

ZHAO SHOUHUI, LIU YONGBING and HONG HUAQING

SINGAPOREAN PRESCHOOLERS' ORAL COMPETENCE IN MANDARIN

(Received 28 July 2006; accepted in revised form 20 November 2006)

ABSTRACT. Singapore's bilingual policy is widely regarded as having been successful. Mostly realized in the educational realm as "language-in-education planning" (Cooper, 1989), where English is legitimized as the medium of instruction and the mother tongues as subjects in the school system (Silver, 2005) the bilingual policy is connected explicitly to overall national goals of economic development and social harmony. In reporting detailed research data collected within the framework of a language in education planning program, this article addresses how language planning can respond to the Singapore context where the dominant spoken language, Mandarin, has been continually constituted as a "problem". Singapore has undertaken frequent reforms of its Chinese language (CL) teaching system, seeking ever greater levels of achievement and more cohesion in the role of Mandarin as the "culturally" unifying form for the majority of its people. Official policy announcing the two most recent reforms in 2000 and 2004 (Silver, 2005; Tan, 2006) has shifted emphasis towards a newly emerging binary that seems to assume increasing importance in Singapore's language planning paradigm: the English-speaking family (ESF) versus the Chinese-speaking family (CSF). The present article reports on the language proficiency of preschoolers, particularly their oral competence in Mandarin, with the objective of both reflecting on the success, assumptions and strategies of the new policy moves as well as proposing new ways to look at language-in-education planning for Chinese acquisition in a society that is, ethnically-speaking, dominantly Chinese.

KEY WORDS: children, Chinese language, language background, Mandarin competence, Singapore, vocabulary

ABBREVIATIONS: CL – Chinese Language; CLCPRC – Chinese Language Curriculum and Pedagogy Review Committee; CSF – Chinese speaking family or families; ESF – English speaking family or families; HL – Home language; NTUC – National Trade Union Congress; PAP – People's Action Party

INTRODUCTION

Singapore's bilingual policy, initially formulated for the purpose of promoting equal treatment of the constituent mother tongues of

Singaporean society and of encouraging what is understood in Singapore as ‘racial harmony’, was also from quite early on in its implementation connected explicitly to economic development (Lo Bianco, 2007). In light of subsequent economic globalization, the policy is now widely perceived within Singapore, but also more broadly, as having been remarkably successful in serving both political and economic purposes (e.g. Goh, 2004; Gopinathan et al., 2003; Shepherd, 2005). Characterized as top-down interventionism, the bilingual policy has been played out most intensively in the education arena (Silver, 2005). In other words, the bilingual policy has been promoted as “language-in-education planning” (Cooper, 1989: 33), where English is legitimized as the medium of instruction and the mother tongues as subjects in the school system (Silver, 2005). “*Language planning is called for wherever there are language problems. If a linguistic situation for any reason is felt to be unsatisfactory, there is room for a program of language planning*” (Haugen, 1966: 52).

In the Singapore context, Mandarin has been continually constituted as a “problem” through government-sanctioned review reports that mark the trajectory of Chinese language development in the national education system. As a result, the country has undertaken frequent reforms of its CL teaching system. The official discourse announcing the two most recent reforms in 2000 and 2004 (Tan, 2006) has shifted emphasis towards a newly emerging binary that appears to be assuming an increasing importance in Singapore’s language planning paradigm: the English-speaking family versus the Chinese-speaking family.

In a 1999 government-sanctioned review report, the Chinese Language Review Committee (1999) formally acknowledged that the student population in Primary 1 from English-speaking families rose from 20% in 1988 to 43% in 1999, arguing that the less CL-able students from the ESF group should be taken care of within a designated CL curriculum with its own pedagogy. In the most recent review report, the Chinese Language Curriculum and Pedagogy Review Committee (CLCPRC, 2004) found that the number of Chinese students entering Primary 1 who speak predominantly English at home has risen to 50% in 2004 compared to 36% in 1994, subsequently predicting that this number will surpass those who speak Mandarin at home in the next few years in the Singapore Chinese community. With regard to the significant shift from a predominant use of Chinese in informal domains to that of

English within the Chinese community, the Review Committee concluded that there are two major distinct groups of Singaporean Chinese children who start to learn Chinese and that these are distinguished by their different home language (HL) backgrounds. Based on this conclusion, the Committee proposed a series of recommendations for the reform of the Chinese curriculum and pedagogy which are to be implemented in 2008.

Among the most important recommendations is the so-called "Module Approach". The module approach consists of Core, Bridging, Reinforcement and Enrichment components from Primary 1 to accommodate the different linguistic backgrounds between children who come from predominantly ESF and those from CSF (see Tan, 2006). Another feature is to place less emphasis on memorizing large numbers of Chinese characters and instead to focus more on listening, speaking and reading skills. In this modular approach, students who start off with "little exposure" to CL will take the Bridging modules (in addition to the Core module) from Primary 1, and students who need further support will take Reinforcement modules (in addition to the Core module) from Primary 3. As such, these proposals are an important variation in Singapore's language in education planning in terms of Chinese competence expectations and reflect the changed sociolinguistic reality of a growing home use of English among Chinese Singaporeans.

However, the recommendations raise a number of questions, both in the area of research as well as of intervention: precisely how different are these two groups of children with different HL exposure in terms of their oral Mandarin competence, given the wider national bilingual context? Is one group markedly different from the other in terms of oral lexical coverage and/or oral fluency of Mandarin when the children begin their study of Mandarin in primary schools? How do the two groups differ in terms of their Mandarin vocabulary repertoires and the complexity of grammar they have acquired?

To answer these questions, a large-scale corpus-driven study was carried out to determine the continuum of children's oral linguistic competence in terms of the number of oral vocabulary items and the complexity of grammar they have acquired before their primary schooling commences; that is to say, the differences in Chinese Singaporean children's oral competence in Mandarin. While the overall study is still in progress, this article reports the findings

which differentiate the children in question into groups in terms of their HL backgrounds and examines how they differ in their oral vocabulary production. As a first step, three major groups were distinguished, based on information gathered on the basis of a questionnaire addressed to the parents of the children in the study (600 respondents), with respect to speaking predominantly English, predominantly Mandarin or both languages (CESF) at home. The differences in oral vocabulary coverage/size were then identified among these groups of children by applying a statistical corpus-driven analysis of about 110 h of Mandarin corpus of the children's oral production as elicited by different methods.

PREVIOUS RESEARCH

Vocabulary is an essential foundation to language learning (Nation, 1990; Schmitt, 2000). Words in a language are often compared to the basic building blocks or bricks by which a building is constructed. The number of words children have acquired naturally is found to have significant consequences on children's cognition and subsequent formal learning in schools (e.g., Clark, 1993; Cummins, 2000; Nation, 2001; Zhu, 1990). Consequently, many studies have been carried out on the development of children's vocabulary, in the context of both first and second language acquisition research on different languages (e.g. Coady & Huckin, 1997; Cook, 1996; MacWhinney, 1991; Nation, 1990, 2001; Singleton, 1999; Zhu, 1990). One strand of the research in vocabulary development focuses on determining the numbers of words that children (or L2 learners) have acquired at a particular age or that adults use in their everyday communication (refer to Adolphs & Schmitt, 2003; Clark, 1996; Li, 1995). However, we are aware of only a small number of studies that focus on an exploration of the lexical coverage of children-spoken Mandarin (see Li, 1995; Zhu, 1990 for surveys of this area). Among these, only two studies are relevant to the current study. One of the studies was "*Shi Shengshi Yanjiu* (Ten Province-and-City Study)", which was carried out in mainland China (Li, 1995). In this study, a team of researchers recorded the speech of more than 2000 Chinese children distributed across ten provinces and cities. The ages of the children ranged from 3 to 6 years. The children's tape-recorded utterances were manually transcribed and a word count was manually tallied. The published report

does not provide information on the size of the database (total collected words), but it did provide a description of the methodology, a listing of the most frequent words in the corpus, the numbers of words used by the 3 to 4 year olds (1730), the 4 to 5 year olds (2583) and the 5 to 6 year olds (3562), a brief discussion of the lexical coverage in terms of content and parts of speech and, finally, a description of the children's general vocabulary developmental tendency.

The second relevant study reports on research conducted on children's oral lexical coverage and on their grammatical complexity in Mandarin. This research project was conducted in Singapore during the 1980s (Ong, 2002). In this study, a team of researchers recorded the speech of 1200 children between the ages of 4 and 6. The researchers stimulated oral language production by showing each of the 1200 children four different pictures and conducting face-to-face dialogues which consisted of communication on the content of the four pictures: (1) in a supermarket, (2) at the beach, (3) in a children's park and (4) at a celebration at home. The speech which was elicited from the participants was recorded, transcribed and analyzed manually afterwards.

In this study, 1411 different words were collected; among which were 180 English words, while the remaining 1231 words were Chinese. It was found that the oral Chinese vocabulary used by the children covered all parts of speech in terms of grammar and were used in both simple and complex syntactical structures. Code switching was evident and natural in these children's talk, with more technical concepts expressed by English terms, such as "cartoon" and "NTUC". Based on these findings, the research team concluded that 80% of the participants could speak Mandarin fluently and that the children had no difficulty in acquiring both Mandarin and English in terms of their listening and speaking in the Singapore bilingual or multilingual context.

These two studies were monumental efforts and up to the present time provided the main evidence for the lexical coverage of pre-school children's Mandarin development in two different linguistic contexts. However, the participants involved in former study were children in a Mandarin-speaking monolingual community, which severely limited its applicability to the Singapore context. Although the latter study involved children in the Singapore context, the fact that it was completed two decades ago means that its findings are no longer relevant to the current Mandarin educational situation. More importantly, when assessed on the basis of today's more

rigorous research methodology, this study has serious shortcomings which reduce the usefulness of its findings. The principal methodological deficiency derives from the manual tallying of the data and that no total collected words (complete corpus) were reported; this latter point is important since the size of the spoken corpora is crucial in deciding the validity and reliability of the findings (Adolphs & Schmitt, 2003). As a result, neither study has direct application for the Singapore context of Mandarin education.

This brief review confirms Schiffman's (2007) observation, among others (e.g. Dixon, 2005), that there are too many untested assumptions in justifying language policy in many countries. Language planners, whether for status or corpus planning, seldom base their justifications of their reform programs on a solid empirical grounding; instead, such reform programs are more often based on self-experience reflections and self-prescription on anticipated language issues. As noted earlier, the differentiation of the Singaporean Chinese children in terms of their HL backgrounds is the major argument used by policy makers to problematize CL education and to justify current Chinese curricular and pedagogical reform, but little research literature is currently available on Singaporean Chinese children's Mandarin competence (for review, also see Liu & Goh, 2004).

As a preliminary report of a corpus-oriented research project, this study attempts to provide some evidence for status and corpus planning from an empirical perspective. For example, the analysis on the correlation between the parents' self-report survey and child's vocabulary acquisition provides a clearer picture of the relationships between the parents' language inclination and the child's oral competence. One of the interesting findings derived from the analysis on questionnaires and child's language acquisition is the neat parallelism between HL and other socioeconomic indicators – i.e. the strong performance of the children's CL perfectly coincides with the socioeconomic data provided by their respective families – behooving us to consider the difficulty of incorporating status planning in terms of language maintenance.

RESEARCH METHODOLOGY

Participants

This study includes 600 preschool children as participants recruited from both kindergartens and childcare centers. The participants are

aged from 5 to 6 years, with an equal number of boys and girls. The children were recruited from 17 public, ten church-run and nine private childcare centers and kindergartens.

Apart from demographic balance (age, gender, etc.), other factors that were taken into account in the population sampling include: kindergarten and childcare center governance, or funding sources, geographical distribution and coverage. At the present time preschool services in Singapore are run in a form of tripartite model, depending on the funding sources: public, church and private.¹ As the preschool service in Singapore is not centrally governed by a unitary official body, there are no reliable data on the demographic characteristics of the kindergarten and the accurate numbers of kindergartens functioning within each of these sectors. Despite these difficulties, we believe that we have gathered relevant data that enable us, as far as reasonably possible under local circumstances, to consider the sampled participants from the three types of kindergartens in Singapore as being broadly reflective of the sociolinguistic diversity in Singapore. The participants are also thought to be proportional to the size of the sector relative to the total number of kindergartens across the Island.

Recruiting the sample from the kindergartens and childcare centers is a convenient sampling procedure that reflects the absence of cost-effective alternatives. We concede that we have not been able to collect or identify systematic data on the social demography of the kindergarten population, let alone the linguistic characteristics of the age group involved because these data are currently not available.

This fact is also problematic for the government justification for the current CL policy changes. As noted earlier, the government has focused its language planning in its formal educational system rather than in preschools and homes where children's cultural and linguistic experiences form the foundation of their future learning (Cummins, 2000). A discussion of this policy 'bias' is not possible here, but it suffices to say that the current CL policy is formulated without empirically derived indicators or any pragmatic evidence from preschools and homes, an essential element in any language planning program. Nonetheless, our understanding is that the

¹ The two major players in the public sector are the PAP (People's Action Party) Community Fund (PCF) and National Trade Union Congress (NTUC). Childcare centers and kindergartens in governmental institutionalized sectors are known as local kindergartens as they mainly provide childcare service for surrounding communities and are normally shunned by middle and professional class families. The privately- and church-owned ones have better trained teachers, and are centrally located and well equipped; as such, they also attract children from the off-circle areas because of their prestige (Ko, 1992).

kindergarten population is broadly representative of the social demography (for a general account, refer to Khoo, 1990; Ko, 1992) and, therefore, of the broad linguistic patterns of the population.

Research Instruments

Parent Survey

With the help of the participating kindergartens and childcare centers, the parents of potential subjects were surveyed by means of a bilingual self-report questionnaire, which consists of 22 question items primarily aimed at family language use and socioeconomic variables, such as parents' education level and occupation. The information obtained from the questionnaire not only allows us to make the necessary adjustment in order to make the sampled population better distributed, but provides rich background information that facilitates the interviewing/elicitation process. Among the 953 questionnaires administered, 600 valid forms were received and processed.

Interview

The actual fieldwork associated with the data collection started with an unstructured interview by four trained female research associates. To make the most of this oral data from the participants, the interview was conducted in a casual manner, which normally started off with a brief warm-up dialogue, followed by pre-determined topics that were assessed to be most relevant and familiar to the child's daily life, such as their personal experience, their interests, family outings and their favorite stories. To maximize spontaneity, the talk was totally open-ended and any child-initiated topic was encouraged. All participants were interviewed individually in a quiet room in order to guarantee the quality of the recording.

Narrative/Picture Elicitation

In this study, there were a number of participating children from ESF or English-Chinese mixed family backgrounds that lacked sufficient ability to talk in Mandarin freely, while some, due to various psychological reasons (e.g. autism), found verbal expression difficult. Under such circumstances, professionally designed pictures were employed to elicit utterances from these children to the maximum extent possible. In cases where such children were not able to produce any sustained utterance, the researchers encouraged these children to name/translate the objects on the pictures. If they still could not speak anything meaningful, or largely used English or

languages other than Mandarin, then the interview was discontinued. At all times the guiding principle was to exhaust all possible means to elicit children's oral production.

Both the interviews and picture elicitations (about 30 min for each participant) were recorded by Olympus DM-20 digital voice recorder.

Classroom Observation

The participants' linguistic behaviors in the classroom were observed with a special focus on what and how they spoke in a 30-min-long Mandarin class. As expected, the oral production obtained from the classroom observations were found to be characteristic in lexical coverage and syntactic complexity, and the gain was especially large in terms of receptive vocabulary in classroom communicative mode, ranging from story-telling, role-play activities to group discussion. Two classes in each sampled kindergarten were observed, and individual background information for each observed child was tagged. This yielded about 36 h of video-recorded data.

Oral Data Transcription and Corpus Construction

The ultimate goal of this project is to compile a multi-modal corpus of Singapore preschool children's oral language in Mandarin. To date, the oral recordings of 600 participants (300 h) and 24 video-taped classroom observations (12 h) have been transcribed. However the data are still in a cleaning-up process. Therefore, the following analysis of vocabulary production is based on the verified data (180 participants) that has been completed so far. About 110 h of transcribed oral data was automatically processed by SCORE, an education corpus software developed by CRPP at Nanyang Technological University (for details, see Hong et al., 2005). The total corpus size is 278,948 words produced by the children (among which 236,107 Mandarin words and 42,841 English words in terms of token words).

RESULTS AND DISCUSSION

Defining the Child's Language Use at Home

As noted earlier, the Singaporean Chinese children of this study were differentiated into two main groups in the official and public discourse, namely the ESF and the CSF groups (also see Tan,

2006). This distinction is typically made on the basis of one question item on the language use at home that appears in the parents' self-questionnaire that is completed with the student's admission to primary school (CLCPRC, 2004: 4); this item commonly consists of the question "which language do you use as a first language or predominantly at home?", or something similar. We assume that the distinction made in this way is too simplistic to capture the complicated phenomenon of language use at home in the Singapore multilingual society.

In order to have a better understanding of which language the preschoolers use predominantly at home, we used a multiple factor approach to define preschoolers' language use at home. The approach includes (see Table 1): (1) parent's language use to their child(ren) (e.g. English or Mandarin or both); (2) parents' and children's TV/radio viewing or listening in terms of language channels (English or Mandarin or both); (3) reading materials in English or Mandarin or both; (4) children's language use with their peers.

We assume that these are important factors that can be considered to contribute to determining children's predominant language use at home, with each factor contributing individually to language use (=contributing value). The contributing values are weighted in descending order as shown in Table 1. The parents' language use to their child was assigned the most weight (50%), followed by, in descending order of weight, viewing and/or listening to TV/radio programs in terms of English or Mandarin (30%), reading materials in English or Mandarin (15%) and children's language use with their peers (5%).

TABLE 1

The contributing factors to a child's home language (HL) use.

Category I (50% weight)		Category II (30% weight)		Category III (15% weight)		Category IV (5% weight)	
Parent language		TV/Radio channels		Reading at home		Playing at home	
Mother (M)	Father (F)	M + F	Child	M	F	Child	Peer ²
25%	25%	15%	15%	5%	5%	5%	5%

² Many Singapore families employ maids as household assistants or caregivers, and most of these maids are from Indonesia and the Philippines. As they are non-Mandarin speakers, they have a negligible influence on the acquisition of Mandarin by the children in their care.

The four factors were computed against the different percentage weighting of the quantifiable variables and the scores plotted with each participant's individual number; as expected, most children distinctly fell within the middle range. Specifically, 30% of the participants were categorized as originating from the CSF group in our sampled population; by the same standard, 24% were from the ESF group and 46% from the ECSF Group.

We must admit that the factors and their contributing values used here may be problematic, but we can argue that, in comparison to a conclusion drawn from a one-question-item on the parents' self-reports, our classification has taken into account the most conceivable factors that will affect a child's language use at home. Consequently, three groups, each consisting of 60 children, were chosen on the basis of this new categorization approach.

Oral Mandarin Competence: Comparative Analyses of the Three Groups

Many measuring tools have been developed and used in vocabulary acquisition studies (e.g. see Malvern et al., 2004 for review). For example, the Reynell developmental language scale (Reynell & Gruber, 1990) was devised to measure a child's passive vocabulary and typically used by language impairment researchers for remedial purposes. Laufer and Nation (1995) developed the Lexical Frequency Profile to estimate productive vocabulary size, which is often used to determine whether a particular text is suitable for a particular group of learners at a particular proficiency level. Other well-known tools include VOCD, or the D value (e.g. Duran et al., 2004) and MLU-5 (Hong Kong University, 1999).

While there are many methodological approaches to measuring and describing children's vocabulary development, for our purpose we use three measurements (e.g. Brown, 1973; Duran et al., 2004; Malvern et al., 2004) to assess oral vocabulary competence in the Mandarin of the three study groups:

- vocabulary size, or NDW (number of different words);
- the mean length of utterance (MLU);
- type and token ratio (TTR).

Table 2 summarizes the results of the four parameters (Mandarin word type, Mandarin word token, English word type and Mandarin sentence number) obtained from the initial results on the 180

TABLE 2

Child's oral production across the three groups of children.

Groups by HL	Chinese types	Chinese tokens	English types	Chinese sentences
CSF ($n = 60$)	2825	94,831	1590	18,049
ECF ($n = 60$)	2259	73,387	2089	17,896
ESF ($n = 60$)	2020	64,289	2989	16,345
Total	3814	236107	4422	52,290

children's oral production.³ Drawing on the tools indicated above, the data reveal language development/competence or language acquisition among these children. The word number in the columns of "Type" is normally considered to be a good indicator of language achievement, while the "token number" is considered to be an indicator for determining verbal productivity or talkativeness. Remarkable differences were found in the four parameters among the three groups. As expected, the CSF children achieved the highest in the Mandarin word type, Mandarin word token and Mandarin sentence number, followed by the other two groups. In contrast, the ESF children achieved the lowest in these three parameters but used many more English words (word type) than the other two groups in their Mandarin oral production.

On average, the CSF group produced 2825 different Chinese words, while the ECSF group produced 2259 different words; the ESF group produced 2020 different words in their total oral vocabulary production (as shown in the column "Type"). There is about a 566 word difference between the CSF and ECF groups, while the difference between the CSF and ESF groups is much larger (805 words). Similarly, there are also notable differences between the three groups in terms of word total production (as shown in "Token"). The CSF group produced 94,831 words in "Token", whereas the ESF group used 73,387 words and the CESF group used 64,289 words. The reverse is true for the differences between the three groups in terms of English word type. While 2989 different English words in total were used by the ESF group in the Mandarin communication, the numbers for the ECSF (2089) and

³ The word segmentation standard used for the present corpus is *Processing Standard for Modern Text Corpus Segmentation and Annotation-973* (Shanxi, 2004). Hesitations, immediate self-repetitions, onomatopoeia, re-formulations and unintelligible utterances were annotated in the corpus but not reported here.

the CSF group (1590) were much smaller. There are nearly twofold more English words used by the ESF children (2989) than by the CSF group (1590 words). However, the total number of Mandarin sentences produced by the CSF group (18,049) is just slightly higher than that by the CESF group (17,896). A comparison of the CSF group with the ESF group (16,345) reveals that the difference is not large – approximately 10%. The differences in terms of sentence type are much smaller than those of both Mandarin vocabulary size and English vocabulary size.

These figures show that the CSF children have a much bigger Mandarin vocabulary repertoire and are much more productive or verbal than their CESF and ESF counterparts in expressing themselves in Mandarin communication. Between the CESF and ESF groups, the former performs better than the latter, but the differences are not very large in terms of their vocabulary repertoire and productivity in their Mandarin communication. Of more significance are the differences between the CSF and the ESF groups in terms of the sizes of the Mandarin word type, Mandarin word token and the English word type. The most significant difference is that the ESF group produced more English words (shown as in type size) than the CESF and CSF groups did in Mandarin communication. This means that the ESF group had to code-mix or/and switch more frequently between Mandarin and English than the other two groups. In other words, the ESF group relied more on frequent code-mixing and/or switching in Mandarin communication. This is correlated with the marginal differences found between the three groups in terms of the numbers of sentence. These marginal differences suggest that the ESF group was as nearly equally communicative as the other two groups in expressing themselves, but that these children relied heavily on English words rather than English syntax.

Comparing Oral Efficiency and Proficiency: Differences in TTR

Vocabulary richness, or lexical diversity, has long been considered one of the important indices for assessing first and second language proficiency. TTR, the best known quantitative tool to measure the diversity and density (Daller et al., 2003: 200), was also used in this study to measure the child's oral vocabulary competence. While word type/NDW indicates word range and token indicates the talkativeness, TTR, as the combination of vocabulary size and the ability to use it effectively, shows a child's linguistic skill of being able to talk a lot by using a wide range of unrepeated vocabulary on

various topics. The predictability of TTR reduces as the token number increases. This is particularly true for oral speech of a young child, as a child's vocabulary repertoire is relatively small and the token can be unlimited given the time they are allowed to talk is long enough. This somewhat reduces the value of TTR. Various mathematical transformations have been explored and designed to compensate for the sample size effect, but no measurement performs perfectly satisfactorily, as concluded by Malvern et al. (2004: 30). Further discussion on methodological issues is beyond the scope of the present theoretical justification; however, taking the viability and our sample features into account, Root TTR (RTTR) was found to be more in line with our expectations and thus chosen to quantify the dimensions of lexical sophistication for its high reliability, excellent performance in measuring child's vocabulary diversity (Daller et al., 2003: 200), and more importantly, for its pertinence to our data. RTTR, which is represented by the index of *G* as it was first used by French researcher Guiraud in 1954, is expressed as $G = \text{types}/\sqrt{\text{tokens}}$. The result of multiple comparisons in terms of the RTTR in Table 3 shows a very strong interaction ($p < 0.001$) between a child's HL and the variety of the words used by different groups of children.

One-way ANOVA testing result ($F = 52.220$, $p < 0.000$) also shows there is an obvious difference between the three groups in terms of verbal creativity. Table 3 shows that once again the *G* score in CSF (8.1897) is higher than that in the other two groups and that the lowest score was that of the ESF group (5.6703). Furthermore, while the mean difference between the CSF and the ECSF is 1.08100, and that between the ECSF and the ESF 1.43835, the biggest disparity of mean difference is found between the ESF group and CSF group (about 2.5-fold). This is an interesting outcome that underlines the bilingual features of the two groups. If we contrast this finding with the difference in MLU in Table 4 between the same two groups

TABLE 3

Multiple comparisons of *G*-value across three groups of family language.

Family language	Mean	Standard deviation	Mean difference	Significance
CSF ($n = 60$)	8.1897	1.30597	1.08100*	000
CESF ($n = 60$)	7.1087	1.35359	1.43835*	000
ESF ($n = 60$)	5.6703	1.38946	-2.51936*	000

* $p < 0.001$.

TABLE 4

Multiple comparisons of MLU between three groups of family language.

Family language	Mean	Standard deviation	Mean difference	Significance
Chinese family ($n = 60$)	4.7633	2.76768	0.51913	0.339
English + Chinese ($n = 60$)	4.2442	1.88574	0.53732	0.31
English family ($n = 60$)	3.7069	1.32590	-1.05645*	0.021

* $p < 0.05$.

at just -1.056 , the discrepancy between the comparable parameters is worth noting. These numbers show that the ESF group had a limited Mandarin vocabulary size and tended to repeat the words within the limited size in their Mandarin oral communication, while the CSF child had a relatively larger Mandarin vocabulary repertoire and tended to use a wider variety of vocabulary at their disposal. In order to further study these findings, we used another important index as supplementary to the above two measurements.

Comparing Development Maturity – Differences in MLU

Mean length of utterance is the summed number of word tokens of a child's utterance divided by the total number of sentences/utterances in a sample. It is considered to be a well-established index for measuring the maturity of a child's language development⁴ (Malvern et al., 2004: 17; Rescola et al., 2000: 464). In other words, MLU can be used as an overall measure of language growth since it reflects the expanding grammatical ability in the child.

As MLU measures the number of words a child can use in the same sentence, it is more of an indicator of syntactic complexity than vocabulary. On the other hand, it is obvious that there is a correlation coefficient between two variable of vocabulary size and sentence length, or syntactic complexity. To justify the validity of MLU in our samples, a two-tailed coefficient of correlation between vocabulary size and sentence length was calculated. The result indicates a strong coefficient correlation (at 0.01 level) for all three groups of children. Specifically, the correlations appear in an ascending order: $r=0.503$ for the ESF group, $r=0.594$ for the

⁴ There is disagreement as to how to apply MLU to analyze Chinese. Some researchers use Chinese characters as an utterance unit (e.g. Peng, 1984); other researchers use words as a measuring unit (e.g. Zhu & Lin, 1986). The unit refers to a transcription unit of child talk, which can be defined by one speaker bounded either by transition in speakers, by grammatical closeness and/or by a long pause.

ECSF group and $r=0.730$ for the CSF group. Table 4 shows the testing results of multiple comparisons of MLU between three groups of children.

The result of One-way ANOVA ($F=30.429$, $p<0.000$) shows that the difference between the three groups in terms of the sentence length is only marginal, with only the ESF group and CSF group significantly different from each other (mean difference is -1.05645 , $p<0.021$). In other words, when it comes to the sentence quality in terms of the word number contained in sentences, the sentences produced by the ESF children were as nearly long as those from those of the ECSF children (mean difference is 0.53732 , literally about $\frac{1}{2}$ word), and this is also true for the difference between the ECSF and CSF children, with a mean difference of 0.51913 , namely, also about $\frac{1}{2}$ word. This marginal difference might result from the use of code-mixing, or more accurately, "intra-sentential alternation" (Torres, 1989). The marginal differences found across the three groups seem to suggest that the structural development has largely reached the similar milestone level at this age cohort among all sampled participants. If code-switching and code-mixing are taken as a normal sociolinguistic reality in the Singapore bilingual society, we would have to assume that Singaporean Chinese children have been developing well bilingually and have no difficulty in expressing themselves appropriately in oral daily communication. It would therefore seem that there is no acute need either to emphatically warn preschool service providers and parents of the seriousness of an unbalanced linguistic development by Singapore children or to worry about the difficulties the children encounter in dealing with the two languages simultaneously (e.g., Ko, 1992; Lee, 1992a, b).

If we return to Table 2, the most notable points are the large differences between the three groups in Mandarin word type, token and English word type, and the marginal differences (ANOVA $F=2.384$, $p<0.095$) between the three groups in the mean number of sentences. If we also take the results of the MLU revealed in Table 4 into account, we can see that there are significant differences in terms of Mandarin word type across three groups ($p<0.05$), but when it comes to the sentence number and sentence length, the differences among the three groups are quite small. This implies that the ESF group is nearly as proficient as the other two groups but that these children rely heavily on English in Mandarin communication. These analyses are quite preliminary, and a better understanding of the different Mandarin competencies among the three groups

children requires further analyses of their oral data, such as syntactic patterns, discourse structures and pragmatic strategies.

CONCLUSION AND DISCUSSION

The Singapore government's approach to language planning and management is generally seen as essentially pragmatic, but policy discourse is changing (e.g. Shepherd, 2005; Tan, 2006). The recent shift in CL policy emphasizes differences between the two groups of children (the ESF and the CSF) in Mandarin learning at schools and rationalizes current policy changes in relation to the perceived problem as described at the beginning of this article. While recognizing that English may begin to dominate the linguistic field of HL use in Singapore, we assume that the policy discourse has overlooked the complexity of children's HL use and their Mandarin competence in a context where bilingual policy has been implemented for some 40 years. Limited data have been used in the determination of policy. If research such as that reported here were to be included in policy discussions, it could counter the binary approach that is applied at the present time. The present policy derives from an over-reliance on a one-question demographic survey that is used as the principal indicator for selection and implementation of programs in early schooling. In this article, we have used a set of variables rather than a single indicator to categorize Singaporean Chinese children into three major groups rather than two groups in terms of their language use at home. We assume that this distinction amounts to a much more robust evaluation of the differences between these children in terms of their language use at home than what is claimed to be in the policy and public discourse. This distinction suggests that the government binary claim of the CSF and ESF groups in their CL planning in education may be less convincing than it is made out to be.

The results show that there is a strong correlation between the children's Mandarin oral competence and their language use at home. As expected, the three groups have developed different Mandarin oral proficiencies in relation to their different exposure to language use at home. The CSF group has a much bigger Mandarin vocabulary repertoire and is much more productive or verbal than the children of the CESF group who, in turn, outperform their ESF counterparts in expressing themselves in Mandarin communication. The ESF group has to code-mix or/and switch more frequently between Mandarin and English than the other two

groups. In other words, the ESF group relies more on frequent code-mixing and/or switching in Mandarin communication.

However, considerable variations within each group in multiple comparisons of vocabulary size seem to suggest that the differences in the vocabulary achievements between the two or three groups of children may not be as great as discussed earlier, particularly if we take into account the marginal differences between the three groups in terms of sentence number and sentence length. These marginal differences seem to suggest that the continuum of CL competence is much broader than assumed in terms of the relationship between these children's linguistic use at home and their actual level of language achievements. In addition, most of the 180 participants of this study fall in the middle range between the two ends; this would seem to confirm that, in a bilingual and multicultural society like Singapore, there are "more bilinguals than monolinguals" (Xu & Li, 2003: 152). Therefore, given the early stage of this study, we are reluctant to make any strong casual statements with respect to the findings; instead, we report here the actual patterns of language achievements within the framework of whether or not these patterns are correlated with these children's language use at home.

As noted earlier, we used more variables and took a holistic stance in differentiating the participants in terms of their language use at home. However, in reality, language use in a bilingual or multilingual society like Singapore is very complex and shifting; for example, the language use can vary from both parents using a mixed code to one parent using one dominant single language and the other parent using another language or a mixed code. In any case, the two-group division of Singaporean Chinese children (ESF: 50%, CSF: 50%) in terms of their language exposure, as claimed by the government and circulated as public policy, is seriously problematic. In this study, only about 4.2% of the children in the ESF group and 0.8% of those in the CSF group can be considered to be limited to a relatively pure monolingual exposure when all the variables used in this study are taken into consideration.

In a multilingual society, ubiquitous multilingual communications will always expose children to more than one language influence. Therefore, it would be reasonable to assume that the more factors being taken into consideration, the more the dividing lines between different groups will be likely to become blurred. However, our discussion here should not be taken as a challenge to the underlying questions that motivated us and the government to identifying

different groups. In our view, it is absolutely correct to identify the major differences between children, both in terms of their language use at home and their actual language proficiency, as these are crucial elements in developing CL policy in education. What we emphasize here is that these major differences between children require a critical understanding if they are to be used to assess the differences of Chinese learning ability between the children from different language backgrounds. We argue here that, despite the obvious differences found between the three groups, more variables are needed to define children's language use at home, but this is clearly an empirical question that needs further study.

We hope that this article will stimulate further debate and study about how we define groups of children in terms of their language use at home and their mother tongue competence in an English-dominant multilingual society as Singapore. In addition, the results may lead to the further development of more plausible approaches and reliable tools which would allow us to make meaningful judgments about the relationship between children's language use at home and their language competence before their formal education actually begins.

ACKNOWLEDGEMENTS

The authors wish to express their gratitude to the project team members, all of the participating kindergartens and the families. Apart from the authors of this article, the main researchers in the project are: Hock Huan, Goh, Sze Win, Gan, Hwee Bin Wendy, Toh and Yimin, Wang. Special thanks go to Wendy and Yimin for their valuable technical support, and to Joseph Lo Bianco for his corrections and comments on earlier versions of this article.

REFERENCES

- Adolphs, S. & Schmitt, N. (2003). Lexical coverage in spoken discourse. *Applied Linguistics*, 24(4), 425–438.
- Brown, R. (1973). *A first language: The early stages*. Cambridge, Mass: Harvard University Press.
- Chinese Language Review Committee (1999). *Report of the Chinese Language Review Committee*. Singapore: Ministry of Education.
- Clark, E. (1993). *The lexicon in acquisition*. Cambridge: Cambridge University Press.

- Clark, E. (1996). Later lexical development and word formation. In F. Paul & B. MacWhinney (Eds.), *The handbook of child language*. Oxford: Blackwell.
- CLCPRC (Chinese Language Curriculum and Pedagogy Review Committee) (2004). *Report of the Chinese Language Curriculum and Pedagogy Review Committee*. Singapore: Ministry of Education.
- Coady, J. & Huckin, T. (1997). *Second language vocabulary acquisition*. Cambridge: Cambridge University Press.
- Cook, V (1996). *Second language learning and language teaching*. London: Arnold.
- Cooper, R. (1989). *Language Planning and Social Change*. Cambridge: Cambridge University Press.
- Cummins, J. (2000). *Language, power and pedagogy: Bilingual children in the crossfire*. Clevedon: Multilingual Matters.
- Daller, H., Van Hout, R. & Treffers-Daller, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24(2), 197–222.
- Dixon, L. (2005). Bilingual education policy in Singapore: An analysis of its sociohistorical roots and current academic outcomes. *International Journal of Bilingual Education and Bilingualism*, 8(1), 25–47.
- Duran, P., Malvern, D., Richards, B. & Chipere, N. (2004). Development trends in lexical diversity. *Applied Linguistics*, 25(2), 220–242.
- Goh, N. (2004). *Huayuwen zai xinjiapo de xianzhuang yu qianjing (The current practice and prospect of Chinese language education in Singapore)*. Singapore: Chuangyiquan Chubanshe.
- Gopinathan, S., Pakir, A., Ho, W. & Vanitha, S. (Eds.) (2003). *Language, society and education in Singapore: Issues and trends*, Singapore: Marshall Cavendish International.
- Haugen, E. (1966). Linguistics and language planning. In W. Bright (Ed.), *Sociolinguistics: Proceedings of the UCLA Sociolinguistics Conference, 1964* (pp. 50–71). The Hague: Mouton.
- Hong Kong University (1999). Xueqian ertong kouyu (yueyu) fazhan de zhuyao guocheng (The major process of preschoolers' oral language (Cantonese) development). Retrieved March 8, 2006, from: http://cd1.emb.hkedcity.net/cd/solar/html/rindex_c02095a_cn.htm.
- Hong, H., Kazi, S. & Sim, T. (2005). SCORE: A multimodal corpus database of education discourse in Singapore schools. *Proceedings of Corpus Linguistics, Birmingham Conference*. July, 14–17, 2005.
- Khoo, K. (1990). An alternative model of childcare: The experience of the National Trades Union Congress of Singapore. In B. P. Chan (Ed.), *Early childhood toward the 21st century* (pp. 205–216). Hong Kong/Gaithersburg: Yew Chung Education Publishing.
- Ko P. (1992). The cognitive and social development of pre-school children in Singapore: What have we learned from the IE-BvLF study? In P. Ko & W. Ho (Eds.), *Growing up in Singapore: The pre-school years* (pp. 135–168). Singapore: Longman Singapore Publishers.
- Laufer, B. & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307–332.
- Lee, F. (1992a). Children's Language: How do our children communicate? In P. Ko & W. Ho (Eds.), *Growing up in Singapore: The pre-school years* (pp. 52–80). Singapore: Longman Singapore Publishers.

- Lee, F. (1992b). Children's talk and play: Do our children's behaviours reflect their parents and homes. In P. Ko & W. Ho (Eds.), *Growing up in Singapore: The pre-school years* (pp. 105–34). Singapore: Longman Singapore Publishers.
- Li, Y. (1995). *Ertong yuyan fazhang (Children language development)*. Wuhan: Huazhong Shifan Daxue Chubanshe.
- Liu, Y. & Goh, H. (2004). *Review of Chinese language education research in Singapore: Making a case for alternative research orientation*. (CRPP Technical Paper), National Institute of Education: Nanyang Technological University.
- Lo Bianco, J. (2007). Advantage and identity: Neat discourse but troubled union: Singapore's medium of instruction policy. In V. Vinito, G. Saravanan, & Y. Liu (Eds.), *Language, capital, culture: Critical studies of language in education in Singapore*. Amsterdam: Sense Publishers (in press).
- MacWhinney, B. (1991). *The CHILDES project: Tools for analyzing talk*. Hillsdale, New Jersey: Lawrence.
- Malvern, D., Richards, B., Chipere, N. & Duran, P. (2004). *Lexical diversity and language development*. New York: Palgrave Macmillan.
- Nation, I. (1990). *Teaching and learning vocabulary*. New York: Newbury House.
- Nation, I. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Ong, Y. (2002). Xinjiapo xueqian ertong huayu kouyu cihui (The oral vocabulary repertoire of Singapore pre-school children). In Y. Ong (Ed.), *Zhonghua yuyan wenhua yu jiaoxue (Chinese language, culture and teaching)* (pp. 110–25). Singapore: Xishan Wenyi Zhongxin.
- Peng, Z. (1984). 3 zhi 6 sui ertong yuyan fazhan de chubu yanjiu, zhongguo xinli xuehui nianhui lunwen ji (A preliminary analysis of the child language development from 3 to 6 years old). *The proceedings of 1984 Annual Conference of China Association of Psychological Studies*.
- Reynell, J. & Gruber, C. (1990). *Developmental language scales*. Los Angeles: Western Psychological Services.
- Rescola, L., Dahlsgaard, K. & Roberts, J (2000). Late-talking toddlers: MLU and IPSyn outcomes at 3 and 4. *Journal of Child Language*, 27 643–664.
- Schiffman, H. (2007). Tamil language policy in Singapore: The role of implementation. In V. Vinito, G. Saravanan, & Y. Liu (Eds.), *Language, capital, culture: Critical studies of language in education in Singapore*. Amsterdam: Sense Publishers (in press).
- Schmitt, N. (2000). *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Shanxi D. (2004). Dangdai hanyu wenben yuliaoku fenci, cixing biao zhu jiagong guifan-973. (The standard for segmentation and speech annotation of modern Chinese text - 973). Retrieved March 8, 2006, from <http://www.chineseldc.org/EN/doc/CLDC-LAC-2003-003/label.htm>.
- Shepherd, J. (2005). *Striking a balance: the management of languages in Singapore*. Frankfurt am Main: Peter Lang.
- Singleton, D. (1999). *Exploring the second language mental lexicon*. Cambridge: Cambridge University Press.
- Silver, R. (2005). The discourse of linguistic capital: Language and economic policy planning in Singapore. *Language Policy*, 4(1), 47–66.
- Tan, C. (2006). Change and continuity: Chinese language policy in Singapore. *Language Policy*, 5(1), 41–62.

- Torres, L. (1989). Code-mixing and borrowing in a New York Puerto Rican community: A cross generation study. *World Englishes*, 8(3), 419–432.
- Xu, D. & Li, W. (2003). Linguistic dirigisme and language management in Singapore. In M. R. Phyllis & R. Terborg (Eds.), *Language: Issues of Inequality* (pp. 122–54). Mexico: Universidad Nacional Autonoma de Mexico.
- Zhu, M. (1990). *Ertong yuyan fazhan yanjiu (Studies on child language development)*. Shanghai: Huadong Shifan Daxue Chubanshe.
- Zhu, Z. & Lin, C. (1986). *Siwei fazhan xinlixue (Psychology of cognitive development)*. Beijing: Beijing Shifan Daxue Chubanshe.

ZHAO SHOUHUI, LIU YONGBING

*National Institute of Education
Nanyang Technological University
1 Nanyang Walk, 637616,
Singapore*

HONG HUAQING

*National University of Singapore, 119077
Singapore*