



Archived at the Flinders Academic Commons:

<http://dspace.flinders.edu.au/dspace/>

This is the authors' version of an article accepted for publication in the *International Journal of Speech Language Pathology*. The original is available by subscription only at:

<http://www.tandfonline.com/doi/full/10.3109/17549507.2011.652171>

Please cite this article as: Teoh, W.Q., Brebner, C.M. and McCormack, P.F., 2012. Assessing the language skills of children from culturally and linguistically diverse backgrounds: The expressive vocabulary performance of Singaporean English-Mandarin bilingual preschoolers. *International Journal of Speech-Language Pathology*, 14(3), 281-291.

DOI:10.3109/17549507.2011.652171

© 2012 The Speech Pathology Association of Australia Limited

Please note that any alterations made during the publishing process may not appear in this version.

Assessing the language skills of children from culturally and linguistically diverse backgrounds: The expressive vocabulary performance of Singaporean English-Mandarin bilingual preschoolers

Authors: Teoh Wei Qin¹, Chris Brebner², Paul McCormack²

¹Department of Child Development, KK Women's and Children's Hospital, Singapore

²Speech Pathology and Audiology, Flinders University, South Australia

Email correspondence: weiqin.teoh@gmail.com

Mailing address: 75 Loyang Rise Singapore 507526

Running head: Assessing the language skills of Singaporean children

5 key words: Singapore, bilingual, cultural & linguistic diversity, language assessment, test bias

**Assessing the language skills of children from culturally and linguistically diverse
backgrounds: The expressive vocabulary performance of Singaporean English-Mandarin
bilingual preschoolers**

ABSTRACT

Standardised language assessments such as the *Clinical Evaluation of Language Fundamentals Preschool 2 United Kingdom (CELF Preschool 2 UK)* (Wigg, Secord & Semel, 2006) are often used in speech-language pathology clinics to determine if a child is at risk of language difficulties. Many of these assessments are designed and standardised for use with monolingual Standard English speaking children. It is thus recommended that these assessments should only be used with the populations they were designed for; if not test bias might result. However, such tests are still selected and used in the clinics of many multicultural and multilinguistic communities (e.g. Singapore). This research aimed to explore the performance of Singaporean English-Mandarin preschool children on the *Expressive Vocabulary (EV)* subtest of the *CELF Preschool 2 UK* and to determine if their performance on the *EV* subtest accurately reflected their language abilities by comparing their performance on a local screening language assessment tool, the *Singapore English Action Picture Test (SEAPT)* (Brebner, 2002). Results showed that local children performed poorly as compared to their UK counterparts. Two plausible reasons for the findings are: 1) the subtest elicited only a single measure in English which ignored the language abilities of these bilingual children in their second language; 2) the presence of culturally and linguistically biased test items.

INTRODUCTION

There has been an increasing demand to provide equitable speech and language services to children from culturally and linguistically diverse (CALD) backgrounds, such as Singapore. International literature has shown that the assessment of these children is often complicated by issues such as: limited understanding of the development of local languages; few, if any, locally developed language assessment tools; and the lack of local normative data for popular standardised language assessments (Brebner, Rickard-Liow & McCormack, 2000; Carter, Lees, Murira, Gona, Neville & Newton, 2005; Pickering & McAllister, 2000).

To make an accurate assessment and diagnosis of a child's language difficulties, speech-language pathologists often select commercially available standardised language assessments to determine whether a child is at risk of language difficulties by comparing the child's performance to that of their peers. In Singapore, such language assessments are also a popular choice among local clinicians. However, these formal assessment tools are often constructed and standardised for monolingual Standard English speaking children who are exposed to different cultural and linguistic settings when compared to the Singaporean children. The use of such culturally and linguistically unfamiliar language assessment tools in Singapore may result in clinical misdiagnosis which could lead to reduced efficiency and efficacy of local speech-language pathology services. Before detailing the specific problems that may arise when such language assessments are used in Singapore, it is important to have an understanding of the language patterns used locally.

English in Singapore

There are two varieties of English spoken in Singapore: Singapore Standard English (StdSE) and Singapore Colloquial English (SCE). StdSE is grammatically similar to the Standard English spoken elsewhere but with variation in accent, whereas SCE is a non-standard form of the English language, differing considerably from other forms of Standard English. Gupta (1994) described SCE as a pragmatically rich language with simple grammar and morphology. The morphology and syntax features of SCE are summarised in Table 1. Although StdSE is the official language of education in Singapore (Teo, 2000), SCE has progressively become the first choice of language spoken in the local community for everyday communication, especially during interactions with and among young children (Gupta, 1994).

The Prevalence of Bilingualism in Singapore

Singapore is a multicultural and multilingual society with four official languages (English, Mandarin, Malay and Tamil) and a variety of unofficial languages (e.g. Chinese dialects) (Gupta, 1994). The languages used by the children's caregivers determine the number of languages the children are exposed to. The diversity of language environments within each family and in Singapore results in most children acquiring more than one language simultaneously. Thus, bilingualism, or even multilingualism, is prevalent. Bilingualism is further emphasised when formal education commences as Singaporean children are required to learn at least two languages, English and a mother tongue, as part of Singapore's bilingual language policy (Ministry of Education, 2008).

Although research suggests that there is no outstanding difference between bilingual children and monolingual children in terms of developmental milestones (Nicoladis & Genesee, 1997; Paradis,

Crago, Genesee, & Rice, 2003; Westman, Korkman, Mickos & Byring, 2008), there are still some subtle differences that need to be taken into account. For example, vocabulary knowledge in bilingual children is said to be distributed between both languages (Bedore, Peña, García & Cortez, 2005; Peña, Bedore, and Zlatic-Giunta, 2002). In other words bilingual children have vocabulary knowledge that is said to be unique to each language, although translated equivalents may exist in both languages. This distribution of vocabulary among both languages can be represented by the Dual System Model. The model states that bilingual children often have a separate linguistic system for each language although some interaction between the two linguistic systems may occur (Genesee, 1989). Many studies (Bedore et al., 2005; Hemsley, Holm & Dodd, 2010; Junker & Stockman, 2002; Pearson, Fernández, Oller, 1993; Peña et al., 2002) have shown that bilingual children often perform more poorly than their monolingual counterparts on a single language vocabulary measure. However, when the lexicon knowledge in both languages is taken into consideration while disregarding translation equivalents, the conceptual scores obtained by the bilingual children are often comparable, or higher, than those of monolingual children.

The use of SCE and prevalence of bilingualism have highlighted the fact that the developmental patterns towards English language competence among Singaporean bilingual children are different from that of their monolingual Standard English speaking counterparts.

Language Dominance in Singaporean Bilingual Children

Although, most bilinguals in Singapore acquire more than one language simultaneously, it is common for them to be dominant, or more proficient, in one language. A language is considered dominant when it is preferred and frequently used in comparison to another language (Law & So, 2006). This is especially true with young Singaporean children as they are still at the early stages

of learning both languages. A Singaporean bilingual child's level of proficiency in each of the languages he or she speaks is dependent on the child's exposure and usage of each language in school, at home and in the community (Bedore et al., 2005; Uchikoshi, 2006).

Issues associated with the use of Standardised Language Assessment Tools in Singapore

Literature indicates that bilinguals often perform below the mean on most standardised assessments (Crutchley, Conti-Ramsden, Botting, 1997; Qi, Kaiser, Milan & Hancock, 2006). This reflects the reality that commercially available standardised language assessments are not suited for use with bilingual children as they are standardised on a very narrow population set of monolingual children (Dollaghan & Horner, 2011; Isaac, 2002; Kohnert, 2010; Pearson et al., 1993). Even when tests claim to include culturally diverse populations and bilinguals in the collection of normative data, these samples are usually made up of only a small percentage of participants (Vaugh-Cooke, 1983).

When a bilingual child is assessed using an assessment designed and standardised for a different population, test bias often results. Test bias occurs when two people of the same ability but from different cultural groups do not have the same probability of success on a test (Goldstein, 2006). There are two forms of test bias (Brebner, 2010) that frequently arise in Singaporean clinical settings. Firstly, cultural bias can occur when Singaporean children are asked to identify unfamiliar content in the standardised assessments. As content in standardised assessments are often designed for the monolingual standard English speaking population, Singaporean children may or may not have been exposed to the same language concepts and vocabulary (e.g., winter clothing such as mittens is not a common sight in tropical Singapore). Secondly, linguistic bias occurs when a local child is expected to respond in Standard English, a second or third language

for some Singaporean children. Most language assessments are designed to measure Standard English development while most young Singaporean children speak SCE, a non-standard form of English. As materials, content and structure of these language assessments may be unfamiliar to Singaporean children, a local child's test result will not be truly representative of his or her language ability (Brebner et al., 2000; Carter et al., 2005; Gutiérrez-Clellen & Peña, 2001; Li, Miller, Dodd & Hua, 2005). Consequently, the chances of misdiagnosis increases as the content and normative data of these language assessments are not sensitive to cultural and linguistic differences.

Purpose of study

This study aimed to examine the performance of Singaporean English-Mandarin bilingual children on the *Expressive Vocabulary (EV)* subtest of the *Clinical Evaluation of Language Fundamentals Preschool 2 – United Kingdom (CELF Preschool 2 UK)* (Wigg, Secord & Semel, 2006) to determine whether the *EV* subtest was able to measure the true language abilities of Singaporean English–Mandarin bilingual children.

In this study, the performance of Singaporean English-Mandarin bilingual children was first compared to the UK standardisation sample as a whole. The researcher acknowledged the fact that these children could not be grouped together as a homogenous group without considering the effect of language dominance. Therefore, the collected data were further divided into two groups: English Dominant (ED) and Mandarin Dominant (MD).

The questions and hypotheses of the research were:

- (i) How did the test performance of the Singaporean sample differ from the UK standardisation sample? *The Singaporean sample would perform poorly when compared to the UK standardisation sample.*
- (ii) Did the difference in language dominance (ED group versus MD group) affect the type/range of scores in the EV test obtained by each group? *The MD group would obtain lower scores than the ED group.*
- (iii) Did the EV subtest measure the true language abilities of the Singaporean sample? *No, the Singaporean sample's true language abilities would not be accurately reflected in the EV subtest.*
- (iv) Was the Singaporean sample's performance affected by culturally and linguistically biased test items present in the EV test? *The poorer performance by the Singaporean sample would be affected by the presence of culturally and linguistically biased items in the EV test.*

METHODOLOGY

Ethical clearance

Before commencing with data collection, ethical clearance was obtained from the Social and Behavioural Research Ethics Committee, Flinders University and Southern Adelaide Health Service of South Australia. A permission license to use the *CELF Preschool 2 UK* was also obtained from the publisher, NCS Pearson.

Permission for testing

In accordance with Singapore's educational protocol, permission for testing was first sought from each centre's principal. Once agreed, the researcher collaborated with the centres' teachers in selecting children that fulfilled the selection criterion. Parental consent forms were then sent to the parents of selected children. Participation was voluntary, as parents chose to either allow or withdraw their child from taking part in the study. Children who did not wish to participate or were unwilling to continue their participation were excluded from the study.

Participants

A total of 79 Singaporean ethnic Chinese children, made up of 42 males and 37 females, aged 4;6 (4 years:6 months) to 5;5 (5 years:5 months), with no known speech, language and/or sensory difficulties as reported by both parents and teachers, were recruited for the study. The mean age of these children was 58.5 months ($SD=3.31$). As there are a large number of official and non-official languages in Singapore, the researcher selected only ethnic Singaporean Chinese children who were bilingual in English and Mandarin. This was necessary so as to minimise the complexity of the heterogeneous language situation associated with a multilingual population. Based on the Census of Population 2000 (Leow, 2000), 76.8% of Singaporeans belonged to the Chinese ethnic group. Thus, this majority ethnic group was selected. Participants were chosen from government and private preschool centres across different socio-economic areas in Singapore to ensure that a representative sample of the population was obtained.

The Singaporean sample was subsequently divided into two groups based on their dominant language (ED group and MD group). The participants were categorised after detailed feedback on each child's language profile from both parents and teachers was received. Using information

from both parents and teachers provided a stronger and more detailed language background check, as compared to using only one informant. Feedback from parents was collected via a Language Dominance Questionnaire (LDQ) that was distributed with the parental consent form. The questions in LDQ were adapted and modified from Tan's (2008) Language Background Questionnaire. In the LDQ, parents were asked to detail information on their child's main communication partners, languages used by each main communication partner, child's proficiency in each language and child's preferred choice of language in different communication situations. This information, along with the teacher's verbal feedback on the child's dominant language, was matched to determine the child's dominant language. Out of the 79 pairs of feedback received, 68 pairs matched, and this related to a high percentage of matching feedback (86%). For the remaining 11 pairs of unmatched feedback, the answers from the LDQ were further analysed and in most cases, parent's feedback on child's dominant language was favoured as they were able to accurately detail information about their child's language profile. In this sample, 38 children had English as their dominant language while the other 41 children had Mandarin as their dominant language; this division is comparable to the first language of Singaporean Chinese children provided by Singapore Census Population 2005 (Leow, 2006).

Validity

One aim of this study was to determine if each participant's performance on the *EV* subtest was representative of his or her true language ability hence the criterion-related validity of the *EV* subtest had to be determined. Kennedy (2002) defined criterion-related validity as 'the extent to which performance on a test is correlated with performance on another instrument'. In this case, it was most ideal to select the *Singapore English Action Picture Test (SEAPT)* (Brebner, 2002) as a criterion-related validity tool. The *SEAPT* is the only locally adapted screening tool that

measures expressive language in the form of information (i.e. expressive vocabulary for nouns and verbs) and grammar (i.e. use of morphological markers, sentence construction etc.). As both the information scores of the *SEAPT* and scores of the *EV* subtest of the *CELF Preschool 2 UK* measured expressive vocabulary, both sets of scores from each participant could be ideally compared to each other. In addition, the *SEAPT* has local normative data pertaining to each of the main language groups as classified in this study.

Procedure

The administration process was followed in accordance to the instructions outlined in each respective test manual. Each participant was tested individually in a quiet room in his or her attending preschool centre. For all participants, the *EV* subtest of the *CELF Preschool 2 UK* was first administered followed by the procedure outlined in the *SEAPT* manual. Responses from all participants were transcribed on official recording sheets. In addition, the responses were also tape-recorded for a second round of transcription. All responses from both tests were then scored according to the instructions outlined in their respective test manuals.

Inter-rater reliability

All tests were conducted by the Singaporean researcher, thus it was essential to determine the accuracy of the transcription and scoring procedure. As such, another Singaporean with a similar amount of experience as the researcher was recruited as a rater. She transcribed and scored 10% of the participants' audio recordings on both tests. Results of 2 Pearson correlations showed that the inter-rater reliability coefficient was high for the *EV* subtest, ($r = +1.00, p < 0.001$) and for the *SEAPT*, ($r = 0.94, p < 0.001$).

Item Analysis

To determine if the test items in the UK designed *EV* subtest were culturally and linguistically appropriate for the Singaporean population, a quantitative and qualitative analysis of the test items in the *EV* subtest was undertaken. The most popular method of evaluation would be to run the Differential Item Functioning (DIF) procedure (Anastasi & Urbina, 1997). However, this level of analysis was beyond the scope of this study. Instead, preliminary analyses based on basic percentage statistics; comparison with local literature (Gupta, 1994; Brebner et al., 2001) and the researcher's judgment were conducted. Each picture stimulus and expected target responses of each test item on the *EV* subtest was analysed to determine if it was culturally and linguistically appropriate for the Singaporean population.

RESULTS

In this study, scaled scores and percentile ranks were selected for comparison between groups in the analysis of the collected data. Given the normal distribution of the data collected, parametric analysis was carried out using the SPSS for Windows version 17.0 (Arbuckle, 2008). Means and standard deviations were used to describe a group's performance. A 1-sample *t* test was carried out to compare the scaled scores obtained by a Singaporean group and the UK standardisation sample while a 2-sample *t* test was carried out when comparison between the scaled scores of two Singaporean groups (MD group and ED group) was made. Lastly, a Pearson correlation was used to obtain bivariate correlations for inter-rater reliability and validity coefficients.

The overall performance of the Singaporean sample

Figure 1 shows the performance of the Singaporean sample on the *EV* subtest based on published UK normative data. Overall, the Singaporean sample obtained a mean scaled score of 6.96 (SD =

3.34) with scaled scores ranging from 1 to 15. This result placed their mean scaled score at the 16th percentile, when compared to the published UK percentile rankings.

To determine if the Singaporean sample performed significantly differently from the UK standardisation sample (Mean=10.0, SD=3.00), a 1-sample *t*-test analysis was conducted. Results showed that the Singaporean sample achieved significantly lower scores than the UK standardisation sample, $t_{(78)}=-8.80$, $p<0.001$. This result was consistent with the hypothesis that the Singaporean sample would perform poorly when compared to the UK standardisation sample.

Singaporean sample: English Dominant (ED) group versus Mandarin Dominant (MD) group

The performances of the ED group and MD group, based on published UK normative data, can be seen from figures 2 and 3 respectively. The ED group obtained a mean score of 8.92 (SD=2.80) with scaled scores ranging for 2 to 15. Whereas the MD group obtained a mean score of 5.15 (SD=2.75) with scaled scores ranging from 1 to 13. Based on the published UK percentile rankings, these results placed the mean scaled score of the ED and the MD group at the 37th and 5th percentile respectively.

To determine if both the ED group and MD group scored significantly differently, a 2 sample *t*-test analysis was conducted to compare the ED group's performance to the MD group's performance. Results showed that the ED group achieved significantly higher scores than their MD counterparts, $t_{(77)}=6.06$, $p<0.001$. This result was consistent with the hypothesis that the MD group would obtain lower scores compared to the ED group.

To determine how each language dominant group performed in comparison to the UK standardisation sample, the mean scaled scores from the ED (M=8.92, SD=2.80) and MD group (M=5.15, SD=2.75) were compared to the UK standardisation sample (M=10.0, SD=3.0)

individually. A 1-sample *t*-test analysis showed that the ED group achieved slightly lower scores at a significant level when compared to the UK standardisation sample, $t_{(37)} = -2.37, p < 0.05$. In contrast, a 1 sample *t*-test showed that the MD group achieved much lower scores at a significant level when compared to the UK standardisation sample, $t_{(40)} = -11.360, p < 0.001$.

Validity and measurement of true language abilities of the Singaporean sample

The *EV* scores and *SEAPT* information scores obtained by the Singaporean sample were correlated to determine if the Singaporean sample's performance in the *EV* subtest was representative of their true language abilities. An overall moderate positive validity coefficient ($r = +0.616, p < 0.001$) was obtained for the Singaporean sample. In individual groups, the measured validity coefficients of ED and MD group were at $r = +0.414, (p < 0.001)$ and $r = +0.514, (p < 0.001)$ respectively. The moderate validity coefficient values obtained demonstrated that the true language abilities of the Singaporean sample were not accurately reflected. Additionally, the percentage of children identified at risk of language difficulties was also determined. The *EV* subtest identified 47% of the Singaporean sample as being at risk of language difficulties (performing 1 standard deviation below the UK mean scaled score) while the *SEAPT* only identified 23% of the Singaporean sample as being at risk of language difficulties (cut off for recommendation of further assessment set at 20th percentile).

Item Analysis – Identification of culturally and linguistically biased test items

One of the aims in the study was to verify whether the *EV* subtest items were culturally and linguistically appropriate for Singaporean preschool children. It was therefore necessary to calculate the percentage of children who gave correct responses (could be either 2 points or 1 point based on the scoring system outlined in the test manual) for each test item, as seen in Table

2, before further analysis of each test item could commence. Although the test items in the *EV* subtest were ranked in order of difficulty, it did not have the same order of effect for the Singaporean sample. The Singaporean sample had more difficulties with earlier test items (e.g. test item number 5, 6, 8: *fireman, wrapping, footprint*) with less than 50% of them obtaining scores as compared to test items that appeared later in the list (e.g. test item number 7, 9: *newspaper, pouring*) with more than 65% of them obtaining scores.

Based on evidence from local literature (Gupta, 1994; Brebner et al., 2001) and the researcher's local knowledge and experience, each test item is classified as appropriate, culturally biased or linguistically biased. Culturally biased test items such as *fireman, telescope, vet* and *binoculars* were identified. In addition, linguistically biased items such as *riding, wrapping* and *pouring* were also identified. A summary of potential culturally and linguistically biased test items is presented in Table 2 as well.

DISCUSSION

The overall performance of the Singaporean sample

Results from statistical analysis showed that as a group, the Singaporean sample obtained a significantly lower mean scaled score on the *EV* subtest as compared to the UK standardisation sample. Using the published UK percentile rankings, the mean scaled score of 7 (≈ 6.96) by the Singaporean sample placed them at the 16th percentile as compared to their UK counterparts whose mean scaled score of 10 placed them at the 50th percentile. This poor performance was consistent with the results from other bilingual studies. Hemsley et al. (2010) found that when comparing only a single language score, English (single language), bilingual children demonstrated significantly lower vocabulary scores in comparison with their monolingual

counterparts. Such a phenomenon can be explained by the Dual System Model (Genesee, 1989). As a result of this division in language abilities, the total knowledge of a single language for bilingual children would not be comparable to that of monolingual children. This was also supported by Bedore et al. (2005) who found that bilingual children often have vocabulary knowledge that is unique to each language. However when conceptual scoring is taken into consideration, Hemsley et al. (2010) and Bedore et al. (2005) found that the composite score obtained by bilingual children would often be comparable to that of their monolingual counterparts. Thus, many studies (e.g. Bedore et al., 2005; Hemsley et al., 2010; Junker & Stockman, 2002; Pearson et al., 1993) have emphasised that a bilingual child's language ability can only be deduced when his or her performances in two languages (i.e. conceptual scoring) are taken into account.

In this study, the Singaporean English-Mandarin bilingual children only had their language abilities measured in a single language, English. Thus, the lower scaled scores obtained by the Singaporean sample in the *EV* subtest were not reflective of their potential language abilities. The single-language measure ignored the fact that bilingual children might choose to use different words from different languages depending on the setting, interlocutor and context (Iglesias, 2002). For example, in this study, some of the participants provided the Mandarin equivalent of the targeted vocabulary names when they did not know them in English. As such, these children were placed at a scoring disadvantage. Although some children had readily provided translation equivalents, they could not be scored as the researchers did not actively seek translation equivalents from all participants. In addition, this procedure was beyond the scope of this research project.

The difference in performance between the English Dominant (ED) group and Mandarin

Dominant (MD) group

Data collected from the Singaporean sample were split into two groups based on language dominance: ED group and MD group. Results from the statistical analysis indicated that the mean scaled score obtained by the ED group (M=8.92, SD=2.80) was slightly lower (still at a significant level) than the UK standardisation group (M=10.0, SD=3.0). However the MD group (M=5.15, SD=2.75) obtained a significantly lower mean scaled score when compared to the UK standardisation sample. Further statistical comparison between the ED group and MD group found that their performance also differed significantly from each. Thus, it could be concluded that the MD group performed poorly in comparison to the ED group.

The ED group performed better than their MD counterparts and this may be for reasons stated in earlier studies (Patterson, 2002; Uchikoshi, 2006) of language input among bilingual children which indicated that vocabulary size and development in the English language are closely related to the amount of exposure a bilingual child has with the English language. In addition to that, local studies in Singapore have also shown that there was a strong correlation between a local child's oral competence in one language and the usage of that particular language at home (Saravanan, 2001; Saravanan, 2004; Zhao, Liu & Hong, 2007). In simpler terms, if a child's dominant language was identified as English, he or she would perform better in an expressive English vocabulary test, as compared to another child whose dominant language was Mandarin. As a result, the ED group's performance was closer to the UK standardisation sample, as both groups spoke mainly English, in comparison to the MD group who obtained scaled scores that fell within the lower range of the scale. These findings further highlighted the fact that Singaporean English-Mandarin bilingual children should not be grouped collectively as a

homogenous group and that it was crucial to consider the effects of language dominance on assessment outcomes.

Validity: Did the EV subtest measure the true language abilities of the Singaporean sample?

The analyses of the results have shown that the Singaporean sample did indeed perform much more poorly than their UK counterparts. However, the question was whether this difference in performance could be attributed to the fact that the Singaporean English-Mandarin bilingual children's language abilities were only measured in one language and not both; or that the presence of culturally and linguistically biased test items in the UK designed *EV* subtest also played a part.

The scores from the *EV* subtest were correlated with their information scores on the *SEAPT* to determine the direction and magnitude of the criterion-related validity coefficient. A moderately positive validity coefficient ($r = +0.616$, $p < 0.001$) was obtained statistically. The positive correlation obtained was expected as the scores from both tests measured expressive vocabulary. However, the validity of the coefficient was only found to be moderate. Even when the validity coefficient was analysed separately, moderate positive validity coefficients were still obtained for both the ED ($r = +0.414$, $p < 0.001$) and MD group ($r = +0.514$, $p < 0.001$). In most psychological texts (e.g. Anastasi & Urbina, 1997), the strength of a moderate coefficient would not be considered as sufficient evidence to support that the *EV* subtest accurately measured what it was purported to measure for the Singaporean sample. In other words, the performance of the Singaporean sample on the *EV* subtest was not reflective of their true language abilities. A main factor that must have contributed to this result was the presence of culturally and linguistically biased items in the UK designed *EV* subtest. Unlike the *EV* subtest, the *SEAPT* is a locally-

adapted language screening tool hence the presence of culturally and linguistically biased test items is minimised, or even absent.

In this study, the *EV* subtest had identified 47% of the Singaporean sample as being at risk for language difficulties, while the *SEAPT* identified 23%. Thus 24% of the Singaporean sample was potentially misdiagnosed on the basis of the *EV* results. In other words, the occurrences of false positive diagnoses were high when the UK designed *EV* subtest was used to measure the language abilities of Singaporean children. Bedore et al (2005) and Westman et al (2008) also found similar results and have recognised that bilingual children are often over-represented and classified at risk of language difficulties instead of typical developing when assessed and compared to monolingual standardised norms. Thus in this study, the *EV* subtest may not have measured the true language abilities of the Singaporean sample.

The presence of culturally and linguistically test biased items

Results and analyses of this study have provided further evidence that, in order to diagnose language impairment through comparing a child's performance with that of their peers, a normative language assessment should only be used with the population that it was designed for. Thus, quantitative and qualitative analysis of the *EV* subtest test items was conducted to single out culturally and linguistically biased test items for the Singaporean population.

Firstly, the difficulty of each test item had to be determined before the identification of culturally and linguistically biased test items could commence. The literature defined the difficulty of a test item in terms of the percentage of people who answered it correctly (Anastasi & Urbina, 1997). As the difficulty of a test item increases, less people are expected to answer it correctly. It is customary for tests to arrange the test items in order of difficulty, so that the test would begin

with relatively easier items before proceeding to harder ones. Likewise, the test items in the *EV* subtest were ranked based on estimates of relative difficulty (Wigg et al., 2006), so that more children could score on earlier items in the *EV* subtest than the later items. However, this was not the case for the Singaporean sample (see Table 2).

The Singaporean sample had more difficulties with some of the earlier test items (*test item numbers 5, 6, 8: fireman, wrapping, footprint*) on the list compared to those that had appeared later in the list (*test item numbers 7, 15: newspaper, calculator*). Hence, the arranged level of difficulty in the *EV* subtest might not have the same order effect for the Singaporean population. In Singapore, most children would have encountered newspapers and calculators in their daily lives as these items can be commonly found in Singaporean homes. On the other hand, *test item number 8: footprint* would not be recognizable to most Singaporean children as many playgrounds in Singapore do not have sandpits. Therefore, Singaporean children would not be familiar with footprints left on the sand.

Further analysis showed that some of the pictures used in the *EV* subtest were also unfamiliar to the Singaporean population. Consequently, the presence of these pictorial culturally biased items resulted in lower scores obtained by the Singaporean sample. For example, *test item number 5 fireman* in the *EV* subtest had a fireman dressed in a yellow uniform with a yellow hat. However, in Singapore, firemen are dressed in blue uniforms and red hats. Thus, the UK's pictorial representation of the target word *fireman* was not culturally recognisable for many of the Singaporean sample, with only 48% of them identifying and naming the picture correctly. Furthermore, there were some pictorial stimuli such as *telescope, vet and binoculars* that had less than 30% of the Singaporean sample identifying and naming them correctly as they were culturally unfamiliar.

In addition, as the target linguistic responses for the *EV* subtest were based only on Standard English, SCE-type of linguistic responses given by the Singaporean sample could not be accepted. As a result of the use of SCE and the influence of other languages, Brebner (2009) reported that Singaporean children only start marking verb tenses regularly from 6 years of age. For the following *test items number 3, 6, 9: riding, wrapping, pouring*, many in the Singaporean sample omitted the present progressive ‘-ing’ verb ending of the target linguistic response and were subsequently penalised although such an omission is typical for Singaporean children. For example, 25% of the Singaporean sample named *item number 9* as *pour* instead of *pouring*, thus 25% of them were given only 1 point instead of 2 full points.

CONCLUSION

Summary

This study aimed to investigate the performance of Singaporean preschoolers on the *EV* subtest *CELF Preschool 2 UK*. In addition, the test items were analysed to determine if the presence of culturally and linguistically biased test items affected the performance of these children. Results showed that the Singaporean sample obtained lower scores on the *EV* subtest as compared to their UK counterparts. Further analysis of the data also showed that within the Singaporean sample, the MD children performed more poorly in comparison to their ED peers. As such, the language background of Singaporean children should be considered carefully as language dominance was found to have significantly affected the outcomes of the assessment. Two plausible reasons for the poorer performance of Singaporean children on the *EV* subtest were: 1) elicitation of only a single language measure in English which ignored existing translation equivalents in Mandarin; and/or 2) the presence of culturally and linguistically biased test items.

Clinical implications

The results found in this study have several implications for clinical practice internationally. To qualify for special education or allied health service support, commercially available language assessments tool are often selected to assess a child's language abilities. The CALD child's performance is often compared the normative data for the monolingual Standard English speaking population and this may lead to misrepresentation of the CALD child's true language abilities. The use of standard scores may lead to an over-representation of these children in special education. This may reduce the diagnostic sensitivity and specificity of clinical disorders (language difference versus language disorder) among the CALD population.

As such, this study recommends that commercially available language assessments that were not originally designed for use with CALD populations should be checked for the presence of biased test items and modified appropriately before use with respective CALD populations. As the modification and re-norming of a commercially language assessment tool can be time and resource consuming, speech-language pathologists working with CALD population should also use non-standard measures of assessment (e.g. dynamic testing, language checklist) to obtain a more representative sample of the language ability of a CALD child. The use of these language assessment tools may still be able to provide useful information regarding a child's language skills if it used as a criterion-referenced measure.

In addition, the findings in this study have shown that due to the prevalent usage of SCE with a mix of different languages, the Singaporean population cannot be grouped as a homogenous community. Even within the same ethnic group, language differences still exist. Literature (McGregor, Williams, Hearst & Johnson, 1997) has recommended speech-language pathologists

to adopt contrastive analysis when assessing non Standard English speakers. The use of contrastive analysis will help determine whether the expressive language output patterns of a Singaporean child is consistent with Singaporean non-standard English speakers when the use of SCE and the influence of dominant language are taken into consideration; or if the expressive output patterns represent true language errors. This will help in obtaining a more accurate and full language profile of Singaporean children, reducing the occurrence of misdiagnosis. This study also encourages the gathering of reliable background information on the languages used by the child with his or her main communication partners and the languages' developmental milestone, as findings in this study have highlighted that language dominance has an effect on the outcome of the assessment.

Limitations of study and future directions

In this study, only Chinese ethnic Singaporean children were recruited. Although the majority of the Singaporean population is made up of the Chinese ethnic group, there are other ethnic groups such as the Malays, Indians and Eurasians which make up 23.2% of the remaining population (Leow, 2006). As such, the conclusive results obtained in this study cannot be said to be the same for the other ethnic groups in the population. The data were only collected from a narrow age band of Singaporean English-Mandarin bilingual children aged between 4;6 and 5;5 months. In addition, the *EV* subtest is only one out of the eight subtests used in the *CELF Preschool 2 UK*. Thus, future research should aim to explore the performance of different populations of bilingual learners (e.g. other ethnic groups and age groups) on the different subtests of the *CELF Preschool 2 UK*.

Another limitation to this study was that stratified sampling was not used. Dixon (2011) found that children from higher socio-economic status (SES) families in Singapore have a larger vocabulary size in comparison to children from lower SES families. The SES of each participant was only assumed and not collected from the participants' families hence the SES status of the participants may not be a representative sample of the Singapore population. Future research should factor SES as a controlled variable.

Future research could investigate if the scores obtained by the Singaporean children would be comparable to the UK published normative data if translation equivalents of test items in their mother tongue were accept. This question should be explored as this study found that when some participants were unable to provide the test item names in English, translation equivalents in Mandarin were given instead. In addition, bilingual studies (Hemsley et al., 2010 & Bedore et al., 2005) have also shown that when conceptual scoring was scored, bilingual children's performance would often be comparable to that of monolingual children.

Lastly, future research could also investigate if the performance of Singaporean children could improve on a locally adapted version of the *CELF Preschool 2*. The performance of Singaporean children on both the original UK version and the locally adapted version could also be compared to determine if adaptation is worthwhile. At the same time, the reliability and validity of the locally adapted *CELF Preschool 2* should also be established. Ideally, local normative data based on the locally adapted version of *CELF Preschool 2* should be collected. This will add to the growing set of clinically useful and sensitive data to be used in Singaporean speech-language pathology clinics.

ACKNOWLEDGEMENT

The authors wish to thank the school principals, teachers, parents and children for their enthusiastic cooperation in this study.

REFERENCES

- Anastasi, A., & Urbina, S. (1997). *Psychological testing*. New Jersey: Prentice Hall.
- Arbuckle, J. L. (2008). *SPSS statistics gradpack 17.0*. Australia: SPSS Inc.
- Bedore, L. M., Peña, E. D., García, M., & Cortez, C. (2005). Conceptual versus monolingual scoring: when does it make a difference? *Language, Speech, and Hearing Services in Schools, 36*(3), 188 - 200.
- Brebner, C. (2002). *The Singapore English Action Picture Test*. Published with permission of Speechmark Publishing Ltd. Available through author, contact chrisandphidugid@yahoo.com.
- Brebner, C. (2009). *Normal Language Development in English for English-Mandarin Bilingual Preschool Children In Singapore: Implications For Clinicians*. Presentation for Speech-Language and Hearing Association (Singapore). Singapore: National University of Singapore.
- Brebner, C. (2010). The development of the Singapore English Action Picture Test: An expressive language screening assessment for Chinese Singaporean preschool children. In M. Cruz-Ferreira (Ed.), *Multilingualism, language norms and multilingual contexts*. (pp. 323 – 342) Bern: Peter Lang Publishing Group.
- Brebner, C., Rickard-Liow, S., & McCormack, P. (2000). Assessment of children's language skills: Challenges facing speech pathologists in Singapore. In C. Lind (Ed.) *Speech Pathology Australia Conference 2000 Proceedings* (pp. 181-187). Adelaide: Speech Pathology Australia.
- Brebner, C., Rickard-Liow, S., & McCormack, P. (2001) The cultural and linguistic modification of The Renfrew Action Picture Test for use in Singapore. In S. Hewatt & L. Wilson (Eds.), *Speech Pathology Australia Conference 2001 Proceedings* (pp. 155-162).

Melbourne: Speech Pathology Australia. Copyright 2001 Speech Pathology Australia.

Adapted with permission.

Carter, J. A., Lees, J. A., Murira, G. M., Gona, J., Neville, B. G. R., & Newton, C. R. (2005).

Issues in the development of cross-cultural assessments of speech and language for children. *International Journal of Language and Communication Disorders*, 40(4), 385 - 401

Crutchley, A., Conti-Ramsden, G., & Botting, N. (1997). Bilingual children with SLI and standardised assessments. *International Journal of Bilingualism*, 1, 117 -134.

Dixon, L. Q. (2011). The role of home and school factors in predicting English vocabulary among bilingual kindergarten children in Singapore. *Applied Psycholinguistics*, 32, 141-168.

Dollaghan, C. A., & Horner, E. A. (2011). Bilingual language assessment: A meta-analysis of diagnostic accuracy. *Journal of Speech, Language, and Hearing Research*, 54(4), 1077-1088.

Genesee, F. (1989). Early bilingual development: One language or two? *Journal of Child Language*, 16, 161 - 180.

Goldstein, B. (2006). Clinical implications of research on language development and disorders in bilingual children. *Topics in Language Disorders*, 26(4), 305 - 321.

Gupta, A. F. (1994). *The step tongue: Children's English in Singapore*. Clevedon: Multilingual Matters.

Gutiérrez-Clellen, V. F., & Peña, E. (2001). Dynamic assessment of diverse children: A tutorial. *Language, Speech, and Hearing Services in School*, 22, 212 - 224.

Hemsley, G., Holm, A., & Dodd, B. (2010). Patterns in diversity: Lexical learning in Samoan-English bilingual children. *International Journal of Speech-Language Pathology*, 12(4),

362-374.

Iglesias, A. (2001). What test should I use? *Seminars in Speech and Language*, 22(1), 3 -16.

Isaac, K. (2002). *Speech pathology in cultural and linguistic diversity*. London: Whurr Publishers Ltd.

Junker, D. A., & Stockman, I. J. (2002). Expressive vocabulary of German-English bilingual toddlers. *American Journal of Speech-Language Pathology*, 11(4), 381 - 394.

Kohnert, K. (2010). Bilingual children with primary language impairment: Issues, evidence and implications for clinical actions. *Journal of Communication Disorders*, 43(6), 456-473.

Law, N. C. W., & So, L. K. H. (2006). The relationship of phonological development and language dominance in bilingual Cantonese-Putonghua children. *International Journal of Bilingualism*, 10(4), 405 - 427.

Leow, B. G. (2000). *Census of Population 2000: Advanced Release*. Singapore Department of Statistics. <http://www.singstat.gov.sg/pubn/popn/c2000adr.html> accessed 5th June, 2008.

Leow, B. G. (2006). *General Household Survey 2005: Statistical Release 1*. Singapore Department of Statistics. <http://www.singstat.gov.sg/pubn/popn/ghsr1.html> accessed 5th June, 2008

Li, W., Miller, N., Dodd, B., & Zhu, H. (2005). Childhood bilingualism: Distinguishing difference from disorder. In M. J. Ball (Ed.), *Clinical Sociolinguistics*. (pp. 193-206) Malden, MA: Blackwell Publishing Ltd.

McGregor, K. K., Williams, D., Hearst, S., & Johnson, A. C. (1997). The use of contrastive analysis in distinguishing difference from disorder: A tutorial. *American Journal of Speech-Language Pathology*, 6(3), 45-56

Nicoladis, E., & Genesee, F. (1997). Language development in preschool bilingual children.

Journal of Speech-Language Pathology and Audiology, 21(4), 258 - 270.

Paradis, J., Crago, M., Genesee, F., & Rice, M. (2003). French-English bilingual children with SLI: How do they compare with their monolingual peers? *Journal of Speech, Language, and Hearing Research*, 46, 113 - 128.

Patterson, J. L. (2002). Relationships of expressive vocabulary to frequency of reading and television experience among bilingual toddlers. *Applied Psycholinguistics*, 23(4), 493 - 508.

Pearson, B. Z., Fernández, S. C., & Oller, D. K. (1993). Lexical development in bilingual infants and toddlers: Comparison to monolingual norms. *Language Learning*, 43(1), 93 -120.

Peña, E. D., Bedore, L. M., & Zlatic-Giunta, R. (2002). Development of categorization in young bilingual children. *Journal of Speech, Language, and Hearing Research*, 45, 938 - 947.

Pickering, M., & McAllister, L. (2000). A conceptual framework for linking and guiding domestic cross-cultural and international practice in speech-language pathology. *International Journal of Speech-Language Pathology* 2(2), 93-106.

Qi, C. H. Q., Kaiser, A. P., Milan, S., & Hancock, T. B. (2006). Language performance of low-income African American and European American preschool children on the PPVT-III. *Language, Speech, and Hearing Services in Schools*, 37(1), 5 - 16.

Saravanan, V. (2001). The significance of bilingual Chinese, Malay and Tamil children's English network patterns on community language usage patterns. *Early Child Development and Care*, 166, 81 - 91.

Saravanan, V. (2004). Bilingual children's languages proficiency and language choice patterns. *Journal of Australian Research in Early Childhood Education*, 11(1), 13 - 25.

Singapore Ministry of Education (2008). *Education in Singapore*. Ministry of Education, Singapore. <http://www.moe.gov.sg/about/files/moe-corporate-brochure.pdf> accessed 29th

June, 2008.

Tan, S. H. (2008). *Language Background Questionnaire IV*. Unpublished Questionnaire. National University of Singapore.

Teo, C. H. (2000). *Global Literacy: The Advantage of Speaking Good English*. Ministry of Education, Singapore. <http://www.moe.gov.sg/media/speeches/2000/sp31032000a.html> accessed 24th August, 2009.

Uchikoshi, Y. (2006). English vocabulary development in bilingual kindergarteners: What are the best predictors? *Language and Cognition*, 9(1), 33 - 49.

Vaughan - Cooke, F. B. (1983). Improving language assessment in minority children. *ASHA*, 25(9), 29 - 34.

Westman, M., Korkman, M., Mickos, A., & Byring, R. (2008). Language profiles of monolingual and bilingual Finnish preschool children at risk for language impairment. *International Journal of Language and Communication Disorders*, 43(6), 699 - 711.

Wiig, E. H., Secord, W., & Semel, E. (2006). *Clinical Evaluation of Language Fundamentals - Preschool Second UK Edition*. London: Harcourt Assessment

Zhao, S. H., Liu, Y. B., & Hong, H. Q. (2007). Singaporean preschoolers' oral competence in Mandarin. *Language Policy*, 6(3), 73 - 94.

Table 1: Morphology and syntax features of Singapore Colloquial English

	Features	Examples
Morphology and Syntax	<ul style="list-style-type: none"> variable noun-plural marking 	<ul style="list-style-type: none"> “two cat_”
	<ul style="list-style-type: none"> variable past tense marking 	<ul style="list-style-type: none"> “go out” instead of “I <u>went</u> out”
	<ul style="list-style-type: none"> pronoun deletion when not required by the context 	<ul style="list-style-type: none"> “_ go shopping” rather than “I go shopping”.
	<ul style="list-style-type: none"> conditional clauses used without subordinating clause when not required by the context 	<ul style="list-style-type: none"> “if / when” deleted, such as “do that, get punish”
	<ul style="list-style-type: none"> the verb “to be” deleted when not required by the context 	
	<ul style="list-style-type: none"> pragmatic particles used to indicate the speaker’s level of commitment to what they have said 	<ul style="list-style-type: none"> “he __ happy”

Note. From title “The Cultural and Linguistic Modification of The Renfrew Action Picture Test for Use in Singapore,” by Brebner, Rickard-Liow & McCormack, 2001, *Speech Pathology Australia Conference 2001 Proceedings*, p. 156. Copyright 2001 by Speech Pathology Australia. Adapted with permission

Table 2: Analysis of each Expressive Vocabulary test item

Item number	Target response	Percentage of Singapore children who obtained 1/2 points	Culturally biased	Linguistically biased
1	Carrot	99%	No	No
2	Flag	86%	No	No
3	Riding	58%	No	Yes
4	Piano	68%	No	No
5	Fireman/Fire fighter	46%	Yes	No
6	Wrapping	38%	No	Yes
7	Newspaper	80%	No	No
8	Footprint	37%	Yes	No
9	Pouring	69%	No	Yes
10	Branch	22%	No	No
11	Telescope	28%	Yes	No
12	Wheelchair	20%	No	No
13	Stamp	14%	No	No
14	Binoculars	15%	Yes	No
15	Calculator	42%	No	No
16	Veterinarian/Vet	10%	Yes	Yes
17	*Calendar	23%	No	No
18	*Trophy	15%	No	No
19	*Audience	1%	No	Yes
20	*Scales	4%	No	Yes

*Vocabulary names that are not yet age-appropriate for children between the ages of 4; 6 – 5; 5

Figure Captions

Figure 1: Overall performance of the Singaporean sample on the Expressive Vocabulary subtest

Figure 2: Performance of the English Dominant group on the Expressive Vocabulary subtest

Figure 3: Performance of the Mandarin Dominant group on the Expressive Vocabulary subtest