Running Head: Differentiating Cantonese preschoolers with SLI

Research note

Differentiating Cantonese-speaking pre-school children with and without SLI using MLU and lexical diversity (D)

Anita M.-Y. Wong, University of Hong Kong

Thomas Klee, University of Canterbury

Stephanie F. Stokes, Curtin University of Technology

Paul Fletcher, University College Cork

Laurence B. Leonard, Purdue University

Contact author:

Anita M.-Y. Wong, Ph.D.

Assistant Professor

Division of Speech and Hearing Sciences

Faculty of Education

University of Hong Kong

Pokfulam

Hong Kong-SAR-China

Email: amywong@hkusua.hku.hk

Abstract

Purpose: This study examined the diagnostic accuracy of a composite clinical assessment measure based on mean length of utterance (MLU), lexical diversity (D) and age (Klee, Stokes, Wong, Fletcher, & Gavin, 2004) in a second, independent sample of 4-year-old Cantonese-speaking children with and without Specific Language Impairment (SLI).

Method: The composite measure was calculated from play-based, conversational language samples of 15 children with and 14 children without SLI. Scores were dichotomized and compared to diagnostic outcomes using a reference standard based on clinical judgment supported by test scores.

Results: Eleven of 15 children with SLI and 8 of 14 children with typical language skills were correctly classified by the dichotomized composite measure. The measure's sensitivity in this second sample was 73.3% (95% CI 48% to 89%); specificity was 57.1% (95% CI 33% to 79%); positive likelihood ratio was 1.71 (95% CI 0.87 to 3.37); and negative likelihood ratio was 0.47 (95% CI 0.18 to 1.21).

Conclusions: The diagnostic accuracy of the composite measure was substantially lower than in the original study, suggesting that it is unlikely to be informative for clinical use in its present form. The value of replication studies is discussed.

Key words: Cantonese Chinese, specific language impairment, language sampling, assessment, diagnostic accuracy

One aspect of clinical assessment involves accurately differentiating individuals with and without disorders. This is an important first step in intervention planning as well as in describing individuals who participate in research involving clinical populations. Clinical assessment of children suspected of having speech or language disorders relies in part on tests and measures that accurately inform clinical judgment (i.e., demonstrate high diagnostic accuracy). Evidence suggests that some language sample measures when used in isolation (e. g., percentage use of finite verb morphemes), or in combination with others (e.g., mean length of utterance (MLU)), can be used to accurately identify English-speaking children with language impairment (see Klee, Gavin, & Stokes, 2007 for a review). The diagnostic potential of language sample measures has also been examined in children learning languages other than English, including Spanish (Simon-Cereijido & Gutierrez-Clellen, 2007) and Cantonese Chinese (Klee, Stokes, Wong, Fletcher, & Gavin, 2004).

Klee, Stokes, Wong, Fletcher, & Gavin (2004) reported that a composite measure based on age, MLU and D, a measure of lexical diversity (Malvern & Richards, 2002), yielded high sensitivity and specificity estimates (> 90%) in their sample. All 15 4-year-old children in the SLI group, all 15 children in a younger language-matched group and all but one of 15 children in an age-matched group were correctly classified by the composite measure based on a discriminant analysis. However, the 95% confidence intervals were wide, due in part to the sample size (Klee et al., 2007), leading the authors to caution that before the diagnostic measure could be recommended for clinical use, its accuracy in another independent sample of Cantonese-speaking children needed to be examined. The purpose of the study reported here is to examine the diagnostic measure in a second, independent sample of children.

Method

Participants. A total of 29 children between 49 and 60 months of age participated in the study, with data coming from two sources. Data were collected from 17 children recruited specifically for this study (8 in the SLI group, 9 in the TD group) and 12 children recruited for previous studies (Fletcher, Leonard, Stokes, & Wong, 2005; Leonard, Deevy, Wong, Stokes, & Fletcher, 2007; Leonard, Wong, Deevy, Stokes, & Fletcher, 2006; Stokes, Wong, Fletcher, & Leonard, 2006; Wong, Leonard, Fletcher, & Stokes, 2004). Fifteen children (13 boys) previously diagnosed with language impairment were referred to the study by speech-language therapists and 14 typically-developing (TD) children (10 boys) were recruited from neighborhood preschools. To ensure that the children in the study sample were similar in age to those of the original study (Klee et al., 2004), the children in this study were selected so that the range and mean age for the SLI and TD groups in both studies were within two months of each other. Children were administered the receptive and expressive subtests of the Cantonese version of the Reynell Developmental Language Scales (RDLS-R and RDLS-E; Hong Kong Society for Child Health and Development, 1987). All children in the SLI group scored below -1 standard deviation (SD) of the mean on the RDLS-R, with seven children scoring below -1.25 SDs. All children in the TD group scored above -0.67 SD on both subtests of the RDLS. Receptive test scores of children in the TD group were significantly higher than those of children in the SLI group ($F(1,27) = 69.24$, $p < .0001$, $d = 3.23$). Similarly, expressive test scores of children in the SLI group were significantly higher than those of the TD group ($F(1, 27) = 12.91$, $p = .001$, $d = 1.36$).

All children in the study scored above -1 SD on the Columbia Mental Maturity Scale (CMMS; Burgemeister, Blum, & Lorge, 1972), a test of nonverbal cognitive ability. The TD group received a slightly higher CMMS score than the SLI group and this difference was approaching significance, $F(1, 27) = 3.75$, $p = .063$. All children also passed a pure-tone audiological screening (.5, 1, 2 and 4kHz presented at 25-30dB HL) and an oral motor screening that was adapted from Robbins and Klee (1987). None of the children had a history of seizure disorder, neurological or psychosocial problems. None of the children in the TD group had a history of speech and language difficulties nor had parental concerns been expressed. Table 1 displays descriptive statistics for the study variables from the original sample (Klee et al., 2004) and the follow-up sample.

Insert Table 1 about here

Language samples. Each child engaged in a 15-20 minute conversation with one of two speech-language pathology research assistants trained in language sampling. These conversations often revolved around, although were not restricted to, theme-based toys the children had chosen to play with. A team of 8 students in speech-language pathology, psychology and Chinese linguistics transcribed the samples after training on the word and utterance segmentation guidelines outlined in Klee et al. (2004). Each transcript was checked against the audio-recording for transcription accuracy and for consistency in word and utterance segmentation by a second experienced research assistant. Orthographic transcripts were then converted to Romanized form (Hong Kong Linguistic Society, 1997) in CHAT format (MacWhinney, 2006a) and checked for accuracy of marking lexical tones for each syllable, and for consistency in the Romanization of variant productions of the same lexeme (e.g., *nei5* and *lei5* with the same meaning: *you*). Transcribers were blind to the language status of the 17 children recruited specifically for this study, but not for the 12 children recruited for previous studies. MLU and D were calculated using the Child Language Analysis X computer program (CLAN-X; MacWhinney, 2006b) following the protocol outlined in Klee et al. (2004).

Index measure and reference standard. The index measure was a composite variable made up of MLU, D and age. Scores were calculated and dichotomously classified (SLI, TD) on the basis of a discriminant function equation derived from the original study data (Klee et al., 2004). Because the discriminant function analysis in the original study was based on three participant groups (SLI, age-matched and language-matched), a new discriminant analysis was run using data from the original SLI and age-matched groups only, consistent with the current study. The resulting discriminant function equation was (-0.037 x Age) + (0.931 x MLU) + (0.099 x D) - 7.269. The centroid for the SLI group was -2.123 and +2.123 for the TD group. The midpoint between the two centroids, 0, served as the threshold for predicting each child's group membership.

The reference standard was defined as the clinical judgment of an experienced speech-language pathologist, whose diagnosis of SLI or TD was based in part on RDLS test scores. However, the individual making the diagnosis was not aware of the child's MLU or D scores at the point at which the diagnosis was made.

Statistical analysis. A child was correctly classified if his/her discriminant score accurately predicted the diagnostic group to which s/he belonged. Diagnostic accuracy measures including sensitivity, specificity, positive and negative likelihood ratios were calculated in order compare the outcomes of the follow-up study to those of the original study. These were calculated using the *Stats Calculator* on the website of the University of Toronto's Center for Evidence-Based Medicine (http://www.cebm.utoronto.ca/).

Results

Descriptive statistics for the language sample measures are presented in Table 1. The TD group produced more complete and intelligible utterances (CIUTT) than the SLI group, and this difference was approaching significance, ($F(1, 27) = 3.55$, $p = .070$). However, the TD group produced significantly more words (TNW), ($F(1, 27) = 16.51$, $p < .001$, $d = 1.52$) than the SLI group, and they demonstrated more vocabulary diversity as measured by number of different words (NDW) ($F(1, 27) = 18.47$, $p < .0001$, $d = 1.60$). Regarding the main language sample variables of interest, the MLU of the TD group was significantly higher that of the SLI group ($F(1, 27) = 12.43$, $p = .002$, $d = 1.30$). Likewise, lexical diversity, as measured by D, was significantly higher in the TD group ($F(1, 27) = 9.24$, $p = .005$, $d = 1.13$).

Using the two-group discriminant function equation derived from the data in the original study (Klee, et al., 2004), 11 of the 15 children in the SLI group were correctly classified, as were 8 of the14 children in the TD group. The composite measure's sensitivity in the follow-up sample was 73.3% (95% CI 48% to 89%); specificity was 57.1% (95% CI 33% to 79%); positive likelihood ratio was 1.71 (95% CI 0.87 to 3.37); and negative likelihood ratio was 0.47 (95% CI 0.18 to 1.21).

Discussion

Results from this study did not replicate the high sensitivity, high specificity, high LR+ and low LR- reported in the original Klee et al. (2004) study. In fact, except for LR-, these diagnostic accuracy indicators fell outside the 95% CI of those reported in Klee et al. (2007). According to Plante and Vance (1994), sensitivity and specificity levels of 90% and above are considered to be good, 80% is considered to be fair and less than 80% is considered unacceptable. Using these criteria, neither the sensitivity nor specificity figures obtained in this study were acceptable. Similarly, neither the positive likelihood ratio (LR+) nor the negative likelihood ratio (LR-) was judged to be clinically useful, as a screening or a diagnostic instrument should have a LR+ greater than 10 and a LR- lower than 0.1 (Dollaghan, 2007).

There are several possible reasons for why the outcome of this study was not as favorable as that of the original study. The first may be related to characteristics of the language samples themselves. As Table 1 shows, the mean difference in average utterance length (MLU) between the SLI and TD groups in the original study was more than twice that of the present study (2.01 and 0.95 respectively). Similarly, the mean difference in lexical diversity (D) between these groups in the original study was 1.6 times of the present study. Therefore, the groups in the original study appeared to differ more on both variables than the groups in the present study. Moreover, the mean MLU of the SLI group in the present study was higher than that of the original study, while the mean D of the TD group in the present study was lower than that of the original study. Our hypothesis is that the diagnostic accuracy of the composite measure appears to change with the distribution of the underlying language production characteristics (MLU and D) of the groups.

A second possible explanation may relate to differences in how the TD and SLI groups were sampled between the original and follow-up studies. In the follow-up study, some of the children with SLI were included on the basis of a slightly lower language criterion. This did not result in major differences in the number of children with SLI who performed more than -1.50 SDs below the mean on RDLS-R (n = 11) when compared to the original sample (n = 10). It is plausible, however, that the two cohorts of children with SLI differed on aspects of language that could not be compared (RDLS-E) or that were not measured by formal tests (e.g., receptive and expressive vocabulary). In this study, all TD children received the entire language and nonverbal assessment battery. In the original study, children in the TD group were only given the RDLS-R but not the CMMS and the RDLS-E, and therefore, it may be that children in this TD group were more heterogeneous with respect to nonverbal cognition and language skills. In fact, there was greater variability in the MLU of Klee et al.'s (2004) TD group (SD = 1.33) as compared to the TD group in this study (SD = 0.71).

The findings of the present study reinforce the notion that just because groups of children with and without a clinical condition, such as SLI, are significantly different on a test or measure does not guarantee that the test or measure will be useful clinically. Earlier works suggest that within-group variability (Goffman & Leonard, 2000) and the overlap of score ranges of the two groups (Hewitt, Hammer, Yont, & Tomblin, 2005) might be the reasons why some of the language sample measures do not appear to be diagnostically useful. In the clinic, the important question is not whether groups differ on an assessment measure but whether an individual child's test (or language sample) results allow an accurate diagnosis to be made. In the case of the composite measure examined here, the outcome of the present study suggests that it may not, despite the positive findings of our original study. Future research into the diagnostic accuracy of clinical assessments might consider whether language sample features such as utterance formulation errors (e.g., Miller, 1991) or turn-taking and other discourse features (e.g., Evans, 1996) reported in English-speaking children with SLI also characterize Cantonese-speaking children with SLI. Research also suggests that measures such as sentence imitation may be useful (Conti-Ramsden, Botting, & Faragher, 2001). Stokes, Wong, Fletcher, & Leonard (2006) reported that their group of Cantonese-speaking children with SLI did significantly poorer than TD age peers on a task of sentence imitation. The sensitivity was found to be 77% and the specificity was 97%. Other promising diagnostic measures include measures of processing speed and working memory. Despite robust findings on English-speaking children (Leonard et al., 2007 for review), future work with Cantonese-speaking children with SLI should first confirm their deficits in these processing domains, since previous work on phonological working memory did not support the application of findings from English-speaking children cross-linguistically (Stokes et al., 2006). And, as the present investigation has demonstrated, it is of paramount importance that measures that look to be promising initially should be put to the test of replication subsequently.

Acknowledgements

We would like to thank Ms. Yvonne Lai at the Heep Hong Society and Ms. Gladys Yan at the SAHK, for their advice in participant recruitment. Thanks also go to the Heep Hong Society Shun Lee and Leung King Centre, the SAHK Chan Tseng Hsi Early Education and Training Centre, the University of Hong Kong Speech and Hearing Clinic, Pamela Youde Nethersole Eastern Hospital, the Hong Kong Christian Services Kwai Hing Centre, Joyful Mill, Kau Yan School Kindergarten Section, Yan Chai Hospital Fong Kong Fai Kindergarten, Rhenish Mission School and Thomas Tam Nursery School for their generous support in data collection. We thank our research coordinator Elaine Yung and research assistants Dorcas Chow, Ginny Lai, Deborah Pun, Penny Lee for their commitment to this project. We acknowledge the Hong Kong Research Grants Council (No. 7264/04H) and the National Institute on Deafness and Other Communication Disorders Research Grant (R01 00-458) for funding support. We could not have completed this project without the children who did their very best and made it fun for us. We thank them all. Portions of this study were presented at the Symposium on Research in Child Language Disorders, University of Wisconsin-Madison, USA, in June 2006 and the Child Language Seminar, Newcastle University, UK, in July 2006.

References

Burgemeister, B., Blum, L., & Lorge, I. (1972). *The Columbia Mental Maturity Scale*. New York: Harcourt Brace Jovanovich.

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.

Evans, J. L. (1996). SLI subgroups: Interaction between discourse constraints and morphosyntactic deficits. *Journal of Speech and Hearing Research, 39*, 655-660.

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.

Goffman, L., & Leonard, J. (2000). Growth of language skills in preschool children with specific language impairment: Implications for assessment and intervention. *American Journal of Speech-Language Pathology, 9*, 151-161.

Hewitt, L. E., Hammer, C. S., Yont, K. M., & Tomblin, J. B. (2005). Language sampling for kindergarten children with and without SLI: Mean length of utterance, IPSYN, and NDW. *Journal of Communication Disorders, 38*, 197-213.

Klee, T., Gavin, W. J., & Stokes, S. F. (2007). Utterance length and lexical diversity in American- and British-English speaking children: What is the evidence for a clinical marker of SLI? In R. Paul (Ed.), *Language disorders from a* *developmental perspective: Essays in honor of Robin S. Chapman* (pp. 103-140). Mahwah, NJ: Lawrence Erlbaum.

Klee, T., Stokes, S. F., Wong, A. M.-Y., Fletcher, P., & Gavin, W. (2004). Utterance length and lexical diversity in Cantonese-speaking children with and without specific language impairment. *Journal of Speech, Language, and Hearing Research, 47*, 1396-1410.

Leonard, L. B., Deevy, P., Wong, A. M.-Y., Stokes, S. F., & Fletcher, P. (2007) Modal Verbs With and Without Tense: A Study of English- and Cantonese-Speaking Children with Specific Language Impairment. *International Journal of Language and Communication Disorders, 42(2),* 209-228.

Leonard, L. B., Ellis Weismer, S., 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.

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.

Linguistic Society of Hong Kong (1994). *The LSHK Cantonese Romanization Scheme*. Hong Kong: Author.

MacWhinney, B. (2006a). CHAT [software manual]. Retrieved May 11, 2006 from http://childes.psy.cmu.edu/manuals/CHAT.pdf

MacWhinney, B. (2006b). Child Language Analysis (CLAN) (version May 11,2006) [Computer software]. Pittsburg, PA: Author.

Malvern, D., & Richards, B. (2002). Investigating accommodation in language proficiency interviews using a new measure of lexical diversity. *Language Testing*, *19*, 85-104.

Miller, J. (1991). Research on language disorders in children: a progress report. In J. Miller (Ed.), *Research on child language disorders: a decade of progress* (pp. 3-22). Austin, TX: Pro-ed.

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

Reynell, J., & Huntley, M. (1987). *Reynell Developmental Language Scales: Cantonese version*. Windsor: NFER-Nelson.

Robbins, J., & Klee, T. (1987). Clinical assessment of oropharyngeal motor development in young children. *Journal of Speech and Hearing Disorders*, *52*, 271-277.

Simon-Cereijido G., & Gutierrez-Clellen V. F. (2007). Spontaneous language markers of Spanish language impairment. *Applied Psycholinguistics*, *28*, 317-339.

Stokes, S. F., Wong, A. M.-Y., Fletcher, P., & Leonard, L. B. (2006). Nonword repetition and sentence repetition as clinical markers of SLI: The case of Cantonese. *Journal of Speech, Language, and Hearing Research*, *49*, 219-236.

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.

Table 1. Mean, standard deviation and range of study variables in Klee et al. (2004) study and in current sample.

	Klee et al. (2004)		Current sample	
	TD (n = 15)	SLI (n = 15)	TD (n = 14)	SLI (n = 15)
Age [a]				
Mean	56.87	56.40	55.71	55.27
SD	3.44	2.59	3.36	2.89
Range	52-61	52-59	49-60	50-60
RDLS-R[b]				
Mean	55.93	42.46	55.64	41.40
SD	3.83	9.98	3.23	5.59
Range	48-61	28-58	50-62	30-50
RDLS-E [c]				
Mean	NA	NA	57.57	49.40
SD			5.02	6.99
Range			59-66	37-62
CMMS [d]				
Mean	NA	NA	108.93	102.80
SD			5.99	10.32
Range			98-120	86-117
CIUTT [e]				
Mean	184.13	133.20	176.79	158.23
SD	51.87	20.99	23.32	28.39
Range	78-267	106-177	154-240	119-219
TNW [f]				
Mean	883.27	378.67	796.29	576.80
SD	333.39	102.63	134.42	154.80
Range	325-1311	251-540	578-1039	300-839
NDW [g]				
Mean	217.73	126.47	193.93	142.40
SD	42.06	21.65	31.98	32.53
Range	136-267	15-90	149-259	98-193
MLU [h]				
Mean	4.65	2.64	4.33	3.38
SD	1.33	0.85	0.71	0.75
Range	3.01-8.20	1.35-3.92	3.39-5.48	2.33-4.53

D [i]					
	Mean	72.26	48.20	57.69	42.92
	SD	12.53	8.69	12.49	13.59
	Range	54.07-97.14	30.96-59.34	40.95-82.38	23.48-68.48

Note. [a] Age in months; [b] RDLS-R: Reynell Developmental Language Scales-Receptive raw score; [c] RDLS-E: Reynell Developmental Language Scales-Expressive raw score; [d] CMMS: Columbia Mental Maturity Scale; [e] CIUTT: number of complete and intelligible utterances; [f] TNW: total number of words; [g] NDW: number of different words; [h] MLU: mean length of utterances in morphemes; [i] D: lexical diversity.