Running title: Differential effects of internal and external factors

Full title: Differential effects of internal and external factors in early bilingual vocabulary learning: the case of Singapore

He Sun

Nanyang Technological University (NTU), Singapore

Bin Yin

Nanyang Technological University (NTU), Singapore

Nur Farina Begum Binte Amsah

National University of Singapore (NUS), Singapore

Beth Ann O'Brien

Nanyang Technological University (NTU), Singapore

Acknowledgements

The authors are indebted to Rasmus Steinkrauss for his comments on an earlier version of this manuscript. The authors would also like to thank research assistants of Singapore Kindergarten Impact Project for collecting data.

Address for correspondence:

He Sun, the Education and Cognitive Development Lab, National Institute of Education, Nanyang Technological University, NIE5-B3-13, 1 Nanyang Walk, Singapore 637616, Singapore, *he.sun@nie.edu.sg*

Abstract

Both internal factors (e.g. nonverbal intelligence) and external factors (e.g. input quantity) are claimed to affect the rate of children's vocabulary development. However, it remains an open question whether these variables work similarly on bilingual children's dual language learning. The current paper examined this issue on 805 Singapore children (4;1 - 5;8 year-old) who are learning English (societal language) and an ethnic language (Mandarin / Malay / Tamil). Singapore is a bilingual society, however, there is an inclination for English use at home in recent years, resulting in a discrepancy of input between English and ethnic languages in many families. In this study, internal and external factors were examined comprehensively with standardized tests and a parental questionnaire. Regression analysis was used to address the questions.

There were statistically significant differences in language input quantity, quality and output between English and ethnic language learning environments. Singapore children are learning English in an input-rich setting while learning their ethnic language in a comparatively input-poor setting. Multiple regressions revealed that while both sets of factors explained lexical knowledge in each language, the relative contribution is different for English and ethnic languages: internal factors explained more variance in English language vocabulary, whereas external factors were more important in explaining ethnic language knowledge. We attribute this difference to a threshold effect of external factors based on the critical mass hypothesis and call for special attention to learning context (input-rich vs. input-poor settings) for specific bilingual language studies.

Keywords: internal variables, external variables, input-rich setting, input-poor setting, bilingual receptive vocabulary knowledge

1. Background

The past decade witnessed a steadily growing interest in child bilingual language learners (Paradis, 2007) and increasing attention has been paid to the internal and external factors that would affect the rate and route of children's linguistic development (Unsworth, Hulk & Marinis, 2011). Internal factors include properties more inherent to the learner, such as biological and cognitive capacities, whereas external factors involve aspects of the learner's environment, such as input quantity and quality. Understanding the impact of these factors is assumed to be crucial in promoting the development of existing theories (e.g., Usage Based Theory) and the formulation of educational policies and advice (e.g., language instructional hours at school; Unsworth, Persson, Prins & de Bot, 2015). Both internal and external factors have been found to significantly predict children's bilingual development (e.g., Paradis, 2011; Unsworth, 2016), however, the relationship between internal and external factors is not fully understood, and few papers have touched upon the issue explicitly (but see Paradis, 2011 and Sunderman & Kroll, 2009). Our understanding of the interdependence between internal and external factors in language acquisition would benefit from further exploration, as it will provide us "a unique window onto the cognitive and experiential underpinnings of language" (Dabrowska, 2016, p. 485). It is important to develop a "consistent research agenda that addresses the interaction of individual cognitive abilities and the differential aspects of learning contexts" (Collentine & Freed, 2004, p. 165) to optimize learning processes and outcomes at an individual level.

The current paper explores this issue with preschool children in Singapore. There are four official state languages (i.e., English, Mandarin Chinese, Malay and Tamil) and three major ethnic groups (i.e., Chinese 74.3%, Malays 13.3% and Indians 9.1% in population; Singapore Department of Statistics, 2016) in this country. Children are expected to acquire both English and their ethnic language (which is denoted publicly as the “mother tongue language” and which we refer to in this paper as MT language); but the two languages have been assigned different functions by the government (Bokhorst-Heng, 1999). English has been promoted as the language of international business, access to science and technology, and the medium of instruction within schools and universities, which in turn has increased its usage within Singaporean families (Pakir, 1999; Dixon, Zhao, Quiroz, & Shin, 2012). As such, children’s exposure to English is rich both outside and inside the home. In contrast, MT languages are mostly relegated to educational contexts, with exposure comparatively limited in amount and variety. The utilitarian attitude and the decreasing proficiency level of parents’ MT languages reduce children’s MT language exposure at home even further (Saravanan, 2001; Dixon et al., 2012). As a consequence, although most children have access to English and their ethnic language at a young age and are assumed to be simultaneous or early sequential bilinguals, there could be a large variation in the learning environment of the two languages: learning English in an input-rich setting (home, school and social domains) while learning their MT language in a comparatively input-poor setting (predominantly restricted to classroom instruction). Such an ethnic language learning environment shares critical features with an instructional learning environment, where 1) the amount of language exposure is limited, 2) teachers have a great impact on children’s learning outcome, 3) children barely use the target language to communicate with their peers, and 4) the target language generally is not spoken outside of the classroom (Muñoz, 2008).

Different learning environments might result in wide variation in external factors, and further influence the impact of internal factors on acquisition of English and MT languages. Research on the acquisition of grammar has shown the critical role of input in the emergence of syntactic rules in children's developing linguistic system. It has been argued, for example, that a threshold amount (or a 'critical mass') of linguistic input works in conjunction with cognitive linguistic principles (such as complexity and salience) to influence the sequence of grammar learning (Gathercole & Hoff, 2007). Similarly, input threshold quantified as the number of exposures is also requisite for the learning of vocabulary items (e.g., Nation, 1990; Meara, 1997). In relation to the current study, then, the different amount of input for English versus MT may translate into the differential effects of internal versus external factors on learning outcome. If relatively limited exposure to the MT language means that input level is below a required "threshold", we would expect students' performance in this language to be subject primarily to external factors, akin to what has been found in instructional settings (e.g., learning English in China as in Sun, Steinkrauss, Tendeiro, & de Bot, 2016). On the other hand, if the relatively ample input to English means that input level has exceeded the requisite "threshold", then we would expect students' performance to be subject less to external, and more to internal factors, similar to what was found in input-rich learning settings (e.g., learning English in Canada as in Paradis, 2011). We explore this issue in the present study.

2. Literature Review

Both internal and external factors are considered essential to child early language development (Clark, 2003). As noted above, internal factors are intrinsic to the learner, including biological characteristics (age, gender) and cognitive abilities (e.g., working memory and intelligence). External factors are those variables existing in the learner’s environment such as the nature and amount of linguistic input, school instruction, and socio-economic situation of the home. The following sections introduce significant internal and external factors on child bilingual vocabulary development found by previous studies (e.g., Unsworth, 2013).

2.1 Internal Factors

Age, gender, and language aptitude are frequently investigated in child bilingual vocabulary development. In general, older bilingual children may demonstrate an initial advantage of learning rate over their younger counterparts in terms of receptive and productive vocabulary (Cummins & Swain, 1986; Goldberg, Paradis, & Crago, 2008), probably because of older children's more advanced cognitive development. However, the younger learners may catch up and obtain a better outcome in the long run (Muñoz, 2008).

Gender differences in child early vocabulary development have been observed in monolingual studies (Pivik, Andres & Badger, 2011; Eriksson et al., 2012) and in some on bilingual studies (Uchikoshi, 2006). In terms of monolingual children, Huttenlocher et al. (1991) found that before 20 months, girls outperformed boys in the rate of vocabulary growth. Such an advantage has been shown to persist over time (e.g., Bauer, Goldfield, and Peznik, 2002; Bouchard, Trudeau, Sutton, Bourdeault, & Deneault, 2009; Eriksson et al, 2012; but see Bornstein, Hahn, & Haynes, 2004). In terms of bilingual children, the gender differences are less clear. Uchikoshi (2006) followed 150 Latino English learners over a preschool year and found boys began with and maintained higher receptive and productive English vocabulary than girls. The author attributed such gender differences to exposure history and cultural value. Parents were assumed to emphasize academic language achievement more for boys than girls. However, in some other studies on bilingual preschoolers' vocabulary development (e.g., Allman, 2005), such gender differences were not observed. In the present study, we explore any possible gender effects on receptive vocabulary measures in bilingual samples.

Language aptitude, which refers to learners' specific capability for language learning (Carroll & Sapon, 2002), has been found to influence child bilingual vocabulary learning as well. Aspects of internal abilities such as phonological short-term memory, nonverbal intelligence, and phonological awareness are assumed to underlie language aptitude. These aspects were found to associate with vocabulary outcomes in child bilinguals (e.g., Alexiou, 2009; Paradis, 2011; Knell et al., 2007). In general, phonological awareness enables children to segmentalize the global lexical representation, accommodating the pressure of children's enlarging vocabulary size (Metsala & Walley, 1998; Pi-Yu & Rvachew, 2007); short-term memory helps children retain the novel sequence of the phonological properties of a language, facilitating word articulation and semantic memory (Gathercole & Baddeley, 1989; Paradis, 2011); and nonverbal intelligence may help children impute structures and reorganize patterns, allowing them to approach the linguistic tasks in an analytic way (Hakuta & Diaz, 1985; Daller & Ongun, 2017).

2.2 External Factors

External factors (i.e., environmental factors) have also been found to significantly impact bilingual children's early vocabulary acquisition (Dixon, Wu, & Daraghmeh, 2012). These factors include input quantity, input quality, children's language output, and family social economic status (SES) (Unsworth, 2013).

"Of all the relevant factors that parent or communities have some control over, quantity of input is the largest" (Pearson, 2007, p. 400), and there is large variation among families. Child bilingual language learners usually vary in overall length of exposure (LoE) and current exposure time (Paradis, 2011). Input quantity has been revealed to affect bilingual children's vocabulary acquisition (Unsworth, 2013). One source of input would come from active communication and children's input quantity is estimated by the amount of talk he/she receives from families, friends and so on. Another source of language input would be in the form of media, which unlike typical home language use, is non-interactional. Research has shown that the development of bilingual children's societal dominant language will benefit from TV exposure (e.g., Latino children's English vocabulary development in the United States; Uchikoshi, 2006). However, such benefit to children's ethnic language development remains unclear. For instance, Scheele, Leseman and Mayo (2010) did not find a correlation between watching TV in the ethnic language and vocabulary development in that language, in 3-year old Moroccan-Dutch and Turkish-Dutch children.

Input quality could be operationalized in different ways. Various interlocutors (Rojas et al., 2016), higher maternal language proficiency (Chondrogianni & Marinis, 2011), and rich literacy environment (e.g., book reading; Paradis, 2011) have all been found to facilitate child bilingual's vocabulary development. For instance, Rojas and colleagues followed 224 Spanish-speaking English learners in US kindergartens and found the presence of older siblings at home predicted children's expressive language skills. The rationale is that older siblings might have more experience with the societal language and consequently might bring more societal language to home settings.

In recent years, some studies have also highlighted the role of children's language use, especially children's language output at home, as a significant factor in their bilingual vocabulary development (Bohman et al., 2010). It is taken as the most functional measure of exposure as it provides a general index of children's overall language experience (Gutiérrez-Clellen & Kreiter, 2003; Bohman et al., 2010). Controlling for the role of input, these studies revealed an "independent contribution" of language output to bilingual children's vocabulary outcome (Bohman et al., 2010, p. 339). According to Swain (2007), output has three main functions in language learning: (1) a noticing function, (2) a hypothesis-testing function, and (3) a metalinguistic function. Output may prompt children to consciously recognize their language limits, allowing them to reflect their own and others' language behaviors. The role of output could be also illustrated from the perspective of the input-proficiency-use cycle (Pearson, 2007). Namely, more input exposure leads to better proficiency, which increases children's confidence in using the language; more use (output) in turn encourages more input, which starts the cycle again. Pearson (2007) argued that this cycle would be more visible in bilingual children's minority language acquisition.

Family socio-economic status (SES), usually estimated from maternal educational levels and family income, has been found to affect bilingual children's vocabulary size (Goldberg, Paradis & Crago, 2008; Dixon et al., 2012). In Goldberg et al.'s (2008) study, bilingual children whose mother had a higher education level (post-secondary) outperformed their peers whose mothers had a lower level of education (secondary) in their English (societal language) receptive vocabulary. On the other hand, SES might have a different effect on the maintenance of minority languages. Oller and Eilers (2002) found that children from working class families did better than their peers from professional families in oral Spanish skills, despite the fact that professional mothers professed to attach more value to learning Spanish than working class mothers (e.g., Lambert & Taylor, 1996). Similarly, in the Singapore context, Saravanan (2001) found that children from higher SES families tended to have lower proficiency in ethnic languages, due to the preference of the parents to use English at home.

The general picture that emerges from the survey of the literature is that both internal and external factors are relevant to child bilingual's vocabulary acquisition. This is consistent with a usage-based approach to understand language acquisition (Dąbrowska, 2016; Ibbotson, 2013), which makes reference to the interaction among domain-general factors, including "physiology, processing, working memory, pragmatics, social interaction, properties of the input, the learning mechanisms, and so on" (O'Grady, 2008, p. 448). One issue which deserves our attention is that most of the existing studies on internal/external factors and child bilingual vocabulary development address the contribution of individual factors to vocabulary acquisition, and very few of them took internal versus external factors as two contrastive sets of predictors nor explored their interdependence on child vocabulary learning. The exploration of such interdependence is both meaningful and necessary as it would reveal the mechanism of cognition and environment in shaping language, providing specific predictions and illustrations under the realm of usage based theory and addressing the individual differences in language learning (Dąbrowska, 2012, 2016).

2.3 The Relationship between Internal and External Factors

As mentioned above, there are not many studies on the relationship between internal and external factors in child vocabulary acquisition, and those conducted by Paradis (2011) and Sun et al. (2016) are among the few. Paradis' research (2011) on immigrant children in Canada found that internal factors together (e.g., short-term memory and analytical ability) were better predictors of linguistic outcome than external factors. On the other hand, Sun et al.'s (2016) study of English L2 learners in China found that, in an instructional setting, it is the set of external factors (e.g., quantity and quality of input at school and home) that explained more variance in L2 outcome. The opposing findings in the studies by Paradis (2011) and Sun et al. (2016) conducted in their specific learning contexts could be understood, in our view, as the external factors (specifically input) exerting a threshold/bottleneck effect on internal factors. That is to say, internal factors begin to influence learning outcome only after the input reaches a certain threshold level. We find relevant insight from research bearing on the "critical mass hypothesis" (Marchman & Bates, 1994; Elman, 2003; Gathercole & Hoff, 2007) which establishes the primacy of the role of input before related patterns/abilities/factors can emerge. Such a supposition is entirely consistent with the central tenet in usage-based thinking, which prioritizes input and general cognitive factors in explaining language learning. A series of studies have been conducted to examine the critical mass hypothesis. For instance, Marchman & Bates (1994) found that children begin to demonstrate productive knowledge on past-tense morphology only after sufficient linguist input enabled their verb vocabulary size to exceed 50. The assumption was that productive morpho-syntactic knowledge emerges after exposure to sufficient instances of inflected verbs in the input. Gathercole & Hoff (2007) also showed a “see-saw” relationship between input and cognitive linguistic principles, in their conjoint effect on grammar learning. For instance, more complex and less salient properties require a higher level of input for acquisition, compared with less complex and more salient properties whose critical mass requirement is lower. Applied to the issues pertinent to the present study (i.e., the relationship between external and internal factors), such a critical mass perspective might suggest that the relative role of internal versus external factors in language acquisition could depend on the setting of learning (i.e., input-rich setting versus input-poor setting). More specifically, in an input-rich environment, learners are exposed to ample linguistic stimuli that provide the requisite amount and quality of input for internal mechanisms to contribute to language learning (e.g., establishing form-meaning mappings in lexical meanings). We can think of this requisite amount and quality of input as a sort of threshold in order for internal mechanisms to manifest their importance. In input-poor settings (e.g., instructional settings) where such a threshold level of input is not reached, then internal mechanisms remain in the background so to speak. In such a scenario, outcome of language learning then is determined primarily by the quantity and quality of input at the initial period. Failing to encounter this critical mass would result in delayed or incomplete learning of language structures.

The current study intends to explore the specific role that each internal factor (i.e., age, language aptitude, gender) and external factor (i.e., input quantity, input quality, output, SES) plays in child bilingual's English and MT vocabulary learning, from the perspective of usage based theory. More importantly, we would like to explore the interdependence of internal and external factors and address the issue when we should give more weight to internal factors and when we should not, shedding light on individual differences within the realm of usage based theory. From our point of view, such an interdependence may be backed up by the critical mass hypothesis in terms of input threshold. Different from the existing literature, we examined both scenarios - one in which internal mechanisms can be expected to play a major role; and the other in which external factors such as input can be expected to play a major role - within the same individual by investigating vocabulary acquisition of bilingual children's dual languages. Specifically, due to English and MT's different learning environments, we expect opposing effects of internal versus external factors on each language. Following Paradis (2011), we expect that learning English for our bilingual children is more affected by internal factors given the abundance and quality of English language input in the school and home environments. Following Sun et al (2016), we expect that learning MT languages (Mandarin, Malay, or Tamil) is more affected by external factors, given the relative deficiency of input for MT in many families.

2.4 Questions and Hypotheses

Question 1: Which individual internal and external factors contribute to bilingual children's vocabulary knowledge in English and mother tongue languages respectively?

Hypothesis 1: Based on findings in input-rich (e.g., Paradis, 2011) and input-poor settings (e.g., Sun et al., 2016), components of both internal factors (i.e., age, language aptitude, gender) and external factors (i.e., input quantity, input quality, output, SES) are expected to explain some variance of English and MT vocabulary knowledge.

Question 2: What is the relative contribution of sets of internal versus external factors to predicting variation in children's bilingual vocabulary knowledge?

Hypothesis 2: Since English is prevalent in Singapore, bilingual children's English will be better predicted by internal factors, while their MT will be better predicted by external factors.

3. Methods

3.1 Participants

The data reported in this study are from part of a larger longitudinal study examining early childhood development in Singapore (Singapore Kindergarten Impact Project; Bull et al., 2014). Children in kindergarten 1 (K1) were recruited in Singapore across a various types of preschool, including public (24), government-run (10), not-for-profit (18), and commercial (2) preschool education providers. During the first wave of data collection, 1123 participants provided full data and 318 children were excluded from analysis because they have been actively exposed to more than two languages at home or their MT was not among the three official languages offered in local preschools. This resulted in a sample of 805 K1 children (4;1-5;8 years old) for the current study: 551 Mandarin-English speaking children, 105 Malay-English speaking children and 149 Tamil-English speaking children. The proportion of the bilingual groups is representative of the demographics of Singapore.

3.2 Measures and Procedures

Internal and external factors have been measured with a series of standard tests and parental questionnaires as follows.

3.2.1 Independent Variables: Measures on Internal Factors

Raven's Colored Progressive Matrices (Raven's; Raven, Court, & Raven, 1995) was used to measure children's nonverbal reasoning ability in the current study. Known to be an unbiased measure of cognitive functioning across various populations, it is highly suitable to allow for an evaluation of fluid intelligence without significant influence by external factors (e.g. cultural, educational). It has three sections: A, AB and B; each section has 12 items. Participants were instructed to choose the missing part of a presented matrix from an array of six options to make the matrix complete. Before administering the test items for Section A, there is a practice item where extensive and detailed feedback was given to the participant to ensure understanding of the task assigned. There were 36 items in total increasing in complexity. Administration of each section was stopped after four consecutive incorrect responses.

Elision and Blending, the sub-tests of the Comprehensive Test of Phonological Processing (CTOPP), were administered to measure children's phonological awareness (Wagner, Torgesen, & Rashotte, 1999). CTOPP is a reliable and valid measure of phonological awareness for individuals between the ages of 5 and 24. Elision is a 20-item subtest where a word is orally presented for the participant to repeat and then is instructed to remove a segment from the word. The participant would then say the resulting word. The items progressively get more difficult as the size of the removed segments gets smaller. Administration was discontinued after three consecutive incorrect test-item responses. Blending Words is a 20-item subtest that assesses young children's ability to combine sounds to form a whole word. The participant was instructed to combine individual sounds into a word by saying it aloud. As in Elision, the items in Blending increased in difficulty and the task was discontinued after three consecutive incorrect responses.

Backward Digit Recall (BDR) test was used to assess young children's working memory. The BDR task started with a series of digits (from 1 to 9) with a visual cue (arrow) to introduce the young participants to count backwards from 9 to 1. After the practice items, participants were instructed to listen to two numbers via an audio clip and to respond by reciting the numbers in backwards (reverse) order. The BDR task consists of six trials per block (total of 6 blocks) and with increasing digit spans after each block of trials. Administration was discontinued after incorrect responses to three test-items in a particular block.

3.2.2 Independent Variables: Parental Questionnaire on External Factors

A parental questionnaire was administered which included items adapted from prior studies to explore children's home language exposure (Lim, Lincoln, Chan, & Onslow, 2008; Sénéchal, LeFevre, Thomas, & Daley, 1998; Bedore et al., 2012; Gutiérrez-Clellen & Kreiter, 2003; Marian, Blumenfeld & Kaushanskaya, 2007) and literacy environment (Farver et al., 2006; Phillips & Lonigan, 2009; Burgess, Hecht & Lonigan, 2002). Parents completed the questionnaire during the child's K1 school year. They provided information about the extent of English spoken at home (1. Never, 2. Rarely, 3. Sometimes, 4. Often, 5. Primary), the amount of English and MT language each family member (i.e., mother, father, grandparents, siblings, others) spoke to children (0, 1/4, 1/2, 3/4, always), the amount of English and MT language children spoke to each family member (0, 1/4, 1/2, 3/4, always), the proportion of English and MT input from media at home , children's onset age of speaking English and the MT language, and the number of English and mother tongue books at home. Children's family background information was obtained from another parental survey, including children's chronological age, number of older siblings, mother's educational level (i.e., highest degree), family income level (1-20, with S$500 increment for each higher level), and children's waking hours spent with each family member on a typical weekday and weekend. Children's weekly input and output in English and MT languages at home was calculated based on the waking hours and the proportion of each language used with family members. The length of English/MT output was calculated using the current age to minus their onset age of speaking the language. Language dominance at home was estimated based on the item of the extent of English dominance at home. The larger the ranking number was, the more English (i.e., the less MT language) was assumed to be used in this family. In the end, there were nine external factors extracted from the parental questionnaires, five quantity oriented in nature (e.g., current weekly home input and output), and four quality oriented (e.g., number of language books in each language and number of older siblings) (Table 1).

3.2.3 Outcome Variables: Bilingual Vocabulary Knowledge Tests

Participants took part in a locally developed, standardized receptive picture vocabulary task, Bilingual Language Assessment Battery (BLAB) (Rickard Liow, Sze & Lee, 2013), in both English language and their MT (based on their school records; Mandarin Chinese, Malay). Previously developed versions of the BLAB in Mandarin and Malay (Rickard Liow and Sze, 2009) were used, and the items were translated into Tamil for the present study. Children participated in three practice items before proceeding to eighty trial items. The task had four-response options and children were asked to identify the picture that best corresponded to the word they heard. Stimuli were presented on an iPad monitor and via headphones.

Approximately 1.6% of data were missing mainly due to parent's overlooking of an item in questionnaires and Expectation-Maximization was used to replace the missing data.

3.3 Data Preparation and Analysis

The fundamental assumption of the current study is that Singapore preschoolers' bilingual language development occurs in two types of settings: an input-rich setting for English learning and an input-poor setting for MT language learning. Paired *t*-tests were used to explore the potential discrepancy in input and output between English and MT language and confirmed that English environment is significantly better than MT environment with regard to media input, number of books, weekly home input and weekly home output across all three MT populations (Table 1). Among these exposure variables, media input has the largest discrepancy between English and MT languages.

Regression models were used to explore the relative contribution of internal and external factors to children's English and MT receptive vocabulary knowledge. Children speaking different mother tongue languages were investigated separately. Non-parametric Spearman's correlations were conducted to check the relationships between the predictors. Those displaying a high correlation with other variables (r=.6-.9) would be averaged or excluded. Weekly home input was found to be highly correlated with weekly home output across three ethnic groups either in English or in MT language (e.g., English input and output of children in Mandarin group: r=.92). Therefore, values of the two variables were averaged into one score (e.g., WeekHomEng) to indicate general language usage at home per week. After the variable combinations, five internal factors were used in the subsequent analysis: age at the testing time ("TesAge"), gender ("Gender"), phonological awareness ("PhoAwa"), analytical reasoning ("NonInt"), and working memory ("WorMem"). Eight external factors were included: language input at home in general indicated by the extent of English dominance ("LanDom"), current weekly language usage at home ("WeekHomLan"), cumulative language output by children over years ("UseLen"), language input from media ("LanMedia"), mothers' educational level

(“MotEdu”), family income (“Income”), number of older siblings (“SibNum”) and number of book at home (“BookNum”). Table 2 presents the correlations between these predictors.

< Insert Table 2 about here >

To answer Question 1, backward regression analyses were conducted to explore the best predictors of children’s English and MT receptive vocabularies. Separate models were built according to language (English/MT) and for sets of children with different mother tongue languages (Chinese/Malay/Tamil), for a total of 6 models. Backward regression has been used to address possible multicollinearity effects (Field, 2013), as some of the predictors were moderately correlated as demonstrated in Table 2. It is vital to make sure that large portions of explained variance of the dependent variable are not shared between predictors, in order to allow interpretability of effects and reliability of estimated coefficients (Sun et al., 2016). Variance inflation factor (VIF) and tolerance indicators were examined for the six models, and all VIF values were below 10 and tolerance figures above 0.2, allowing us to safely conclude that there is no issue with multicollinearity in our final models (Field, 2013).

To answer Question 2, significant predictors selected by each final model in the backward regression analysis were categorized into two groups (internal factors vs. external factors) and were put in forced entry regression models. The aim is to demonstrate how internal factors and external variables as two distinct groups contributing to English and MT receptive vocabularies.

4. Results

4.1 Regression Analyses for English Receptive Vocabulary

In order to address the first research question on the contribution of individual internal and external factors to English receptive vocabulary, children using different MTs were analyzed separately and all 13 predictors were entered into each of the three backward regressions to find the best fitting models. The results are presented in Table 3 by mother tongue language group (Part I: MT=Mandarin, Part II: MT=Malay and Part III: MT=Tamil). All three models are significant and explained a large portion of the variance in English receptive vocabulary scores.

Specifically, in terms of Mandarin speaking children, approximately 34% of the variance ($F(6,544)=46.08$, $p < .001$) could be explained by six predictors in the besting fitting model: phonological awareness, nonverbal intelligence, working memory, English use length, number of English books at home and weekly English usage at home. Nonverbal intelligence made the largest contribution to English vocabulary ($\beta = .20$, semipartial $r = .18$). For Malay speaking children, approximately 31% of the variance ($F(6,544)=46.08$, $p < .001$) could be explained by four predictors in the besting fitting model: phonological awareness, working memory, English use length, and number of older siblings. Phonological awareness made the largest contribution to English vocabulary ($\beta = .28$, semipartial $r = .25$). In terms of Tamil speaking children, the final model explained about 37% of the variance ($F(2,146)=42.86$, $p < .001$) and two predictors were selected: phonological awareness and nonverbal intelligence. As in the model of Malay speaking children, phonological awareness made the largest contribution to English vocabulary ($\beta = .37$, semipartial $r = .33$).

As a further step, hierarchical regression was conducted to answer the second research question that relates to the relative contribution of internal and external factors when taken as two different sets for predicting variation in children's English receptive vocabulary. For Mandarin speaking children, a hierarchical regression analysis was conducted in which the internal factors (PhoAwa, NonInte, WorMem) were entered as a first block and the external factors (EngUseLen, EngBookNum and WeekHomEng) as a second block. Results revealed that the internal factors alone explained about 25% of the variance in English receptive vocabulary scores (R^2 = .25, F(3,547)=59.64, p < .001), and just 9% of additional variance was explained by entering the external factors (R^2 change = .09, F(6,544)= 46.08, p < .001). The same analysis was conducted on predictors selected by Malay speaking children's best fitting model. Internal factors (PhoAwa and WorMem) were entered as a first block and the external factors (EngUseLen and SibNum) were put in as a second block. It was found that the internal factors alone explained about 21% of the variance in English receptive vocabulary scores (R^2 = .21, F(2,102)=59.64, p < .001), and 7% of additional variance was explained by entering the external factors (R^2 change = .070, F(4,100)= 11.31, p < .001). In terms of Tamil speaking children, since no external factors were found to be significant predictors for English receptive vocabulary, no forced entry regression was conducted to compare the variance explained by internal and external factors. Across all the three groups, it was found that internal factors explained more variance than external factors.

< Insert Table 3 about here >

4.2 Regression Analyses for MT Receptive Vocabulary

As with the English receptive vocabulary skills, backward regression was used on MT receptive vocabulary scores by language group (Table 4; Part I: MT=Mandarin, Part II: MT=Malay and Part III: MT=Tamil) to address Question 1. In terms of Mandarin speaking children, seven predictors were maintained in the final model: phonological awareness, nonverbal intelligence, media input in MT, the extent of English dominance at home, length of language output in MT, number of MT books at home, and number of older siblings. The model accounted for 31% of the variance in Mandarin receptive vocabulary (F(7,543)=35.29, p<.001). Among the seven predictors, home language dominance had the strongest contribution to Mandarin vocabulary (β

= .24, semipartial r = .21). For Malay children, only two external factors, media input in MT and length of language output in MT, were maintained in the final model. They explained about 13.6% of the variance in Malay receptive vocabulary ($F(2,102)=8.06$, $p< .001$), with media input in Malay having the strongest contribution to Malay vocabulary (β = .27, semipartial r = .26). In terms of Tamil speaking children, the best fitting model held six predictors and all are significant: gender, nonverbal intelligence, media input in MT and length of language output in MT, number of books in MT, and family income. The final model captured approximately 33% of the variance in Tamil receptive vocabulary ($F(6,142)=11.73$, $p< .001$). Number of books in Tamil showed the largest contribution to Tamil vocabulary (β = .34, semipartial r = .33).

As the next step, forced entry regression was used again to compare the impact from internal and external factors, respectively, addressing Question 2. For Mandarin speaking children, internal factors from the best fitting model (PhoAwa and NonInte) were entered as a first block, followed by external factors (MTMedia, HomEngInput. MTUseLen, SibNum and MTBookNum) as a second block, to calculate the R^2 change. The data show that internal factors alone explained 5% of the variance in Mandarin receptive vocabulary scores ($F(2,548)= 15.694$, $p <.001$), and 26% of additional variance was explained by entering external factors ($F(7,543)= 35.287$, $p <.001$). In terms of Malay speaking children, since no internal factors were selected in the final model, all variance was explained by the external factors. For Tamil children, internal factors from the final model (Gender and NonInte) were entered as a first block, followed by external factors (MTMedia, MTUseLen, MTBookNum and Income) as a second block. The data show that internal factors alone explained 7% of the variance in Tamil receptive vocabulary scores ($F(2,146)= 5.54$, p = .005), and 26% of additional variance was explained by entering external factors ($F(6,142)= 11.73$, $p <.001$). Therefore, contrary to significant factors of English receptive vocabulary, external factors were found to explain more variance than internal factors of MT receptive vocabulary across all the ethnic-language groups. This interesting distinction will be discussed in detail in the Discussion session.

< Insert Table 4 about here >

5. Discussion

The current study examined how internal and external factors individually and as two sets affect the acquired vocabulary in English and MT language of very young Singapore bilinguals. This is one of the few studies that takes a comprehensive view to examine both internal and external factors on bilingual preschooler's dual language learning. More importantly, it aims at demonstrating how the learning environment may modulate the effects of internal and external factors on vocabulary learning. A common pattern for English receptive vocabulary was that predictors related to internal abilities (e.g., phonological awareness and working memory) explained more variance than external factors (e.g., length of language use) which was found to be consistent with findings of studies in input-rich environments (Paradis, 2011). In contrast, MT receptive vocabulary has an opposite pattern applied for all groups. Namely, external factors explained more variance in the data than internal factors, a pattern consistent with observations regarding language learning in input-poor environment (Sun et al., 2016).

5.1 Internal and External Factors

5.1.1 Internal Factors in English and MT Receptive Vocabulary

Phonological awareness, a measure of metalinguistic knowledge of phonemes and phonological structure, was found to significantly predict receptive English vocabulary for all bilingual learners and receptive Chinese vocabulary for Mandarin speaking learners. Previous research has demonstrated the central role of phonological awareness in word reading and spelling (e.g., Gillon, 2004) and the present study extended such an association to vocabulary knowledge. General analytical reasoning ability, measured with Raven’s Coloured Progressive Matrices, was a significant predictor for both Mandarin-speaking children’s and Tamil-speaking children’s English and MT receptive vocabulary. Working memory, measured by backward digit recall, also emerged as a significant predictor for English vocabulary with Mandarin-speaking and Malay-speaking children. These findings pertaining to the role of language aptitude are generally consistent with previous findings from child bilingual studies (French & O’Brien, 2008; Paradis, 2011), demonstrating the robustness of language aptitude in predicting preschool children’s bilingual receptive vocabulary knowledge. Furthermore, gender emerged as a significant predictor for Tamil vocabulary. Since such an association has been only observed in one ethnic group, we are reluctant to attribute this to biological reasons as explained in some studies (e.g., Pivik et al., 2011; Eriksson et al., 2012), but suggest future studies to explore it from a social-cultural perspective.

5.1.2 External Factors in English and MT Receptive Vocabulary

External factors are less consistent than internal factors in predicting English and MT receptive vocabulary, with only length of language output being common across the languages. We therefore discuss English and MT separately for the effect of external factors.

Language output in English (i.e., EngUseLen) was significantly associated with Mandarin-speaking and Malay-speaking children’s English receptive vocabulary. Current weekly English usage, which included the average of English input and output at home, was also significant for Mandarin-speaking children’s English receptive vocabulary. In general then, this finding points to the crucial role that language exposure, especially output, might play in child bilingual vocabulary knowledge, keeping in line with previous studies (Bohman et al., 2010;

Montrul, 2008; Paradis, 2011). The number of English books at home was found to significantly predict Mandarin-speaking children's English vocabulary knowledge. This echoes previous findings that a rich and sound literacy environment at home is important for vocabulary growth (e.g., the Turkish-Dutch children in Scheele et al., 2010). Lastly, the number of older siblings was found to be positively related to Malay-speaking children's English vocabulary, consistent with findings on the importance of child-to-child communication in English learning (Gutiérrez-Clellen & Kreiter, 2003).

Similar to what was found for English, language output in MT was also associated with children's MT receptive vocabulary across the three ethnic groups. Even though previous research has argued for the importance of language use (i.e., output) in facilitating the development of accuracy and automaticity in grammar knowledge (Bohman et al., 2010), the present study found that output is also important for vocabulary growth, as was observed for MT learning in our participants just as for English. Media input emerged as a significant factor across all three groups in predicting MT vocabulary and the extent of English dominance at home was found to negatively correlate to Mandarin-speaking children's Chinese vocabulary. This generally confirms the importance of input quantity to child bilingual's vocabulary learning found by other studies (e.g., Place & Hoff, 2011; Unsworth, 2013). In an environment where children could only receive limited MT language input from home and school, media offers a particular channel for children to get valuable exposure. As was observed for English learning, number of MT books at home was also found to significantly predict Mandarin-speaking children's Chinese receptive vocabulary, and Tamil-speaking children's Tamil receptive vocabulary. Differing from English, the number of older siblings was found to be negatively associated with Mandarin-speaking children's Chinese vocabulary. The seemingly contradictory findings for English and MT might be explained from a sociolinguistic perspective. Older siblings in an immigrant family have been argued to introduce more society language to the home domain (Paradis, 2011). In the present study conducted in a multilingual society, it did appear that siblings prefer to speak English, the societally dominant language, at home. Based on paired t-tests, we have found a significant difference between siblings' amount of time spent on English use and on MT use (t(15.28)=, p =.000). Regarding the role of family income, previous findings on monolingual populations showed a positive relationship between SES and vocabulary skills (i.e., children from richer homes command a larger vocabulary than those from poorer homes). However, in the present study, family income was negatively related to Tamil vocabulary, a finding in Dixon's study (2012) as well, which may be explained by the preference of Singapore high-income Indian families to use English at home (e.g., Saravanan, Lakshmi, & Caleon, 2007). In the current study, Tamil speaking children's family income is indeed found to be significantly and negatively correlated to children's weekly MT output at home (r = -.23**).

To sum up, both internal and external factors have been found to significantly predict bilingual children's vocabulary knowledge. Specifically, language aptitude and language exposure (i.e., length of language output in months) functioned in a similar manner across the three ethnic groups on both English and MT languages. In terms of language aptitude, the current study demonstrates the underlying acquisition mechanism shared by the dual languages of a bilingual (Cummins, 1979) and verifies the fundamental role that phonemic coding ability, language analytic ability and verbal memory plays in bilingual children's vocabulary knowledge (Paradis, 2011). The current study also highlights the crucial role of language output in bilingual children's vocabulary development. Language output as part of human psychological processes originates from collective behaviors (e.g. discourse) and facilitates the internalization of language knowledge as part of one's mental activity (Stetsenko & Arievitch, 1997; Swain, 2007). In attempting to produce language, learners may begin to notice the gap between what they want to express and what they can express. In addition, erroneous language produced by the learner may invite implicit/explicit feedback from the interlocutors. These aspects of language output may therefore enable children to notice their linguistic limitations and to mitigate their language problems via conversations with competent speakers.

5.2 Internal Factors vs External Factors

The current study also shows that, at least in the learning of English, as a socially dominant language in Singapore, internal resources play a more significant role than external resources in vocabulary acquisition. This echoes findings of studies conducted in similar learning settings (e.g., Paradis, 2011). At the same time, our study shows that in the learning of MTs, that are considered less socially dominant, external resources should be given more weight, which also echoes previous findings (Sun et al., 2016). It revealed that learning contexts play an important role in determining the relative weight of internal versus external factors. Specifically, an input-rich context characterized by abundant language exposure allows the working of internal resources to manifest. In contrast, an input-poor context characterized by impoverished language input suppresses the role of internal factors. Paired contrasts showed a significant difference between English and MT learning contexts in terms of the quantity and quality of input, which we take to mean that the context for the English is more input-rich than that for the ethnic language.

In sum, there seems to be a "see-saw" effect between internal and external factors, depending on the learning context. This, as mentioned earlier in the paper, is reflective of assumptions in the critical mass hypothesis, in that we can see the input (external factors) as exerting a constraining effect on internal factors unless and until the former surpasses a threshold level (Marchman & Bates, 1994). Other studies showing the effect of critical mass include Bates, Bretherton, & Snyder (1988), who found that vocabulary size was the best predictor of mean length of utterance at a later age, which can be taken to mean that complexity in language is contingent upon sufficient exposure to linguistic input enabling the extraction of linguistic patterns. Our explanation in terms of the difference in learning context finds broad support in the language acquisition literature as well. For instance, with regard to first language learning (an input-rich context), the argument has sometimes been that linguistic cognition (internal factor) trumps language input (external factor), illustrated in the so-called Poverty-of-Stimulus situations where grammatical competence emerges despite lack of decisive evidence in the input (Crain, 1991; Singleton & Newport, 2004). On the other hand, observations of instructed adult second language learners have frequently pointed to the lack of a role of internal factors (such as working memory, Juffs, 2004; Clahsen & Felser, 2006) and to an increased importance of external variables (such as instruction and practice, Bley-Vroman, 1990).

Findings for question 1 and question 2 are compatible with the fundamental assumptions within the usage-based tradition, but expand on these assumptions. Findings for the first question in general support the assumption that language learning is an interaction between input material and domain-general characteristics. Bilingual children's vocabulary emerges from their experience with language (e.g., input and output), together with the general cognitive skills (e.g., phonological awareness) to make categorization and linguistic representations based on a rich memory of exemplars. Findings for the second question demonstrated the modulation of general learning context on the relationship between external and internal factors in bilingual children's vocabulary learning. While some researchers within the tradition of usage based theory give primary recognition to the role of input properties such as frequency information (e.g., Ellis, 2002), we might consider adopting a more eclectic view about the weight of internal and external factors on language acquisition, answering the call of Dąbrowska (e.g., 2012, 2016), who addresses the importance of paying more attention to examining individual differences in learners' internal and external resources while examining learning outcomes (also see Verhagen & Mos, 2016). Cognitive processes, such as categorization, chunking, rich memory, analogy and cross-modal association, have been found to influence language development (Bybee, 2010); yet it remains unclear under which circumstances these factors better demonstrate their prominence. The current study revealed the dominant role of input environment at the initial stage of learning, showing that while both internal and external factors were crucial for bilingual children's vocabulary learning, they have different degrees of importance depending on the specific input context.

6. Limitations

There are several limitations of the present study. First of all, no causality between internal/external factors and receptive vocabulary knowledge could be generated from the current cross-sectional study. Researchers could adopt a longitudinal design to examine the association bidirectionally in the future. Moreover, only receptive vocabulary knowledge has been examined and future studies could include knowledge from various language domains (e.g., expressive vocabulary and receptive grammar) to verify the impact of language learning context on these linguistic domains. Furthermore, language environment has been only comprehensively scrutinized at home whereas contextual factors at preschool (e.g., MT instruction hours and teacher's proficiency level) may be similarly influential and should also be considered in future studies (Unsworth et al., 2015). In addition, estimating children's language environment based on parental survey is less objective and therefore not necessarily accurate. Environmental assessing technology, such as the Language Environment Analysis (LENA) system, might be adopted in future studies to more precisely capture the input and output quantity between a child and his/her family members. Last but not least, it remains unclear where the threshold is or when the critical mass is reached. Future studies might longitudinally follow older children from the same groups to explore the potential threshold.

7. Conclusion and Implications

In the present study, we examined the influence of external and internal factors on Singapore bilingual children's receptive vocabulary knowledge in both of their languages. We found that, while both sets of factors explain lexical knowledge in each language, the relative contribution is different for English (the societal language) compared with ethnic languages. Specifically, internal factors explained more variance in English language vocabulary, whereas external factors were more important in explaining ethnic language knowledge. We attribute this difference to the crucial role of exposure in language learning, as suggested by the critical mass hypothesis. Our findings in general are consistent with previous research pointing to the role of the home language environment in the maintenance of heritage/ethnic languages (e.g., Dixon et al., 2012). The significance role that external factors played in all three MT vocabulary supports Pearson's (2007) comment that "learning to speak two language in one's personal or professional life is always a possibility, but never a given" (p.399). If children could not enjoy ample input from the environment, their learning might be greatly affected. The results in the current study regarding language output imply that children can be encouraged to start to use their dual languages actively at a young age, as doing so would help build up their receptive vocabulary above and beyond the impact of language input (Sun, et al., 2016). Future research may follow a group of preschoolers longitudinally and verify whether the discrepancy of effects of internal and external factors on English and MT learning would last over time. It would be necessary to further specify the amount of input and output children need to become fluent bilinguals in the interest of making concrete policy suggestions.

References

Allman, B. (2005).Vocabulary size and accuracy of monolingual and bilingual preschool children. In J. Cohen, K.T. McAlister, K. Rolstad and J. MacSwan (eds). *Proceedings of the 4th International Symposium on Bilingualism* (pp. 58-77). Somerville, MA: Cascadilla Press.

Alexiou, T. (2009). Young learners' cognitive skills and their role in foreign language vocabulary learning. In M. Nikolov (Ed.), *Early learning of modern foreign languages: Processes and outcomes* (pp. 46–61). Clevedon: Multilingual Matters.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms.* New York: Cambridge University Press.

Bauer, D. J., Goldfield, B. A., & Reznik, J. S. (2002). Alternative approaches to analyzing individual differences in the rate of early vocabulary acquisition. *Applied Psycholinguistics, 23*, 313-335.

Bedore, L. M., Peña, E. D., Summers, C. L., Boerger, K. M., Resendiz, M. D., Greene, K. A. I., Bohman, T. M., & Gillam, R. B. (2012). The measure matters: Language dominance profiles across measures in Spanish–English bilingual children. *Bilingualism:Language and Cognition, 15*(3), 616-629. doi: 10.1017/S1366728912000090

Bley-Vroman, R. (1990). The logical problem of foreign language learning. *Linguistic Analysis, 20*, 3–49.

Bohman, T. M., Bedore, L. M., Pena, E. D., Mendez-Perez, A., & Gillam, R. B. (2010). What you hear and what you say: Language performance in Spanish-English bilinguals. *International Journal of Bilingual Education and Bilingualism, 13*(3), 325-344. doi: 10.1080/13670050903342019

Bokhorst-Heng, W. D. (1999). Bilingual education and the dialectics of national integration. In J. E. Alatis & A.-H. Tan (Eds.), *Proceedings of Georgetown University round table on languages and linguistics* (pp. 356–370). Washington, DC: Georgetown University Press. Retrieved from: http://aladinrc.wrlc.org/bitstream/1961/3455/35/gurt_1999_fm.pdf

Bornstein, M. H., Hahn, C.-S., & Haynes, O. M. (2004). Specific and general language performance across early childhood: Stability and gender considerations. *First Language*, *24*(3), 267-303.

Bouchard, C., Trudeau, N., Sutton, A., Boudreault, M.-C., & Deneault, J. (2009). Gender differences in language development in French Canadian children between 8 and 30 months of age. *Applied Psycholinguistics, 30*(4), 685-707. doi: 10.1017/S0142716409990075

Bull, R., Lee, K., Karuppiah, N., Poon, K., O'Brien's, B., Towndrow, P., Bautista, A. (2014). Preschool impact upon child outcomes: A longitudinal study of transition from preschool to primary, Office of Educational Research Education Research Funding Programme, National Institute of Education, Nanyang Technological University, Singapore.

Burgess, S., Hecht, S., & Lonigan, C. (2002). Relations of the home literacy environment (HLE) to the development of reading-related abilities: A one-year longitudinal study. *Reading Research Quarterly, 37,* 408-426.

Carroll, J. B. & Sapon, S. (2002). *Modern Language Aptitude Test: Manual 2002 Edition.* Bethesda, MD: Second Language Testing, Inc.

Clahsen, H. & Felser, C. (2006). Grammatical processing in language learning. *Applied Psycholinguistics 27,* 3– 42.

Clark, E. V. (2003). *First Language Acquisition*. Cambridge: Cambridge University Press.

Chondrogianni, V., & Marinis, T. (2011). Differential effects of internal and external factors on the development of vocabulary, morphology and complete syntax in successive bilingual children. *Linguistic Approaches to Bilingualism, 1*(3), 223-248.

Collentine, J., & Freed, B. (2004). Learning context and its effects on second language acquisition: Introduction. *Studies in Second Language Acquisition, 26*(2), 153-172.

Crain, S. (1991). Language acquisition in the absence of experience. *Behavioral and Brain Sciences, 14,* 597–650.

Cummins, J. (1979). Linguistic Interdependence and the Educational Development of Bilingual Children. *Review of Educational Research 49*: 222–251.

Cummins, J., & Swain, M. (1986). *Bilingual Education: Aspects of Theory, Research and Practice.* London: Longman.

Dąbrowska, E. (2012). Different speakers, different grammars: Individual differences in native language attainment. *Linguistic Approaches to Bilingualism 2*, 219-253

Dąbrowska, E. (2016). Cognitive Linguistics' seven deadly sins. *Cognitive Linguistics, 27*(4), pp. 479-491. doi:10.1515/cog-2016-0059

Daller, M., & Ongun, Z. (2017). The threshold hypothesis revisited: bilingual lexical knowledge and non-verbal IQ development. *International Journal of Bilingualism.* ISSN 1756-6878 doi: 10.1177/1367006917690835

Dixon, L. Q., Wu, S., & Daraghmeh, A. (2012). Profiles in bilingualism: Factors influencing kindergartners' language proficiency. *Early Childhood Education Journal, 40*(1), 25-34. doi: 10.1007/s10643-011-0491-8

Dixon, L. Q., Zhao, J., Quiroz, B. G., Shin, J. (2012). Home and community factors influencing bilingual children’s ethnic language vocabulary development. *International Journal of Bilingualism, 16*(4) 541–565.

Ellis, R. (2002). Does form-focussed instruction affect the acquisition of implicit knowledge? A review of the research. *Studies in Second Language Acquisition, 24,* 223-236.

Elman, J. L. (2003). Generalization from sparse input. *Proceedings of the 38th Annual Meeting of the Chicago Linguistic Society.*

Eriksson, M., Marschik, P. B., Tulviste, T., Almgren, M., Pérez, P. M., Wehberg, S., & Gallego, C. (2012). Differences between girls and boys in emerging language skills: Evidence from 10 language communities. *The British Journal of Developmental Psychology, 30,* 326-343. doi: 10.1111/j.2044-835X.2011.02042

Farver, J. M., Xu, Y., Eppe, S., & Lonigan, C. J. (2006). Home environments and young Latino children's school readiness. *Early Childhood Research Quarterly, 21,* 196-212.

Field, A. (2013). *Discovering statistics using IBM SPSS Statistics (4th edition)*. London: Sage. ISBN 9781446249178

French, L. M., & O'Brien, I. (2008). Phonological memory and children's second language grammar learning. *Applied Psycholinguistics, 29*(3), 463-487. doi: 10.1017/S0142716408080211

Gathercole, S. E., & Baddeley, A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*, 200-213.

Gathercole, V.C.M., & Hoff, E. (2007). Input and the acquisition of language: Three questions. In E. Hoff & M. Shatz (Eds.), *The handbook of language development* (pp. 107–127). Oxford: Blackwell Publishers.

Gillon, G. (2004). *Phonological awareness: From research to practice*. New York: The Guilford Press.

Goldberg, H., Paradis, J., & Crago, M. (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics, 29*(1), 41-65. doi: 10.1017/S014271640808003

Gutiérrez-Clellen, V. F., & Kreiter, J. (2003). Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics, 24*(2), 267-288. doi: 10.1017/S0142716403000158

Hakuta, K., & Diaz, R. M. (1985). The relationship between degree of bilingualism and cognitive ability: A critical discussion and some new longitudinal data. *Children's language, 5*, 319-344.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. *Developmental Psychology, 27*, 236-248.

Ibbotson, P. (2013). The Scope of Usage-Based Theory. *Frontiers in Psychology, 4,* 255. http://doi.org/10.3389/fpsyg.2013.00255

Juffs, A. (2004). Representation, Processing and Working Memory in a Second Language. *Transactions of the Philological Society, 102*(2), 199-225. doi: 10.1111/j.0079-1636.2004.00135.

Knell, E., Qiang, H.Y., Pei, M., Chi, Y. P., Siegel, L. S., Zhao, L., Zhao, W (2007). Early English immersion and literacy in Xi'an, China. *Modern Language Journal, 91,* 395-417.

Lambert, W. E., & Taylor, D. M. (1996). Language in the lives of ethnic minorities: Cuban-American families in Miami. *Applied Linguistics, 17*, 477-500.

Lim, V. P., Lincoln, M., Chan, Y. H., & Onslow, M. (2008). Stuttering in English-Mandarin Bilingual Speakers: The influence of language dominance on stuttering severity. *Journal of Speech, Language and Hearing Research, 51*, 1522-1537. doi: 1092-4388/08/5106-1522

Marchman, V. A., & Bates, E. (1994). Continuity in lexical and morphological development: A test of the critical mass hypothesis. *Journal of Child Language, 21*, 339–366.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language and Hearing Research, 50*(940-967). doi: 1092-4388/07/5004-0940

Meara, P. (1997). Towards a new approach to modelling vocabulary acquisition. In Schmitt, N., & Mc Carthy, M. (eds) *Vocabulary: Description, Acquisition and Pedagogy* (pp. 109-121). Cambridge: Cambridge University Press.

Metsala, J.L. & Walley, A.C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J.L. Metsala and L.C. Ehri (eds.), *Word recognition in beginning reading.* Mahwah, NJ: Lawrence Erlbaum Associates, pp. 89–120.

Montrul, S. (2008). *Incomplete acquisition in bilingualism: Reexamining the age factor*. Amsterdam: John Benjamins.

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics, 29*(4), 578-596.

Nation, I. S. P. (1990). *Teaching and Learning Vocabulary.* New York: Newbury House.

O’Grady, W. (2008). The emergentist program. *Lingua, 118*, 447-64.

Oller, D. K., & Eilers, R. E. (2002). *Language and literacy in bilingual children*. Clevedon, UK: Multilingual Matters.

Pakir, A. (1999). Bilingual education with English as an official language: Socio-cultural implications. In J. E. Alatis & A.-H. Tan (Eds.), *Georgetown roundtable of linguistics archive* (pp. 341–349). Washington, DC: Georgetown University Press. Retrieved from: http://aladinrc.wrlc.org/bitstream/1961/3455/26/gurt_1999_25.pdf

Paradis, J. (2007). Second language acquisition in childhood. In E. Hoff and M. Shatz (Eds.), *Handbook of Language Development* (pp. 387-406). Oxford: Blackwell.

Paradis, J. (2011). Individual differences in child English second language acquisition: Comparing child-internal and child external factors. *Linguistic Approaches to Bilingualism, 1,* 213–237.

Pearson, B. Z. (2007). Social factors in childhood bilingualism in the United States. *Applied Psycholinguistics, 28*, 399-410.

Phillips, B. M., & Lonigan, C. J. (2009). Variations in the home literacy environment of preschool children: A cluster analytic approach*. Scientific Studies of Reading, 13*(2), 146-174. doi: 10.1080/10888430902769533

Pi-Yu, C., & Rvachew, S. (2007). English-French bilingual children's phonological awareness and vocabulary skills. *Canadian Journal of Applied Linguistics, 10*(3), 292-307.

Pivik, R. T., Andres, A., & Badger, T. M. (2011). Diet and gender influences on processing and discrimination of speech sounds in 3- and 6-month-old infants: A developmental ERP study. *Developmental Science*, *14*(4), 700-712. doi: 10.1111/j.1467-7687.2010.01019

Place, S., & Hoff, E. (2011). Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development, 82*(6), 1834-1839. doi: 10.1111/j.1467-8624.2011.01660

Raven, J. C., Court, J. H., & Raven, J. (1995). *Coloured progressive matrices manual.* London: Lewis.

Rickard Liow, S.J., & Sze, W.P. (2009). *Bilingual language assessment battery (BLAB).* Unpublished measure. Singapore: National University of Singapore, Department of Psychology and Division of Graduate Medical Studies.

Rickard Liow, S. J., Sze, W. P and Lee L.C. (2013). *Bilingual Language Assessment Battery (BLAB) Manual (Unpublished measure).* Singapore: National University of Singapore, Department of Psychology and Division of Graduate Medical Studies.

Rojas, R., Iglesias, A., Bunta, F., Goldstein, B., Goldenberg, C., & Reese, L. (2016). Interlocutor differential effects on the expressive language skills of Spanish-speaking English learners. *International Journal of Speech-Language Pathology, 18*(2), 166-177.

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

Saravanan, V., Lakshmi, S., & Caleon, I. (2007). Attitudes towards literary Tamil and standard spoken Tamil in Singapore. *International Journal of Bilingual Education and Bilingualism, 10*(1), 58-79.

Scheele, A. F., Leseman, P. P. M., & Mayo, A. Y. (2010). The home language environment of monolingual and bilingual children and their language proficiency. *Applied Psycholinguistics, 31*(1), 117-140. doi: 10.1017/S0142716409990191

Sénéchal, M., LeFevre, J. A., Thomas, E. M., & Daley, K. E. (1998). Differential effects of home literacy experiences on the development of oral and written language. *Reading Research Quarterly, 33*(1), 96-116. doi: 10.1598/RRQ.33.1.5

Singapore Department of Statistics. (2016). *Population Trends 2016*. ISSN: 1793-2424

Singleton, J. L. and Newport, E. L. (2004). When learners surpass their models: The acquisition of American Sign Language from inconsistent input. *Cognitive Psychology 49*(4), 370–407.

Stetsenko, A., & Arievitch, I. (1997). Constructing and deconstructing the self: Comparing post-Vygotskian and discourse-based versions of social constructivism. *Mind, Culture and Activity, 4,* 160–173.

Sun, H. E., Steinkrauss, R., Tendeiro, J., & De Bot, K. (2016). Individual differences in very young children's English acquisition in China: Internal and external factors. *Bilingualism: Language and Cognition*, *19*(3), 550-566. doi: 10.1017/S1366728915000243

Sunderman, G., & Kroll, J. F. (2009). When study-abroad experience fails to deliver: The internal resources threshold effect. *Applied Psycholinguistics*, *30*(1), 79-99.

Swain, M. (2007). *The Output Hypothesis: Its History and Its Future*. Beijing: Beijing Foreign Studies University.

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

Unsworth, S. (2013). Assessing the role of current and *cumulative* exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism: Language and Cognition, 16*, 86-110. http://dx.doi.org/10.1017/S1366728912000284

Unsworth, S. (2016). Quantity and quality of language input in bilingual language development. In E. Nicoladis & S. Montanari (Eds.), *Lifespan perspectives on bilingualism* (pp. 136-196). Mouton de Gruyter/APA.

Unsworth, S., Hulk, A., & Marinis, T. (2011). Internal and external factors in child second language acqusition. *Introduction to Spotlight Issue of Linguistic Approaches to Bilingualism, 1*, 207-212.

Unsworth, S., Persson, L., Prins, T. & de Bot, K. (2015). An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics, 36*, 527-548. http://dx.doi.org/10.1093/applin/amt052.

Verhagen, V. & Mos, M. (2016). Stability of familiarity judgments: Individual variation and the invariant bigger picture . *Cognitive Linguistics*, *27*(3), pp. 307-344. Retrieved 14 Aug. 2017, from doi:10.1515/cog-2015-0063

Wagner R. K., Torgesen J. K., Rashotte C. A. (1999). *Comprehensive Test of Phonological Processing (CTOPP)*. Austin, TX: PRO-ED.

Table 1

Raw scores of predictors and outcome variables

	Descriptive Statistics: Average (SD)								
	Mandarin			Malay			Tamil		
	English	MT	*t*	English	MT	*t*	English	MT	*t*
Internal variables									
TesAge	57.23 (3.77)			57.26 (3.88)			57.68 (3.84)		
Gender	0.50 (0.50)			0.50 (0.50)			0.48 (0.50)		
PhoAwa	14.66 (7.72)			9.96 (5.54)			14.84 (9.51)		
NonInte	15.96 (4.84)			13.54 (4.60)			13.39 (4.61)		
WorMem	2.83 (3.42)			1.57 (2.56)			2.46 (3.49)		
External variables									
HomDom	3.41 (0.89)			3.30 (0.82)			3.12 (0.87)		
WeekHomInput	14.50 (9.11)	8.45 (7.10)	12.29***	16.58 (8.92)	12.24 (8.77)	4.08***	10.59 (7.48)	8.96 (6.92)	1.88.
WeekHomOutput	15.38 (9.57)	8.33 (7.39)	13.29***	16.88 (8.78)	12.64 (9.38)	3.69***	11.67 (8.36)	8.71 (7.59)	3.03**
HomMedia	78.(21.07)	21.50(20.81)	31.88***	80.29 (18.10)	19.42 (18.12)	17.23***	77.09 (21.19)	20.50 (20.7)	16.85***
BookNum	3.11 (1.56)	1.99 (1.14)	22.16***	2.29 (1.13)	1.15 (0.85)	10.79***	2.48 (1.35)	1.14 (1.01)	12.78***
UseLen	32.32(12.78)	28.5 (13.04)	5.56***	25.13 (13.95)	26.76 (14.45)	-0.94	27.74 (12.47)	27.37(13.73)	0.25
SibNum	0.78 (0.82)			1.15 (1.11)			0.67 (0.71)		
MotEdu	7.22 (2.34)			5.49 (2.09)			8.33 (2.19)		
Income	13.45 (5.56)			6.58 (4.88)			11.51 (5.41)		
Outcome variable									
RecVoc	36.45 (8.72)	30.24 (7.82)		33.01 (7.85)	29.38 (7.00)		33.54 (8.61)	24.28 (7.11)	

Note. TesAge = age at the time of testing in months; Gender = children's gender (0=male, 1=female); PhoAwa = phonological awareness based on sum of elision and blending tasks ; WorMem = working memory score based on backward digits recall; NonInte = non-verbal IQ scores as a measure of analytic reasoning; HomDom = the extent of English dominance at home; WeekHomInput = the amount of language input to children at home in hours per week; WeekHomOutput = the amount of children's language output at home in hours per week; HomMed = the percentage of Media input in English and mother tongue; UseLen= number of months of children using English or mother tongue actively; BookNum = number of English or mother tongue books at home on a 1-7 point scale; SibNum = number of older siblings at home; MotEdu = mothers' highest educational level; Income = monthly family income on a 1-20 increasing point scale; Vocab = receptive vocabulary size based on BLAB tests. .p<1, **p<.005; ***p<.001.

Table 2

Non-parametric correlations of the predictors in the model analysis

	English variables		1	2	3	4	5	6	7	8	9	10	11	12	13
1	TesAge		1												
2	Gender	MT=1	-0.03												
		MT=2	0.02	1											
		MT=3	-0.04												
3	PhoAwa	MT=1	.31**	0.02											
		MT=2	0.02	-.22*	1										
		MT=3	.26**	.17*											
4	NonInte	MT=1	.32**	-0.04	.32**										
		MT=2	.32**	0.03	.30**	1									
		MT=3	.26**	0.06	.45**										
5	WorMem	MT=1	.22**	0.02	.44**	.30**									
		MT=2	0.17	-0.04	.32**	.33**	1								
		MT=3	.20*	0.11	.36**	.29**									
6	HomMed	MT=1	0.02	-0.12	.14**	.17**	.13**								
		MT=2	-0.14	-0.10	0.06	0.09	0.03	1							
		MT=3	0.05	.17*	0.16	.26**	0.09								
7	HomDom	MT=1	0.08	-0.04	.20**	0.07	.11*	.39**							
		MT=2	-0.10	0.00	.23*	0.18	0.01	.43**	1						
		MT=3	-0.02	0.06	0.08	0.04	0.08	0.08							
8	UseLen	MT=1	.33**	-0.01	.29**	.14**	.25**	.25**	.41**						
		MT=2	0.07	0.07	.20*	0.17	0.15	.24*	.40**	1					
		MT=3	.44**	.17*	.40**	.31**	.20*	0.12	.38**						
9	SibNum	MT=1	.11**	0.06	0.03	-0.03	-0.02	0.02	.14**	0.07					
		MT=2	0.09	-0.10	0.00	-0.14	-0.08	-0.04	-0.04	-0.05	1				
		MT=3	0.03	-0.10	0.08	-0.06	0.02	0.00	-.29**	-0.09					
10	BookNum	MT=1	.11**	-0.05	.35**	.18**	.22**	.21**	.29**	.29**	.17**				
		MT=2	-0.03	0.03	0.10	0.08	.22*	0.07	0.01	.31**	0.08	1			
		MT=3	0.05	0.06	.37**	0.15	0.11	0.10	.35**	.36**	0.02				
11	MotEdu	MT=1	0.01	-0.02	.31**	.14**	.21**	.15**	.14**	.18**	0.02	.38**			
		MT=2	-0.04	-0.02	-0.01	0.12	0.17	.29**	0.12	0.13	-0.19	.21*	1		
		MT=3	0.05	0.07	.18*	.17*	0.14	.25**	-.30**	-0.01	.19*	0.04			

12	Income	MT=1	.11**	-0.02	.33**	.16**	.24**	.18**	.27**	.27**	0.08	.30**	.55**		
		MT=2	-0.03	-.24*	0.16	.27**	.30**	.20*	0.16	0.15	-0.06	.35**	.42**	1	
		MT=3	.23**	0.12	.33**	.21*	0.15	0.14	-0.02	0.12	0.07	.34**	.42**		
13	WeekHom	MT=1	.09*	-0.02	.19**	0.03	.09*	.34**	.51**	.35**	.26**	.26**	.12**	.22**	
		MT=2	0.00	0.02	.26**	.29**	0.03	.26**	.47**	.27**	-0.01	.23*	0.08	.25**	1
		MT=3	-0.09	0.01	0.05	-0.04	0.07	.18*	.30**	.24**	-0.08	.17*	-.19*	-0.10	

	MT variables		1	2	3	4	5	6	7	8	9	10	11	12	13
1	TesAge														
			1												
2	Gender	MT=1	-0.03												
		MT=2	0.02	1											
		MT=3	-0.06												
3	PhoAwa	MT=1	.30**	0.02											
		MT=2	0.05	-.22*	1										
		MT=3	.26**	.17*											
4	NonInte	MT=1	.32**	-0.04	.32**										
		MT=2	.33**	0.03	.30**	1									
		MT=3	.26**	0.06	.45**										
5	WorMem	MT=1	.21**	0.02	.44**	.30**									
		MT=2	0.17	-0.04	.32**	.33**	1								
		MT=3	.19*	0.11	.36**	.29**									
6	HomMed	MT=1	-0.02	.13**	-0.13**	-0.16**	-0.12**								
		MT=2	0.13	0.09	-0.06	-0.09	-0.01	1							
		MT=3	-0.05	-0.12	-0.15	-.18*	-0.10								
7	HomDom	MT=1	.08*	-0.04	.20**	0.07	.11*	-0.38**							
		MT=2	-0.10	0.00	.23*	0.18	0.01	-.433**	1						
		MT=3	-0.03	0.06	0.08	0.04	0.08	-0.12							
8	UseLen	MT=1	.17**	0.05	.09*	0.05	0.06	.23**	-0.30**						
		MT=2	.33**	0.14	-0.09	0.14	0.07	.33**	-.25*	1					
		MT=3	.34**	-0.02	.19*	.17*	.19*	.20*	-.29**						
9	SibNum	MT=1	.12**	0.06	0.03	-0.03	-0.02	-0.02	.14**	-0.02					
		MT=2	0.10	-0.10	0.00	-0.14	-0.08	0.02	-0.04	-0.03	1				
		MT=3	0.04	-0.10	0.08	-0.06	0.02	0.12	-.29**	0.12					
10	BookNum	MT=1	.11*	0.01	.23**	.12**	.15**	0.07	0.02	.18**	.16**				
		MT=2	0.12	-0.07	-0.02	0.07	0.15	.22*	-.30**	0.03	0.07	1			
		MT=3	-0.08	-0.05	0.16	0.03	-0.05	.26**	0.09	0.12	0.16				
11	MotEdu	MT=1	0.01	-0.02	31**	.14**	.21**	-0.15**	.14**	0.01	0.02	.29**			
		MT=2	-0.05	-0.02	-0.01	0.12	0.17	-.29**	0.12	-0.02	-0.19	0.13	1		
		MT=3	0.05	0.07	0.18*	.17*	0.14	-.21*	-.30**	.25**	.19*	0.09			

12	Income	MT=1	.11*	-0.02	.33**	.16**	.24**	-0.18**	.27**	0.03	0.08	.16**	.55**		
		MT=2	-0.04	-.24*	0.16	.27**	.30**	-.19*	0.16	-.22*	-0.06	.30**	.42**	1	
		MT=3	.22**	0.12	.33**	.21*	0.15	-0.11	-0.02	0.11	0.07	0.08	.42**		
13	WeekHom	MT=1	-0.06	.10*	-0.18**	-0.10**	-0.09**	.37**	-0.49**	.32**	-0.05	-0.01	-0.29**	-0.23**	
		MT=2	0.12	0.16	0.06	0.18	-0.09	.29**	-.33**	.38**	0.05	0.10	-0.14	-0.10	1
		MT=3	-0.06	0.03	-0.09	-0.08	-0.05	.30**	-.42**	.30**	0.00	0.13	0.15	-.19*	

Note. Numbers of MT languages refer to Mandarin (1), Malay (2), and Tamil (3).

Table 3

Backward regression model results for English receptive vocabulary

Mandarin	B	SE	β	T	Sig	95% CIs	Semipartial r
(Constant)	18.88	1.30		14.52	0.00	16.33 - 21.44	
PhoAwa	0.21	0.05	0.19	4.34	0.00	0.12 - 0.31	0.15
NonInte	0.36	0.07	0.20	5.13	0.00	0.22 - 0.49	0.18
WorMem	0.26	0.10	0.10	2.52	0.01	0.06 - 0.47	0.09
EngUseLen	0.10	0.03	0.15	3.74	0.00	0.05 - 0.15	0.13
EngBookNum	0.93	0.22	0.17	4.29	0.00	0.50 - 1.35	0.15
WeekHomEng	0.13	0.04	0.13	3.54	0.00	0.06 - 0.20	0.12

Note. R^2=.34, $F(6,544)=46.08$, $p < .001$

Malay	B	SE	β	T	Sig	95% CIs	Semipartial r
(Constant)	23.00	1.84		12.52	0.00	19.35 - 26.64	
PhoAwa	0.39	0.13	0.28	3.05	0.00	0.14 - 0.64	0.25
WorMem	0.80	0.28	0.26	2.88	0.00	0.25 - 1.35	0.24
EngUseLen	0.14	0.05	0.25	2.88	0.00	0.04 - 0.23	0.24
SibNum	1.21	0.59	0.17	2.05	0.04	0.04 - 2.39	0.17

Note. R^2=.31, $F(4,100)=11.31$, $p < .001$

Tamil	B	SE	β	T	Sig	95% CIs	Semipartial r
(Constant)	20.04	1.74		11.49	0.00	16.59 - 23.48	
PhoAwa	0.34	0.07	0.37	5.09	0.00	0.21 - 0.47	0.33
NonInte	0.63	0.14	0.34	4.60	0.00	0.36 - 0.90	0.30

Note. R^2=.37, F(2,146)=42.86, p < .001

Table 4

Backward regression model results for MT receptive vocabulary

Mandarin	B	SE	β	t	Sig	95% CIs	Semipartial
(Constant)	25.37	1.87		13.57	0.00	21.70 - 29.04	
PhoAwa	0.19	0.04	0.18	4.58	0.00	0.11 - 0.27	0.16
NonInte	0.22	0.06	0.13	3.45	0.00	0.09 - 0.34	0.12
MTHomMed	0.07	0.02	0.18	4.49	0.00	0.04 - 0.10	0.16
HomDom	-2.15	0.37	-0.24	-5.84	0.00	-2.87 - -1.43	-0.21
MTUseLen	0.13	0.02	0.22	5.74	0.00	0.09 - 0.18	0.20
SibNum	-0.81	0.35	-0.09	-2.34	0.02	-1.49 - -0.13	-0.08
MTBookNum	0.74	0.26	0.11	2.87	0.00	0.24 - 1.25	0.10

Note. *R^2=.31, F(7,543)=35.29, P<.001*

Malay	B	SE	β	T	Sig	95% CIs	Semipartial r
(Constant)	24.92	1.40		17.77	0.00	22.13 - 27.70	
MTHomMed	0.10	0.04	0.27	2.78	0.01	0.03 - 0.18	0.26
MTUseLen	0.09	0.05	0.19	1.97	0.05	0.00 - 0.18	0.18

Note. *R^2=.13.6, F(2,102)=8.06, p < .001*

Tamil	B	SE	β	T	Sig	95% CIs	Semipartial r
(Constant)	13.60	2.01		6.76	0.00	9.62 - 17.57	
Gender	1.89	0.99	0.13	1.92	0.06	-0.06 - 3.84	0.13
NonInte	0.44	0.11	0.28	3.86	0.00	0.21 - 0.66	0.27
MTHomMed	0.06	0.03	0.18	2.47	0.01	0.01 - 0.11	0.17
MTUseLen	0.10	0.04	0.20	2.75	0.01	0.03 - 0.18	0.19
MTBookNum	2.40	0.50	0.34	4.83	0.00	1.42 - 3.39	0.33
Income	-0.26	0.09	-0.20	-2.74	0.01	-0.44 - -0.07	-0.19

Note. R^2=.33, F(6,142)=11.73, p < .001