Title	Examining the diagnostic accuracy of a delayed sentence imitation test for Cantonese-speaker preschoolers
Author(s)	Ng, Ka-man; 吳嘉雯
Citation	
Issued Date	2009
URL	http://hdl.handle.net/10722/173679
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Examining the diagnostic accuracy of a delayed sentence imitation test for Cantonese-speaker preschoolers

Ng Ka Man

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, 2009.

Abstract

The study investigated the diagnostic accuracy of a delayed sentence imitation (DSI) test in identifying 3-year to 5-year-11-month-old Cantonese-speaking children with language impairment. Ten children with a diagnosis of language impairment and 10 typically developing children were tested by the DSI test and a reference test by two independent examiners blinded to the results of the tests. The DSI test was found to have fair accuracy with sensitivity and specificity of 80%. The children's short-term memory was measured using a digit span test. A significant and moderately positive correlation was revealed between their performance on the DSI and the digit span test. Recommendations were suggested for further study to validate the DSI test as a locally applicable diagnostic tool and to investigate the relationship between working memory and language processing in Cantonese-speaking preschoolers.

Elicited imitation (EI) refers to the "repetition of a model sentence presented in a context calling for imitation" (Slobin & Welsh, 1973, p. 486). Slobin and Welsh (1973) suggest that it can be a "fruitful" (p. 496) method to examine a child's knowledge of grammatical structures. The theoretical rationale for this suggestion is that as the child imitates a sentence upon request after an adult model, s/he has to recognize and re-construct its underlying structure. The child's linguistic system is therefore involved in an active filtering process, and hence EI reveals the child's knowledge, or lack of knowledge, of the structure, especially when errors are demonstrated during imitation. The validity of EI as a language assessment procedure is supported by research which has shown that children's EI scores significantly correlated with their performance in tests of general verbal ability (Gallimore & Tharp, 1981). In fact, EI is one of the more commonly-used procedures in formal tests on expressive language, including Test of Language Development (TOLD-P:4; Hammil & Newcorner, 2008) and Clinical Evaluation of Language Fundamentals (CELF-4; Semei, Wiig, & Secord, 2006).

Recent works have examined the potential use of EI for the diagnosis of specific language impairment (SLI) and the results are positive. Conti-Ramsden, Botting, and Faragher (2001) compared the performance of 160 children with a history of SLI and 100 typically developing age-matched (TDAM) children, aged 11 years, on four tasks, including a third person singular task, a past tense task, a nonword repetition (NWR) task, and an EI task. The EI task used in that study was the recalling sentences subtest of the Clinical Evaluation of Language Fundamentals-Revised (CELF-R; Semel, Wiig, & Secord, 1994), which

required the children to repeat a model sentence word in word immediately after they heard it from the adult examiner. The series of sentences in the recalling sentences subtest of the CELF-R increase in length and complexity (Semel, Wiig, & Secord, 1994). Results revealed that the EI task was the most useful marker for discriminating children with SLI from their TDAM peers, with a sensitivity of 90%, a specificity of 85% using a -1 S.D. (16^{th} percentile) cut-off point. In addition, the EI task was also found to be able to identify children with a history of SLI, whose language scores now fell within the normal range and language impairments (LI) resolved (Conti-Ramsden et al., 2001). In the study by Botting and Conti-Ramsden (2003), four groups of children with LI (autistic spectrum disorders, and two groups of primary pragmatic language impairment, and SLI) were tested on three of the marker tasks employed in Conti-Ramsden et al.'s (2001) study. The EI task was also found to be more efficient than the NWR task and the past tense task in discriminating between SLI group and the other LI groups and their TDAM peers.

Poor performance in EI tasks is also a psycholinguistic marker for SLI in Cantonese-speaking children (Stokes, Wong, Fletcher, & Leonard, 2006). In Stokes et al.'s (2006) study, performance of 14 children with SLI, aged 4;02 to 5;07 (year; month), was compared with 15 TDAM children and 15 typically developing younger (TDY) children matched with mean length of utterance on a NWR task and an EI task. The EI task in the study consisted of only two sentence types (aspect marker and passive) and all sentences were of similar length. Results indicated that the EI but not the NWR task discriminated children with SLI from their TDAM peers. The EI task showed a sensitivity of 77%, a

specificity of 97%, a positive likelihood ratio (LR+) of 25.66, and a negative likelihood ratio (LR-) of 0.24. The diagnostic accuracy of the EI task in Stokes et al.'s study has been considered as clinically acceptable according to the criteria suggested by Dollaghan (2004), in which LR+ should be 10 or more and LR- should be 0.2 or less. The EI task however was not acceptable according to Plante and Vance's criteria (1994) since its sensitivity was below 80%. Plante and Vance (1994) have recommended that sensitivity and specificity above 90% indicates a good diagnostic accuracy while between 80-89% indicates a fair diagnostic accuracy. Sensitivity is the proportion of which children who have a diagnosis of LI according to a reference standard are identified as being so according to the results in the EI task. Specificity is the proportion of which children who are diagnosed as free of LI according to a reference standard are identified as being so according to the results in the EI task. LR+ is an index reflecting the confidence that a tested positive case is truly language impaired. LR- is an index reflecting the confidence that a tested negative case is truly free of LI (Dollaghan, 2007).

Although poor performance in EI has been reported to be a potential clinical marker for LI in Cantonese-speaking children (stokes et al., 2006), further studies are required to ascertain this finding with a wider range of sentence types. In addition to the one reported in Stokes et al. (2006), there is another Cantonese EI task developed for children from a wider age range. In this delayed sentence imitation (DSI) test, Mok (1995) developed the test using part of Cheung's (1993) original 104 sentence stimuli that were designed to examine 26 different Cantonese grammatical structures. In the DSI test, children first heard two sentences

in each trial and then asked to repeat the first sentence. Mok (1995) administered this test to six groups of 20 Cantonese-speaking children, in six-monthly intervals, between the ages of 3;0 and 5;11, and reported on their performance. Good construct validity has been reported for the DSI test (Mok, 1995).

Before the DSI test can be used for the assessment of children suspected with language disorders, the DSI test needs to be examined for its psychometric properties (McCauley & Swisher, 1984). The DSI test only met six of the ten, including description of the standardization sample, item analysis, reporting of means and standard deviations of scores, test-retest reliability, inter-rater reliability, and description of test administration procedures. Criteria that concern test validity, including concurrent validity and predictive validity, were not reported. Apart from a psychometric review, evidence-based practice (EBP) is another approach to critically appraise a diagnostic measure (Dollaghan, 2007). The EBP framework is originated from evidence-based medicine, which requires integration of "individual clinical expertise with the best available external clinical evidence from systematic research" (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996, p. 71). With the use of the EBP approach, a diagnostic measure being evaluated (i.e. index measure) and other established tests (i.e. reference standard) will be administered to a recognizable and representative group of participants with blinding of the examiners. Results obtained from the index measure will then be compared with the defined reference standard to examine the diagnostic accuracy of the index measure (Dollaghan, 2007). Given earlier evidence from English-speaking as well as Cantonese-speaking children, it is however plausible that this

DSI test might also show good diagnostic accuracy. The primary purpose of this study therefore is to examine the accuracy of Mok's (1995) DSI test in identifying Cantonese-speaking children with LI.

There is evidence that some English-speaking children with LI also have deficits in short-term memory (STM) and working memory (WM) (Archibald & Gathercole, 2006). STM refers to the transitory storage and recall of unprocessed materials and an example task that assesses STM is forward digit recall. WM refers to the storage and recall of material that has been processed and an example task for assessing WM is backward digit span (Vance, 2008). An interactive and bidirectional relationship between a child's linguistic knowledge and WM system has been proposed to explain the coexistence of LI and memory deficits (Montgomery, 2002). Leonard et al. (2007) also investigated the relationship between LI and processing factors such as WM and processing speed. It was reported that the processing factors, especially verbal WM, contributed to the prediction of children's language test scores. Montogmery (2002) described the two major models of the verbal WM, which are the phonological loop model (Baddeley, 1986) and the capacity theory of comprehension (Just & Carpenter, 1992). In the phonological loop model, Verbal WM is capacity-limited system with multiple components, including a central executive, a phonological loop, and a visual-spatial sketchpad (Montgomery, 2002). Verbal VM according to Baddeley's model is known as the phonological WM as it "relates to the amount of acoustic-phonetic material a listener can accurately phonologically encode and temporarily hold in the phonological store at any given moment" (Montogmery, 2002, p. 79). Just and Carpenter's model of verbal WM

is a single, capacity-limited system for both storage and processing functions and is referred as functional WM (Montgomery, 2002).

Vance (2008) suggested that EI requires interactions between STM and the language processing system. She suggested that as a child imitates a model sentence upon request, s/he has to keep all the words and their orders s/he has heard in memory while processing the linguistic structures of the sentence. Baddeley (2000) has also accounted for this relationship using the episodic buffer component in his working memory model. Episodic buffer is a temporary storage system with a limited capacity for information that is integrated from the components of the WM (phonological loop and visual-spatial sketchpad) and long-term memory (Baddeley, 2000). Despite this evidence from English-speaking children, the group of Cantonese-speaking children with SLI in Stokes et al. (2006) performed significant worse than their age controls on the EI task, but not on a task on non-word repetition, a measure of phonological WM. In addition, little is known about the relation between STM and LI in Cantonese-speaking cases. Such an inconsistency in the findings concerning WM and lack of studies about STM warrants a follow-up investigation on STM in Cantonese-speaking children.

In summary, the primary aim of this study was to investigate the diagnostic accuracy of Mok's (1995) DSI test for Cantonese-speaking preschoolers with LI in Hong Kong. The second aim was to explore the relationship between children's STM and performance on the DSI test.

Method

Participants

Twenty-one Cantonese-speaking children were invited to participate in this study initially. Eleven of them were between 3;00 (year;month) and 5;05, with a mean age of 4;06. Among these 11 children, six were recruited from a special child care centre, four from the Internal Child Clinic of the Speech and Hearing Division, University of Hong Kong, and one from a neighborhood preschool that implementing the Integrated Programme offered by the Social Welfare Department. Ten of the 11 children had a diagnosis of language impairment given on the basis of the clinical judgment of a speech therapist, and they were all currently receiving language therapy or had received language therapy in the last six months as reported in the case history form completed by the parents. One participant was reported to have a history of language impairment but his language ability was clinically assessed to be caught up with his peers after receiving language therapy. This participant was excluded from the study due to his resolved language ability. The rest of the ten children made up the language impaired (LI) group.

Ten children, aged between 5;00 and 5;11 (with a mean of 5;6), were recruited from two mainstream kindergartens and through colleagues of the author as the control group. No reports of speech and language problems or speech therapy services were indicated in the case history form completed by the parents. Given that this is the first study investigating the diagnostic accuracy of the DSI test, to maximize its specificity, we chose children from the age range which had the smallest standard deviation value for the group mean in the

normative data. Children who are bilingual were excluded since the tests used in this study were designed for monolingual Cantonese-speaking children. Bilingual children are those who communicate with people using Cantonese less than 70% of time as indicated in the case history from by their parents (Fong, 2006).

Measures

The DSI test used from Mok (1995) contains 67 items, covering 17 different types of Cantonese grammatical structures. There are two sentences in each item, a target sentence and a distractor. The target sentence and the distractor are the same in grammatical structure but different in the specific lexemes used. The two pictures describing the sentences were presented to the children with the sentences to reduce memory load. Each item was scored using the 0-3 scoring system (Mok, 1995). Three points were given if the child produced a sentence that revealed the same underlying grammatical structure and lexemes as the target modeled. Two points were given if the child produced a sentence that could be identified as the underlying structure with one lexeme missing or incorrect, whereas one point was given if the child used some lexemes from the distractor sentence or there were some errors in the target structure being examined. Zero point was given if the child imitated the distractor or the response was inadequate to show the use of the target structure or irrelevant to the picture.

The Cantonese adaptation of the Reynell Developmental Language Scales (RDLS; Hong Kong Society for Child Health and Development, 1987) is the only published norm-referenced test for assessing the general language ability of Cantonese-speaking preschoolers and is widely used by speech therapists in Hong Kong (Klee, Wong, Stokes,

Fletcher, & Leonard, 2009). The RDLS comprises of a receptive and an expressive scale. Both scales of the RDLS were administered to assess the children's language ability formally. Since the RDLS only meets five psychometric criteria in the psychometric review (Klee et al., 2009) and there is no report of its diagnostic accuracy, the results of the RDLS and the clinical judgment of speech therapist would together form the reference standard of the diagnostic status of the LI group in this study (Dollaghan, 2007). Since most of the participants in the control group had never received language assessment or therapy before, the clinical judgment of a speech therapist was not available for them. The reference standard of the control group would be formed on the basis of a lack of family concern on their language development and their performance on the RDLS.

There is not much guidance available for deciding the cut-off scores of diagnostic tests (Paul, 2007) and there is no cut-off score reported in both the RDLS and DSI test for language impairment. The diagnostic accuracy of the DSI test was therefore calculated using two commonly adopted cut-off scores, which were 1 S. D. below the mean (approximately 16^{th} percentile) (Records & Tomblin, 1994) and 1.25 S.D. below the mean (approximately 10^{th} percentile) (Fey, 1986). Therefore, the reference standard criteria for language impairment in the LI group in this study included 1) performance below the cut-off score in either the receptive or expressive or both scales of the RDLS and 2) clinical judgment of a speech therapist as language impaired. The reference standard criteria for normal language skills in the control group included 1) performance above the cut-off score in both scales of the RDLS and 2) absence of family concern of language impairment. Children who scored

below the cut-off score in the DSI test (the index measure) were classified as tested positive.

A forward digit span (F-DS) test was adopted from the Wechsler Intelligence Scale for Children-III (WISC-III; Wechsler, 1991) to measure the STM of the children. The test comprises of 16 randomly generated series of single-digit numbers. The length of the series ranges from two to nine digits. There are two series for each of the eight digit lengths. The F-DS test was administered according to the instructions in Wechsler (1991). One point would be given if the child repeated all the digits in a series correctly and in the same order as presented. For example, if the original order of a series was 6-1-5-8 and the child recalled as 6-1-8-5, the child did not get a point due to the incorrect order. The score of the test was the sum of points the child got.

Procedure

The children were seen individually twice for testing in a quiet room in the Internal Child Clinic of the Speech and Hearing Division or in the preschools. Testing included the RDLS, the DSI test, and the F-DS test. Since the DSI test was relatively long, it was split into two parts and carried out in two separate meetings. All three tests were randomized in the order of administration. All responses in the tests were audio-recorded for subsequent scoring and reliability checks. The three tests were administered independently by the author and another year four undergraduate student from Speech and Hearing Sciences at the University of Hong Kong. One test administrator carried out the RDLS and the author administered the DSI test and F-DS test. Both administrators were blind to the findings of the tests they did not administer in order to control for subjective bias and to ensure independent decisions.

Inter-rater reliability

Two year four undergraduate students from Speech and Hearing Sciences at the University of Hong Kong independently rated the DSI test from three randomly-selected participants (15 % of the total number of participants) using the 0-3 scoring systems. Pearson Product-Moment Correlation Coefficient *r* was calculated to examine the correlation of the scores among the two raters and the author and yielded *r* of .851 and .836 ($p < .005$). The strong and positive correlations revealed good inter-rater reliability on the scoring of the DSI test.

Results

Diagnostic accuracy of the DSI test

The diagnostic accuracy of the DSI test was examined using two cut-off scores: -1 S.D. (approximately 16th percentile) and –1.25 S.D. (approximately 10th percentile) below the mean for age. These cut-off scores were applied to both the RDLS and the new DSI test, resulting in four possible combinations ((1) -1 S.D. for both the RDLS and DSI test; (2) -1.25 S.D. for both the RDLS and DSI test; (3) -1 S.D. for the RDLS and -1.25 S.D. for DSI test; (4) -1.25 S.D. for the RDLS and -1 S.D. for DSI test). Table 1 illustrates the relationship between results of the DSI test and those of the reference standard. Based on the cut-off score of -1 S.D. for both the RDLS and the DSI test, eight of the ten participants (80%) in the control group were found to be true negative, while two participants (20%) were found to be false positive. The false positive cases indicated that the DSI test wrongly assessed the children as language impaired. For the LI group, eight of the ten participants (80%) were found to be true

positive, while two participants (20%) were found to be false negative. The false negative cases indicated that the DSI test wrongly reported the children who were truly language impaired as free of language impairment.

Table 1. The relationship between results of the DSI test and those of the reference standard.

	Reference standard	
DSI test results	Positive	Negative
Tested Positive	True Positive	False Positive
Tested Negative	False Negative	True Negative

The analysis of the diagnostic accuracy of the DSI test involved calculating its sensitivity (true positive/ [true positive + false negative]), specificity (true negative/ [false positive + true negative]), LR+ (sensitivity/ [1 – specificity]), and LR- ([1 – sensitivity]/ specificity) (Dollaghan, 2007) using a statistical calculator. Results showed a sensitivity of 80%, specificity of 80%, LR+ of 4.00, and LR- of 0.25 and indicated a fair diagnostic accuracy of the DSI test according to the criteria suggested by Plante and Vance (1994) and Dollaghan (2004). The diagnostic accuracy of the DSI test using the other three combinations of cut-off points of the tests was analyzed as well and the results were presented in Table 2. The diagnostic accuracy of the DSI test was the highest when the cut-off score of the DSI test was set at -1 S.D. and was the lowest when the cut-off score of the DSI test was set at -1.25 S.D. The diagnostic accuracy of the DSI test was not affected by the change of the cut-off scores of the RDLS in this study.

Table 2. Results of sensitivity, specificity, LR+, and LR- of the DSI test at four combinations of cut-off points.

RDLS cut-off	-1 S.D.	-1.25 S.D.	-1 S.D.	-1.25 S.D.
DSI test cut-off	-1 S.D.	-1.25 S.D.	-1.25 S.D.	-1 S.D.
Sensitivity	80%	60%	60%	80%
Specificity	80%	90%	90%	80%
LR+	4.00	6.00	6.00	4.00
LR-	0.25	0.44	0.44	0.25

Note. RDLS = the Cantonese adaptation of the Reynell Developmental Language Scales; DSI test = Delayed sentence imitation test (Mok, 1995); LR+ = positive likelihood ratio; LR- = negative likelihood ratio

Relation between performance in the RDLS and the DSI test

The diagnostic accuracy of the DSI test was revealed to be fair. An individual who scored below the cut-off in the DSI test was likely to be truly language impaired according to the reference standard, and an individual who scored above the cut-off in the DSI test was likely to be free from LI. Given this finding, it was predicted that there would be a high correlation between the children's performance in the two scales of the RDLS and the DSI test. Pearson Product-Moment Correlation coefficient *r* was run on the data to test this hypothesis. Results revealed a strong, positive and significant correlation ($r = .890$, $p < .005$) between the children's standard scores in the DSI test and the receptive scale of the RDLS.

Participants who scored higher in the DSI test also tended to scored higher in the receptive scale of the RDLS. A strongly positive and significant correlation was also found between the standard scores in the DSI test and in the expressive scale of the RDLS ($r = .742$, $p < .005$). The scatter plots of the standard scores of both scales of the RDLS and the DSI test (Figure 1 and Figure 2) also showed the spread of scores of the LI and control groups, with the scores of LI group mainly clustered in the bottom part and the scores of control group clustered in the upper end. Such findings confirmed what was expected based on the fair diagnostic accuracy of the DSI test.

There were two interesting observations in Figure 1 and 2. One observation was that the data point of one participant in the control group (marked by a darken triangle ▲) lied in the cluster of data points of the LI group. This participant scored below -1.5 S.D. in both scales of the RDLS but was not reported to be language impaired and her parents did not show concern on her language ability. Figure 2 illustrates another interesting observation. Three data points were clearly a distance away from the regression line (marked by a square □), and all three belonged to the LI group. These LI participants scored higher than -1 S.D. in the expressive scale of the RDLS but lower than -2 S.D. in the DSI test, suggesting that the expressive scale of the RDLS might not be sensitive to discriminate children with language impairment especially if it were the only test used for the diagnosis of LI.

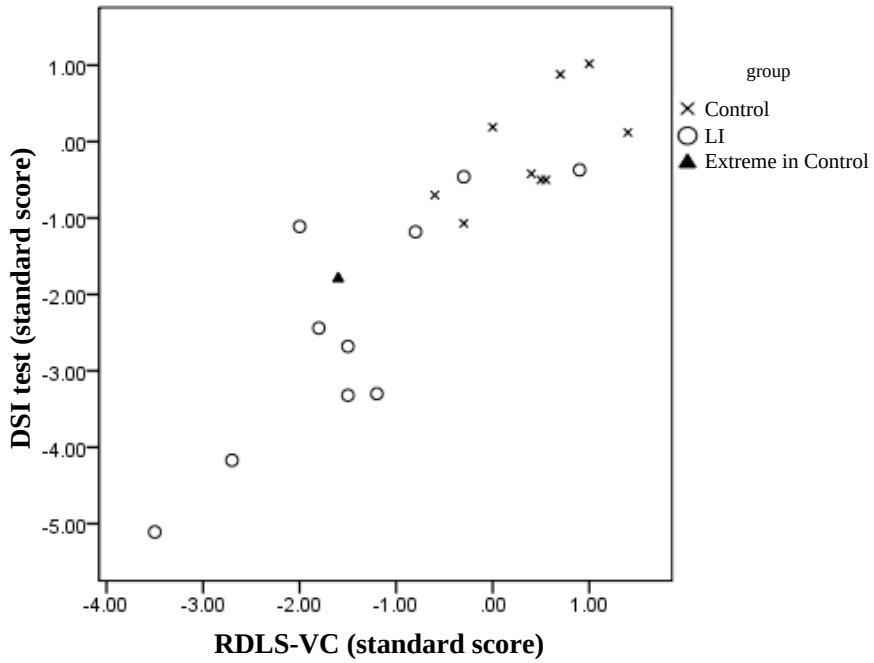

Figure 1. Scatter plot of standard scores of the delayed sentence imitation test (DSI test) and the receptive scale of Cantonese adaptation of the Reynell Developmental Language Scales (RDLS-VC).

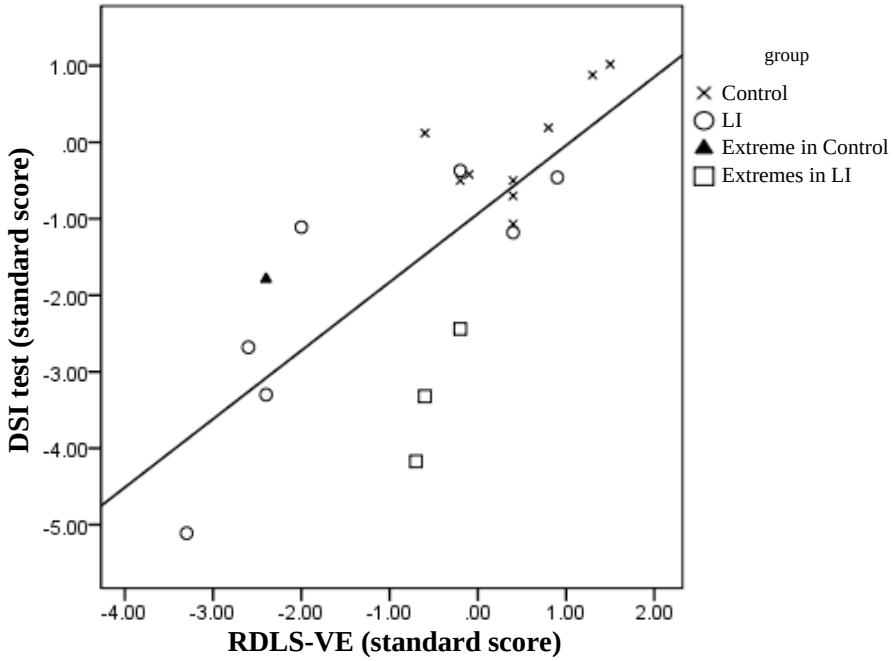

Figure 2. Scatter plot of standard scores of the delayed sentence imitation test (DSI test) and the expressive scale of Cantonese adaptation of the Reynell Developmental Language Scales (RDLS-VE), with a regression line.

Relation between short-term memory and performance in the DSI test

The correlation between performance in the DSI test and short-term memory (STM) was analyzed to examine the nature of the relation between the two variables. The participants' STM was measured by the forward digit span (F-DS) test in terms of the number of digit series that the child repeated correctly. The Pearson Product-Moment Correlation coefficient *r* was calculated and revealed a moderate and positive correlation between the DSI and the F-DS test score ($r = .694$, $p = .001$). Figure 3 shows the scatter plot of the two variables. It was observed from Figure 3 that most of the data points of the control group were clustered at the upper part and these data points did not reveal a linear relationship. The relation between STM and DSI test performance was therefore examined for two groups separately. The nonparametric test of Spearman's Rank Order Correlation *rho* was used due to the small sample size in each group (n = 10). Results showed that the correlation between STM and the DSI test performance was significant and moderately positive in the LI group ($r_s = .679$, $p = .031$) but insignificant in control group ($r_s = .524$, $p = .120$). The results suggested that STM did not relate to every child's performance in the DSI test equally. Children with a lower language ability seemed to depend on STM when performing the DSI test while those with a higher language ability might not depend on STM for the same test.

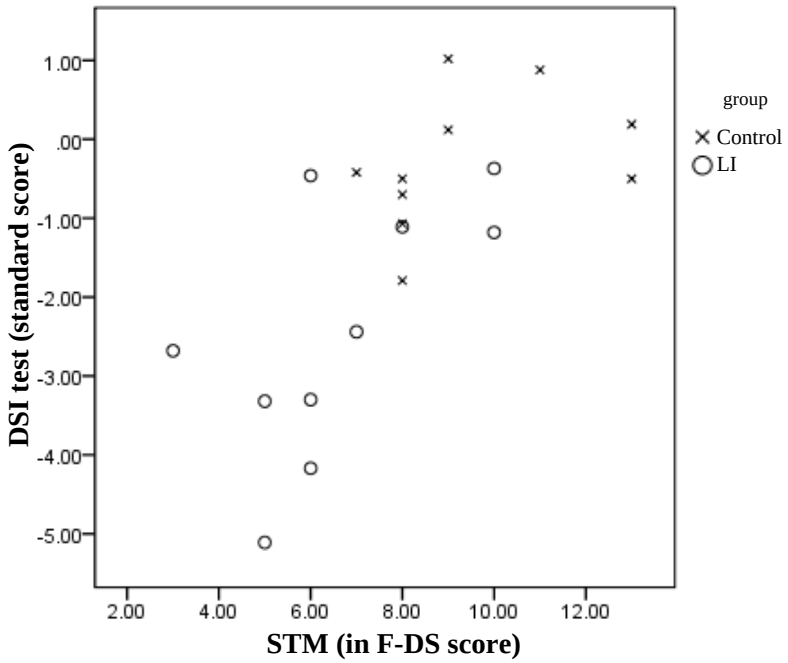

Figure 3. Scatter plot of the delayed sentence imitation test (DSI test) standard scores and forward digit span (F-DS) test score, for children in language impaired group and control group.

Discussion

Diagnostic accuracy of the DSI test

The optimal diagnostic accuracy of the DSI test was found to be 80% sensitivity, 80% specificity, 4.00 LR+, and 0.25 LR- when the cut-off scores were set at -1 S.D. for the DSI test and -1 or -1.25 S.D. for the RDLS. These results indicate a fair diagnostic accuracy of the DSI test according to the criteria suggested by Plante and Vance (1994). Based on Dollaghan's (2004) criteria, the intermediate value of LR+ and LR- indicates that the test result using the DSI test are suggestive but insufficient to diagnose or exclude the disorder (Dollaghan, 2007). The results are similar to the findings of previous studies that EI tasks are potentially useful in identifying children with language impairment (Botting &

Conti-Ramsden, 2003; Conti-Ramsden et al., 2001; Stokes et al., 2006). The fair diagnostic accuracy of the DSI test and the significant, strong and positive correction between the children's performance in the DSI test and both scales of the RDLS provide evidence of the concurrent validity of the DSI test, in which the DSI test is now found to meet one more psychometric criterion suggested by McCauley and Swisher (1984).

When the cut-off score of the DSI test was changed to -1.25 S.D., the diagnostic accuracy of the DSI test was lowered with unacceptable sensitivity (60%) and intermediate LR+ (6.00) and LR- (0. 44). Such an inconsistency in the diagnostic accuracy of the DSI test when two cut-off scores were used for both the RDLS and the DSI test may be due to the fact that the diagnostic accuracy of the RDLS itself has not been examined (Klee et al., 2009). There is no empirically-derived and precise cut-off point of the referent test, the RDLS, for defining the language status of the children (i.e. whether a child is language impaired or non-impaired) at different ages. Although we included clinical opinions as part of the reference standard, undoubtedly the RDLS took a reasonably heavier weight in the diagnostic decision as it is used by 88% of speech therapists in Hong Kong for diagnostic purpose (Klee et al., 2009). Plante and Vance (1994) have addressed the importance in deriving a score for each of the standardized tests empirically from statistically analysis instead of using arbitrary cut-off scores across tests. In their study, the empirically-derived cut-off score for each of the four standardized tests in assessing 4- to 5-year-old children ranged from -3.25 S.D. to 0.06 S.D. The rate of misidentifying children's language status was high if a single, arbitrary cut-off score such as -1 S.D. was used across tests. In order to ascertain that the RDLS is a

good reference test, an optimum cut-off score should be determined statistically. The receiver operating characteristic (ROC) curve is one of the statistical methods to evaluate the performance of a diagnostic test (Fawcett, 2006). The ROC curve is a plot of sensitivity against [1-specificifity] for different possible cut-off points of a diagnostic test and the accuracy of the test can be measured by the area under the ROC curve (Park, Goo, & Jo, 2004). Conti-Ramsden et al. (2001) demonstrated the use of the ROC curve to identify the optimum cut-off score for the marker tasks in their study.

Despite its lack of empirically-derived cut-off scores and diagnostic accuracy information, the RDLS has been employed for more than 20 years since its publication in 1987. Speech therapists in Hong Kong expressed in a survey that the norm of the RDLS seems to be outdated and the RDLS tends to over-estimate children's language ability compared to their clinical judgment (Klee et al., 2009). In this study, three participants in the LI group scored above the pre-determined cut-off on the expressive scale of the RDLS while scoring more than two standard deviations below the mean in the DSI test. This result confirms with the speech therapists' views in the survey that the expressive scale of the RDLS with outdated norm may not be sensitive for the identification of children with language impairment. Therefore, there is a need to determine the diagnostic accuracy and derive optimum cut-off scores for the RDLS. This information will help to make the RDLS become a more valid language assessment tool for clinical purposes and for evaluating the diagnostic accuracy of other language tests.

There is also another limitation of the reference standard used in this study. The

reference standard for the control group was scoring above the cut-off score in both scales of the RDLS and no family concern on language development. One participant in the control group, who did not have a clinical report or family concerns of speech and language problem, was unable to pass the reference test (i.e. scored below the cut-off point of both scales of the RDLS). The disagreement between the components of the reference standard for this child in the control group indicates that this reference standard is not ideal. Apart from uncertainties with the validity of the RDLS discussed above, the accuracy of parents' impressions about children's language ability is also questionable. Parents usually lack professional training in language development thus they may not be aware of subtle language problems in their children which could be spotted out by speech therapists (Dale, 1996; Fong, 2006). Parent may also lack the ability to critically assess impressions about their children's language ability due to pride in their children (Dale, 1996; Fong, 2006). It is recommended that the language status of children in the control group should be determined by professionals in the same way as LI group in future studies.

Limitations of the current DSI test

Although the DSI test was found to have a fair diagnostic accuracy, the author encountered some difficulties during data collection and analysis. Some refinements of the test may be required before it can be used for clinical practice. Firstly, the DSI test may be too long and too demanding for some young preschoolers and for some children with LI. In the current test, each grammatical structure is elicited three to four times to allow item analysis, which is the error analysis of individual grammatical structures, for intervention

planning. Children in the control group generally took 20 to 30 minutes to complete the DSI test while children in the LI group took about 40 to 50 minutes to finish the test. One extreme case was that one LI participant, aged 3;07, took 75 minutes to complete the test. Although the test could be administered in two separate sessions, it may still be too long for some children, especially those with a short attention span. A shorter test, such as eliciting each grammatical structure two times, is recommended. Although there will be a trade-off between the administration time and the depth of item analysis when a shortened test is adapted, it can be supplemented with follow-up probes on specific structures during therapy if a child is tested positive in the shortened test.

Secondly, although Mok (1995) provided a set of comprehensive scoring criteria so that examiners can refer to them for most of items during actual scoring, some variations in the children's responses still have not been accounted for. For example, there is no suggestion for scoring productions which are similar to the target sentences but not grammatically appropriate or acceptable, such as "個女人唔單止追住隻牛同埋狗" (The woman is not only chasing the cow and the dog) for the target sentence of item 53 "個女仔唔單止追緊隻牛仲追緊隻狗" (The girl is not only chasing the cow but also chasing the dog). The scoring scheme allowed the substitution of an additive coordinator with a similar meaning "同埋" (and) for the target "仲" (also) but did not mention the scoring criteria for the grammatically inappropriate productions after substitution. Another limitation is that no explicit scoring criteria are provided for word order problems, such as "個女仔追緊隻狗 唔單止追緊隻牛" (The girl is chasing the dog, not only chasing the cow) for the target sentence of the same

item “個女仔唔單止追緊隻牛仲追緊隻狗” (The girl is not only chasing the cow but also chasing the dog). In this case, the order of the phrases “唔單止追緊隻牛” (not only chasing the cow) and “追緊隻狗” (chasing the dog) was reversed. In summary, if some modifications can be made on the DSI test, the clinical application of the DSI test for assessing the language ability of preschool-aged children will be more appropriate and user-friendly.

Short- term memory and the DSI test

A significant and moderately positive correlation between STM and the children’s performance in the DSI test was revealed in the LI group but not in the control group. The results indicate that STM does not relate to the language ability of every child in the same way. Children with LI tend to score higher in the DSI test if they have better STM. The STM of typically developing children, on the other hand, does not relate to their language ability. Children with higher language ability do not rely on STM as much as children with LI when performing the DSI test. The moderate strength of the correlation between STM and performance of children with LI on the EI task is in line with previous findings that EI involves more than purely temporal storage of the model sentences (Vance, 2008). EI requires the dual functions of storage and processing of the target sentences heard at the same time, and functional working memory (WM) described by Montgomery (2002) is likely to be involved.

In the functional WM, the limited amount of resources is shared by both storage and processing functions (Montgomery, 2002). When the demand of information storage and processing in a task exceeds the limited amount of resources, a trade-off between storage and

processing of information is resulted (Montgomery, 2002). For instance, the resource for storage will be reduced when the demand for processing during comprehension of the model sentence increases. Problems in storage will then lead to loss of some semantic or syntactic information of the model sentences in recall. Therefore, it is important to be aware of the memory load that the DSI test may impose on some children, especially those with language impairment. The DSI test has intended for addressing this memory demand issue by providing picture support. With the help of the pictures, the demand on the temporarily storage of the semantic information of the target sentences reduces. A greater proportion of the resources will be available for comprehending the target sentences and for processing the syntactic structures of the target sentences.

Clinical implications and further investigations

This study demonstrates the use of the EBP framework for evaluating the diagnostic accuracy of the DSI test. Given the fair diagnostic accuracy of the DSI test for identifying language impairment in this small sample of preschool age children, the clinical use of the DSI test for assessing children with language impairment is rather promising. The findings from this study encourage future studies to critically appraise norm-referenced tests or other new measures for diagnostic accuracy and to promote the application of EBP in our clinical work. This study also draws attention to the needs for new language assessment tools for clinical use in Hong Kong (Klee et al., 2009). Further investigations are recommended to ascertain the usefulness of the DSI test with a larger sample size and a more diverse group of children with language impairment in terms of different severity and medical diagnoses. In

addition, further studies can investigate the effect of working memory such as functional WM, phonological WM and processing speed on the performance of EI tasks in Cantonese-speaking children, which will in turn contribute to the understanding of the relationship between language processing and working memory in Cantonese case.

Acknowledgements

The author would like to express my deep gratitude to Dr. Anita Wong for her kind supervision, and valuable advice throughout the study. The author would like to gratefully acknowledge the SAHK for the support in participant recruitment. The author is grateful to all the participants and their parents for their participation in this study. Thanks are given to the principals, teacher and staff of the following preschools and clinic for their support in data collection: Apleichau Pre-school Centre, Truth Baptist Church Kindergarten, Salvation Army Tin Ka Ping Kindergarten, and Internal Child Clinic of the Speech and Hearing Division, University of Hong Kong. Thanks are also extended to Miss Rita Chan, Miss Sally Wong, Miss Julia Yuen, Miss Vivian Wong, and Miss Flora Lee for their help during data collection and analysis.

References

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

Baddeley, A. (1986). Working memory and comprehension. In D. Broadbent, J. McGaugh, M. Kosslyn, N. Mackintosh, E. Tulving, & L. Weiskrantz (Eds.), *Working memory* (pp. 75–107). New York: Oxford University Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417 – 423.

Botting, N. & Conti-Ramsden, G. (2003). Autism, primary pragmatic difficulties, and specific language impairment: Can we distinguish them using psycholinguistic markers? *Developmental Medicine & Child Neurology, 45*, 515-524.

Cheung, P. M.-P. (1993). *Comprehension of grammatical structures in Cantonese-speaking preschoolers: A Cantonese adaptation of the Test for Reception of Grammar.* Unpublished Honors B. Sc. Thesis, Department of Speech and Hearing Sciences, University of Hong Kong.

Conti-Ramsden, G., Botting, N., & Faragher, B. (2001). Psycholinguistic markers for specific language impairment (SLI). *Journal of Child Psychology and Psychiatry, 42*(6), 741-748.

Dale, P. S. (1996). Parent report assessment of language and communication. In K. N. Cole, P. S. Dale, & D. J. Thal (Eds.), *Assessment of communication and language* (pp. 161-182).

Baltimore: Paul H Brookes.

Dollaghan, C. A. (2004). Evidence-based practice in communication disorders: what do we know, and when do we know it? *Journal of Communication Disorders, 37,* 391-400.

Dollaghan, C. A. (2007). *The handbook for evidence-based practice in communication disorders*. Baltimore: Paul H. Brookes Publishing Co.

Fawcett, T. (2006). An introduction of ROC analysis. *Pattern Recognition Letters, 27,* 861-874.

Fey, M. (1986). *Language intervention with young children.* San Diego: College-Hill Press.

Fong, K. Y. -Y. (2006). *Validation of a screening tool for identifying three-year-old Cantonese-speaking children with language impairment.* Unpublished Honors B. Sc. Dissertation, Division of Speech and Hearing Sciences, University of Hong Kong.

Hammil, D. D., & Newcorner, P. L. (2008). *Test of Language Development: Primary* (4th ed.). Austin: PRO-ED.

Hong Kong Society for Child Health and Development. (1987). *Manual of the Reynell Developmental Language Scales, Cantonese (Hong Kong) version*. Hong Kong: Author.

Just, M., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.

Klee, T., Wong, A. M.-Y., Stokes, S. F., Fletcher, P., & Leonard, L. B. (2009). Assessing Cantonese-speaking children with language difficulties from the perspective of evidence-based practice: Current practice and future directions. In S.-P. Law, B. S.

Weekes, & A. M.-Y. Wong. (Eds.), *Language disorders in speakers of Chinese* (pp. 89-111). Bristol: Multilingual matters.

Leonard, L. B., Weismer, S. E., Miller, C. A., Francis, D. J., Tomblin, J. B., & Kail. R. V. (2007). Speed of processing, working memory, and language impairment in children. *Journal of Speech, Language, and Hearing Research, 50*, 408-428.

McCauley, R. J., & Swisher, L. (1984). Psychometric review of language and articulation tests for preschool children. *Journal of Speech and Hearing Disorders*, *49*, 34-42.

Mok, C. K.-F. (1995). *Comprehension and expression of grammatical structures in Cantonese-speaking preschoolers.* Unpublished Med. Thesis, Faculty of Education, University of Hong Kong.

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.

Park, S. H., Goo, J. M., & Jo, C.-H. (2004). Receiver operating characteristic (ROC) curve: Practical review for radiology. *Korean Journal of Radiology, 5*(1), 11-18.

Paul, R. (2007). *Language disorders from infancy through adolescence: Assessment and intervention* (3rd ed.). St. Louis: Mosby Elsevier.

Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25,* 15-24.

Records, N., & Tomblin, J. B. (1994). Clinical decision making: Describing the decision rules of practicing speech-language pathologists. *Journal of Speech and Hearing Research,*

37, 144-156.

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal, 312*(7023), 71-72.

Semel, E., Wiig, E. H., & Secord, W. (1994). *Clinical Evaluation of Language Fundamentals-Revised*. San Antonio: The Psychological Corporation.

Semei, E., Wiig, E. H., & Secord, W. (2006). *Clinical Evaluation of Language Fundamentals* (4th ed.). London: The Psychological Corporation.

Slobin, D. I., & Welsh, C. A. (1973). Elicited imitation as a research tool in developmental psycholinguistics. In C. A. Ferguson & D. I. Slobin (Eds.), *Studies of child language development* (pp. 485 – 497). New York: Holt, Rhinehart & Winston.

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.

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: From theory to practice* (pp. 23-38). Hove: Psychology Press.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-III*. San Antonio: Psychological Corporation.