------	-------	------	------	-----	-----	-------

## Experimental items

1. 有個女仔俾個男仔打

Jau5	Go3	Neoi2	Zai2	Bei2	Go3	Naam4	Zai2	Daa2
------	-----	-------	------	------	-----	-------	------	------

2. 妹妹俾哥哥搶咗把梳

Mui4	Mui2	Bei2	Go4	Go1	Ceong2	Zo2	Baa2	So1
------	------	------	-----	-----	--------	-----	------	-----

3. 戴帽個女仔俾隻貓咬

Daai3	Mou2	Go3	Neoi5	Zai2	Bei2	Zek3	Maau1	Ngaau5
-------	------	-----	-------	------	------	------	-------	--------

4. 姐姐俾三隻斑點狗吠

Ze4	Ze1	Bei2	Saam1	Zek3	Baan1	Dim2	Gau2	Fai6
-----	-----	------	-------	------	-------	------	------	------

5. 長頭髮個男仔俾隻狗追

Ceong4	Tau4	Faat3	Go3	Naam4	Zai2	Bei2	Zek3	Gau2	Zeoi1
--------	------	-------	-----	-------	------	------	------	------	-------

6. 有個男仔俾個女仔踢

Jau5	Go3	Naam4	Zai2	Bei2	Go3	Neoi5	Zai2	Tek3
------	-----	-------	------	------	-----	-------	------	------

7. 妹妹俾兩隻黑花貓/zit1/

Mui4	Mui2	Bei2	Leong5	Zek3	Hak1	Faa1	Maau1	Zit1
------	------	------	--------	------	------	------	-------	------

8. 姐姐俾弟弟攞咗個袋

Ze4	Ze1	Bei2	Dai4	Dai2	Lo2	Zo2	Go3	Dai2
-----	-----	------	------	------	-----	-----	-----	------