Individual differences in very young Chinese children's English vocabulary breadth and semantic depth: Internal and external factors

Abstract

This study examines the English vocabulary development of 43 very young child English as a Foreign Language learners (age 3;2-6;2) in China. They were tested twice for vocabulary breadth (reception and production) and semantic depth (paradigmatic and syntagmatic vocabulary knowledge). The development of the English vocabulary knowledge between these two measurements was predicted using a series of internal factors and external factors.

An exploratory mixed-effects regression analysis revealed that English use, interacting with age of English onset, significantly predicted such growth. Older children benefitted more from practicing English to enhance their English vocabulary. The amount of English input at school was found to positively impact the development of English syntagmatic knowledge. Chinese paradigmatic knowledge significantly influenced the growth of English paradigmatic and syntagmatic knowledge. These findings indicate that at least at an early stage, external factors play an important role in child foreign language (FL) vocabulary development, particularly in children with a later age of English onset. Furthermore, the transfer of concepts from the first language to the second language (L2) might be more pronounced with respect to L2 semantic depth than L2 vocabulary breadth. This conceptual transfer is relevant to FL learners as young as three years of age.

Keywords: internal and external factors, receptive vocabulary, productive vocabulary, paradigmatic knowledge, syntagmatic knowledge, early foreign language learning

1. Introduction

Building vocabulary is one of the most fundamental linguistic developments in early childhood. Both breadth and depth of vocabulary knowledge are important because the former indicates how much children know and the latter indicates how well they know it. Previous studies demonstrated that both internal factors (e.g., short-term phonological memory) and external factors (e.g., input quality) tend to affect the growth of vocabulary knowledge (overview in Unsworth et al. 2014).

Most of the previous studies on child L2 vocabulary knowledge only focused on breadth of knowledge and only a few tapped into the depth aspect (e.g., Proctor et al. 2009). Moreover, many of the target children are in primary school (e.g., as in the studies listed in Rolstad, Mahoney, and Glass 2005) and little is known about whether these results could be applied to very young children. Furthermore, the participants of such studies were usually children acquiring a L2 in a natural setting, where the language exposure is meaning-driven and sufficiently large, and little is known about their peers who are learning a L2 as a foreign language (FL) in an instructional setting. This group may significantly differ from children in a natural setting because their L2 input is much more limited and classroom-based (Copland and Garton 2014). According to Muñoz (2008), typical child FL learners in the instructional settings usually have some or all of the following features: 1) the L2 exposure is limited and the L2 instructional time is approximately 50 minutes per week at school; 2) the L2 exposure in class, both its quantity and quality, is heavily influenced by the teachers´ L2 proficiency and the amount of L2 use in class; 3) children do not use L2 among each other in their daily communication; and 4) the practice and exposure of L2 outside the classroom is limited. Due to the different L2 environment (e.g., differences in input quantity and quality outlined by Muñoz in 2008), the internal and external factors found to be significant in the naturalistic settings could play out differently in instructional settings.

The current study explores the development of both the breadth and semantic depth of the vocabulary knowledge of very young child English as a foreign language (EFL) learners, and examines the impact of internal and external factors on this development. Chinese very young learners of English are chosen as participants because they are a large proportion of all very young EFL learners worldwide (Butler 2013). According to a report by the Chinese department of education, approximately 210 million child English learners are taking English courses in more than 50,000 private English institutes in China (Li 2013).

2. Vocabulary knowledge

As researchers noted that "knowing a word" involves knowing many different aspects, such as phonology, orthography, syntactic constructions and semantic representations (Richards 1976; Nation 1990, 2001; Meara 1996). Vocabulary knowledge should be conceptualized as a continuum ranging from none to complete, with incremental nature (Nagy and Scott 2000). The assessment of vocabulary knowledge should therefore be multifaceted and comprehensive. However, it seems that only vocabulary size (i.e., breadth of vocabulary knowledge) has received sufficient attention in child L2 studies and is considered a key dimension of vocabulary competence (August et al. 2005; Schwartz 2014). It is usually estimated from the number of words children can comprehend and produce. In recent years, the depth of vocabulary knowledge has also drawn researchers' attention because it has been found from some intervention programs that the establishment of deep meaning connections between words facilitate vocabulary growth and reading comprehension (August et al. 2005; Proctor et al. 2009). In general, the breadth and the depth of vocabulary knowledge are positively related (Schmitt and Meara 1997; Rashidi and Khosravi 2010). For instance, in Schmitt's and Meara's (1997) study, Japanese learners' amount of word associations was found to correlate significantly with their vocabulary size, supporting the hypothesis that these two dimensions are interconnected. The current study will examine both breadth and depth and depict a picture about child EFL learners' quantity and quality of vocabulary knowledge.

2.1 Meaning levels of a word: paradigmatic and syntagmatic knowledge

Depth of vocabulary knowledge is a broad construct, including various domains, such as the phonology and orthography of a word, its morphology, its semantic representation, its pragmatics and its etymology (Proctor et al. 2009). Therefore, it is difficult to measure it with a single test or a battery of tests (Schmitt 2014; Read 2004; but see the attempts of Scott et al. 2008; Deane et al. 2014). Traditionally, this construct is broken down into separate elements, using a component approach (Read 2000), and one of its important components for child L2 learners is semantic relations (i.e., meaning levels of a lexicon; Schwartz, Moin, and Leikin 2012). The meaning levels of a word include two fundamental types of knowledge: paradigmatic and syntagmatic (Cruse 1986). The former refers to the knowledge of hierarchical relations (e.g., a rabbit is an animal) and the latter refers to the knowledge of horizontal relations (e.g., a cute grey rabbit). In general, paradigmatic relations include three types of relations: superordination (class-inclusion relations, e.g., fish is a type of animal),

subordination (types of fish, e.g., trout, whale, catfish, shark) and part-whole relations (e.g., a fin is a part of a fish). This type of knowledge reflects children's cognitive ability to conceptualize, categorize and de-contextualize certain objects, and it increases with richer educational experiences (Anglin 1985). In contrast, syntagmatic relations refer to descriptive information of the object's form (e.g., a long fish), color (e.g., a red fish), material, function and location. This level of vocabulary knowledge relates to children's ability to describe and associate the distinctive attributes of an object. Syntagmatic knowledge has been found to develop prior to paradigmatic knowledge (Anglin 1985), and the latter is more likely accumulated while children are getting older (Anglin 1993) and having more instructional experience (Snow 1990). Children's performance on paradigmatic or syntagmatic tasks also depends on their knowledge of the target word in question (Wolter 2001).

2.2 Factors that impact the development of vocabulary knowledge

The following sections will introduce findings on the effects of internal and external factors on very young children's L2 vocabulary development. In line with Clark's (2003) view that both social factors and cognitive status are crucial in early language acquisition, the current study also considers them essential to very young children's early EFL vocabulary development. As discussed in the introduction, most of the existing studies have only focused on the development of vocabulary breadth in naturalistic settings. Therefore, most factors listed below were significant predictors on L2 receptive/productive vocabulary size of immigrant children or L2 learners in immersion programs. The current study extends the scope to the effect of these internal and external factors on vocabulary depth in instructional settings.

Internal factors

Regarding internal factors, time-related factors (e.g., Snedeker 2007, 2012), language aptitude (e.g., Unsworth et al. 2014) and children's first language (L1) (e.g., Paradis 2011) were found to impact the speed and ultimate outcome of children's L2 vocabulary acquisition.

Among the time-related factors, age of onset (AoO) draws most attention. In general, an older AoO has been found to be advantageous for vocabulary knowledge development in both naturalistic and instructional L2 settings in the short run (Muñoz 2014; Miralpeix 2006; García Mayo and García Lecumberri 2003; Chondrogianni and Marinis 2011). Golberg, Paradis and Crago (2008) found that ESL children build their L2 vocabulary faster when the

AoO is later than age five, compared to peers with a younger starting age. This probably indicates the facilitating role of an advanced cognitive ability on vocabulary acquisition (Muñoz 2014). However, most previous studies only examined the breadth aspects of lexical knowledge and hardly explored the effect of AoO on very young foreign language learners' semantic depth.

Language aptitude is another crucial factor that has been found to significantly impact the development of children's L2 vocabulary knowledge (Abrahamsson and Hyltenstam 2008). It is defined as the specific talent for language learning that a learner is assumed to have (Dörnyei and Skehan 2003) and is thought to consist mainly of phonemic coding ability, language analytic ability and verbal memory (Skehan 1986). Some studies have found that short-term memory and analytical reasoning ability are two important components of language analytic ability for child L2 vocabulary acquisition (e.g., Alexiou 2009; Paradis 2011). For instance, Paradis (2011) investigated the English outcome of 169 immigrant children in Canada after 3-62 months following arrival. She found that both good phonological short-term memory and nonverbal intelligence (as the indicator of children's analytical reasoning) predicted the children's receptive vocabulary size with short-term memory being the stronger predictor.

L1 and L2 vocabulary knowledge were positively correlated in a series of studies (e.g., Sparks and Ganschow 1991). Researchers such as Cummins (1979), Sparks and Ganschow (1991, 2001) attributed this connection to the common underlying acquisition mechanism shared by L1 and L2, leading to the moderation of the level of L2 attainment by the level of L1 knowledge. In other words, whatever aptitude enables a learner to acquire good L1 vocabulary knowledge is probably available for the development of L2 vocabulary knowledge as well. Researchers such as Snow and Kim (2007) attributed this positive correlation to the transfer of concepts from L1 to L2 and to higher metalinguistic sophistication. They argued that if vocabulary development is about acquiring new concepts and new phonological forms, then learners with better L1 vocabulary knowledge should benefit from the transfer of concepts from L1 to L2 and could focus more on the new form. As a result, they might learn L2 vocabulary faster than learners who have limited L1 vocabulary knowledge. L2 learners might also hope to express in the L2 what they can express in the L1. The efforts to match their L1 and L2 levels, at least for vocabulary, generate some metalinguistic sophistication and facilitates L2 vocabulary acquisition. For instance, a larger vocabulary size entails a better understanding of polysemy and morphological analysis and could also enable L2 learners to

exploit cognates shared by the two languages better. It is worth noting that such speculative explanations for positive L1-L2 vocabulary relationships mainly apply to those L2 learners who already have a good command of L1 vocabulary knowledge. For instance, the strong positive transfer from L1 to L2 vocabulary breadth found in adult L2 learners was not confirmed by the studies on child L2 learners (e.g., Lindsey, Manis, and Balley 2003; Verhoeven 1994). For very young L2 learners who are developing both their L1 and L2 vocabularies, the correlation between the L1 and L2 vocabulary remains unclear. Because both positive correlations (e.g., Leseman 2000) and negative correlations (e.g., Ordoñez et al. 2002) between L1 and L2 vocabulary have been found, Snow and Kim (2007) argue for more studies on child L2 learners with different AoO, in various learning environments, and investigating different types of vocabulary knowledge. The current paper follows this call and explores the effect of four aspects of Chinese vocabulary knowledge on the L2 vocabulary development of young EFL learners in China.

External factors

A series of studies have found the amount of input to influence the speed of L2 vocabulary acquisition (Thordardottir 2011; David and Wei 2008; Vermeer 2001; Muñoz 2014). The calculation of input quantity is a current topic of debate because the traditional approach of measuring length of exposure (LoE) is unable to distinguish individual differences among L2 speakers (Muñoz 2014). In most cases, one year of L2 exposure for one child is not the same as that for another. Unsworth (2013) called for the attention to the concept of "cumulative length of exposure" brought up by Gutierrez-Clellen and Kreiter (2003) for child bilingual studies and proposed measuring children's L2 input with their daily schedules. By calculating the proportion that each language is used during children's total waking hours, the amount of L2 input over time could be estimated more accurately. The current study adopted this approach to measure the cumulative exposure in the school setting. Weekly exposure to English at home is also calculated (e.g., Chondrogianni and Marinis 2011; Gathercole and Thomas 2005).

Another external factor is input quality, which reflects the authenticity and richness of children's L2 input (Jia and Aaronson 2003). Young SL learners' learning setting may vary with respect to teachers' L2 proficiency (Unsworth et al. 2014), whether the input is from native or non-native speakers (Place and Hoff 2011), a variety of L2 input sources, e.g., English cartoons, readings, games (Uchikoshi 2006), and parental L2 ability (Chondrogianni

and Marinis 2011). All of these aspects have an effect on the acquisition rate and outcome of children's L2 vocabulary knowledge. Uchikoshi (2006) tracked 150 Latino English language learners in preschools over an academic year for their growth rates in vocabulary and found that children who watch TV programs such as *Arthur* and *Between the Lions* at home have an advantage in English vocabulary score at the start of kindergarten over children who do not watch these programs. She argued that how new words were introduced made the difference. For instance, while watching the new words being introduced at the beginning of the show, children might raise questions or repeat the words after the cartoon figure, which could enhance their memory of the words and enlarge their vocabulary size. It is worth noting that children's language input in different settings could be qualitatively different (Cummins 1984) and the current paper explores the English input quantity and quality at school and at home separately.

In recent years, some studies have also highlighted the role of children's L2 use, especially children's L2 output at home, as a significant factor of their L2 vocabulary development (Paradis 2011; Bohman et al. 2010; Sun et al. 2015). According to Swain (2007), output has three main functions in SL learning: 1) a noticing function, 2) a hypothesis-testing function, and 3) a metalinguistic function. In the current case, the noticing function refers to children's attention paid to items that they do not know how to convey in the L2. In other words, language output prompts children to consciously recognize their linguistic limits. The hypothesis-testing function is that child SL learners sometimes take the output as a "trial run" to test their hypothesis of how to pronounce or write what they wish to express. Children might be able to produce the targeted linguistic items correctly by actively seeking feedback through hypothesis testing. The metalinguistic function means that children use language to reflect their own and others' language behaviors, mediating SL learning. Sun et al. (2015) examined 71 preschoolers' English acquisition in China and found that the number of situations of English use significantly predicted children's English receptive vocabulary size and receptive grammar knowledge. They argued that in an environment where L2 input is limited, L2 use could not only help children gain new knowledge, as proposed by Bohman et al. (2010), but also help children to maintain their learning motivation. The increase in knowledge could be attributed to the noticing function of L2 output and the maintenance of motivation might be related to the metalinguistic function of L2 output.

Family socio-economic status (SES) has been found to relate to children's vocabulary outcome in some studies (e.g., Golberg et al. 2008; Paradis 2009). SES is usually estimated

from parental educational levels in such studies. Golberg et al. (2008) followed 19 immigrant children in Edmonton, Canada for two years and found that those whose mothers had post-secondary education had a larger vocabulary size than the others whose mothers had only secondary education. Estimating SES is not unproblematic, however. If it is measured based on maternal education only, children in the current study seem to come from different classes, but if it is measured by Butler's approach (2013), where SES is estimated by parental education, occupations and income together, the majority of the children in the current study are from the middle class. The current study adheres to using maternal educational level as an indicator to family SES, keeping in line with other studies on child L2 vocabulary development (Golberg et al. 2008).

Overall, a series of internal factors, such as AoO and L1 vocabulary knowledge, and a number of external factors, such as input and L2 use, were found to significantly predict L2 vocabulary development. However, most of the related studies focused on child L2 learners in a naturalistic setting and few have looked into young FL learners in an instructional setting. Those who did research the latter group only explored the development of L2 vocabulary breadth and left the semantic depth almost untouched. Furthermore, few such studies adopted a comprehensive view to include both internal and external factors as predictors. The current study will address these limitations by exploring the four aspects of vocabulary knowledge development- receptive vocabulary, productive vocabulary, paradigmatic knowledge and syntagmatic knowledge- of young EFL learners in instructional settings and examine the impact of both internal factors and external factors on the L2 vocabulary development. The following questions are addressed:

- To what extent do children develop their vocabulary skills over seven months during their onset period of instruction?
- To the extent that such development exists, is it related to differences in children's internal and external factors? And what are the best predictors?

This study employs paired *t*-tests to analyze children's L2 vocabulary development and uses mixed-effects models (Willett 1994; Krueger and Tian 2004) to explore significant predictors.

3. Methods

3.1 Participants

A total of 43 (age at the first time of testing: 3.2-6.2, mean=4.5) Chinese children from two branches of Happy English participated in the study. Happy English is one of the largest English language institutes in southeastern China, targeting children 2-12 years old. Most of the students in Happy English are preschoolers and they are taught English using a series of textbooks designed for very young English beginners (*Yippees*: red, green and blue), published by MM publications. The textbooks focus on children's listening and speaking ability first and gradually shift the focus to reading and writing. The Total Physical Response approach (Asher 1996) is used to teach children and maintain their motivation. Each week, children are required to come to Happy English twice for approximately two hours: one time for the main course and the other time for the activity class. The parents are allowed to select more activity classes if they have time. In the main class, a native English teacher and a Chinese teacher work together to teach children new words, phrases and songs following the syllabus. In the activity class, a Chinese teacher works alone to help children review and practice what they learnt during the main class. Child friendly words, such as colors, animals, shapes, food, numbers and transportations, are taught in class. The foreign teachers are required to obtain a TESOL certificate before employment and the Chinese teachers must at least have a bachelor's degree. Most of the Chinese teachers majored in English during college.

There were two selection criteria for the participants. First, they should have had little English instruction and input before the study, and second, they should have no history of language impairment. Most of the children included in the study have been followed since they started their English class in Happy English around September 2012. Two rounds of data collection were conducted. One round was conducted in March 2013 (38 children) and the other in October 2013 (35 children). Of the 43 children who were involved in the current study, only 31 of them have been tested twice. Children missed one of the test rounds when they were ill or because their parents were too busy to bring them to the tests. In each round, there were two sessions of tests and each session lasted for approximately 45 minutes. In one of the sessions, Chinese and short-term working memory were tested, while in another English knowledge and nonverbal intelligence were tested. To avoid a priming effect between L1 and L2 as well as a fatigue effect, the two sessions were separated by a week. In addition,

during the tests, children were asked whether they wanted to have a break every 25 minutes. Most of the children finished the tests without a break.

3.2 Instruments

As mentioned, children's vocabulary knowledge was operationalized as English receptive vocabulary size, English productive vocabulary size, English paradigmatic knowledge and English syntagmatic knowledge (Figure 1). These were measured with four tests that were carried out following standard procedures. Children were individually tested in a quiet room using a computer. Before formal testing, examples were given to make sure that children understood the requirements of each task.

[Figure 1 near here]

English receptive vocabulary size was examined with the Peabody Picture Vocabulary Test (PPVT-4; Dunn and Dunn 2007), which is suitable for children aged from two and a half years old and onwards. Children were shown an array of four pictures and were asked to point out the image that best corresponded with the word they heard. For instance, a child saw an array of four images: a spoon, a fish, a cake and a glass, and was asked to indicate the "spoon". The PPVT has two parallel versions, version A and version B. In each version, there were 204 items in the test. When the child was tested with version A in the first testing round, the child was tested with version B in the second round, and vice-versa.

English productive vocabulary size was measured with the Expressive One-Word Picture Vocabulary Test-4 (EOWPVT-4; Brownell 2010). After seeing a picture, the child was asked to say the word (an object, action or adjective) in English. The test consisted of 196 items in total, and all children took the same test in both testing rounds. If there were practice effects, we assume these affected each participant to the same extent.

English paradigmatic and syntagmatic knowledge were evaluated with word description tasks. Four frequently used nouns (i.e., "car", "watermelon", "dog", and "table") were chosen as stimulus words, and for each of them a child was first asked to describe the subject in great detail (e.g., "What is a dog? Tell me everything you know about a dog"). According to the teachers' interviews, the participants were familiar with the stimulus words in English. To elicit all possible word attributes and semantic dimensions, the child was then asked to answer a series of standard questions related to the super-/subordinate relation (e.g., "which category does a car belong to?", "What types of car do you know?"), distinctive features (e.g., "What

does a car usually look like in terms of size, color and shape?"), part-whole constituents (e.g., "Can you tell me about the different parts of a car?"), function and use (e.g., "What can you do with a car?"), location and time ("Where can you usually find a car?"). The procedure was in line with Vermeer (2001), and Schwartz, Moin and Leikin (2012). The questions were formulated in English first, and were then translated into Chinese to ensure participants' full comprehension. All children's responses were transcribed, coded and scored following the model proposed by Verhallen and Schoonen (1993). English paradigmatic knowledge was examined through answers related to hierarchical taxonomy (e.g., super-/subordinate and part-whole relations). English syntagmatic knowledge was assessed using the descriptive answers about the object (e.g., appearance, taste, use, location and function). One point was given per piece of hierarchical taxonomy or the descriptive information mentioned.

English paradigmatic knowledge was also measured with a semantic fluency task. Food and animals were chosen as the semantic categories, as previous studies have shown their effectiveness in testing young children's verbal fluency (e.g., Schwartz et al. 2012). A child was asked to name as many items as possible in 60 seconds for each semantic category. Each retrieved item was scored with one point.

Because there are no commonly used standard tests for Chinese vocabulary knowledge, we used translated (into Chinese) versions of the PPVT and the other three English tests. Words and sentences were literally translated from English to Chinese. The same testing procedure was followed as for the English tests: children were tested one by one in a quiet room with a computer.

Tests related to language aptitude, which include the tests on phonological short-term memory and nonverbal intelligence, were administered individually on a computer in a quiet room. As with the vocabulary tests, examples were provided to make sure that children understood the requirements of each task before launching the tests. Children's phonological short-term memory was measured with the digit span task and the non-word repetition task, two sub-tests of the Comprehensive Test of Phonological Processing (CTOPP; Wagner, Torgesen, and Rashotte 1999). The tests consist of hearing and having to repeat an increasingly longer (i.e., more difficult) list of digits or non-words in English. The total score was calculated by adding up the scores of the two tasks. When these tasks are conducted in an L2, performance may be lower than in an L1 (Thorn and Gathercole 1999), but given the

feasibility demonstrated in other studies (e.g., Paradis 2011; Unsworth et al. 2014) and the reasonable score range obtained in the current study, this should not affect data analysis.

Nonverbal intelligence was tested to assess the preliterate learners' analytical reasoning ability (Genesee and Hamayan 1980; Paradis 2011). It was measured with the colored version of the Raven's Standard Progressive Matrices (Sets A, B and C; Raven, Court, and Raven 1995). Children were shown pictures with a missing part and they needed to choose a piece from six options to complete the picture. In total, there were 36 items with increasing complexity. In the current study, all of the children were at the age of fast cognitive development; therefore, all of the internal factors were tested twice to reflect such growth. It is worth noting that a small portion of the children were under the youngest age suggested by the aptitude tests, however, all of them seemed to enjoy doing the task and no child demonstrated uneasiness or made complains even in the most difficult session. This probably indicates that the tasks were challenging but manageable.

A parental questionnaire was used to estimate children's English input and use outside of Happy English as well as to assess other background information. The questionnaire was designed based on the language background questionnaire of the Early Language and Intercultural Acquisition Studies (ELIAS) and the Utrecht Bilingual Language Exposure Calculator (UBiLEC; Unsworth 2013), two effective questionnaires widely used in large-scale studies on young foreign language learners in Europe. In contrast to these questionnaires, the questionnaire used in the current study more deeply explored English media use at home. Information about the children's age of English onset, parental education (highest degree), parents' self-reported English fluency (five-point scale), weekly amount of English input in the family-setting (mostly from media), number of sources of English input at home per week, weekly frequency of English media use at home, and English use in general (number of settings in which oral English was used regularly) was collected with the questionnaire (see Appendix I).

Happy English has an electronic system to track each learner's class attendance and performance. These online school records were used to calculate children's overall input in the instructional setting. As Unsworth (2013) argued, LoE is too general to estimate children's actual amount of language exposure and therefore the current study investigated children's exact English class attendance through the online attendance system of Happy English. By doing so, we obtained a more precise indication of the accumulated input in the instructional

setting for each child. Moreover, the teaching time of both the foreign teachers and the Chinese teachers were also registered. By using this information, we could estimate the proportion of native English input of the total instruction time. The same approach was not used to calculate the cumulative input at home. This was because the English exposure at home in an EFL setting was neither stable in quantity nor similar in quality over months.

3.3 Statistical analysis

First, a descriptive analysis was conducted including all predictors and the four aspects of English vocabulary knowledge. Subsequently, a correlation analysis was used to assess the relationships between the predictors. To avoid multicollinearity, when two predictors correlated highly (r >.8), one of the two predictors was excluded. Age at testing time was highly correlated with age of onset (r = .93) and weakly but significantly correlated with total amount of school input (r = .38), so it was not included into the final dataset. Next, the two rounds of tests were compared using paired *t*-tests to see whether the performance on the tests increased significantly, independently of any predictors. Finally, linear mixed-effects regression modeling (using the lme4 package in R) was used to assess which predictors significantly influenced English vocabulary development. The number of the testing round was included as a predictor in this model to investigate the development of the outcome variable over time.

Compared to a traditional approach such as ANOVA, mixed-effects regression is more flexible as it is well suited to analyze an imbalanced dataset or a dataset containing missing values (Baayen 2008 Ch. 7; Jaeger 2008). Given that each participant provides multiple test scores, the variability associated with the participant needs to be taken into account when analyzing the data. In mixed-effects regression, a distinction is made between fixed-effect factors and random-effect factors. Fixed-effect factors generally have a small number of levels (e.g., gender) that are exhausted in the sample and would be the same when the experiment would be replicated (i.e., in both experiments both boys and girls would participate). The only random-effect factor in our study is a participant. Blom, Paradis and Duncan (2012) explain random effects as follows; “the effect of the participating children is typically considered a random-effect variable. This is because the sample of children is drawn

from a larger population, and each participating child has unknown properties that will influence the measurements. Hence, the effect of participating children cannot be measured without error and if a new study is undertaken, other participants will be included, with again unknown properties" (p. 978). Some children may perform better than others, and these differences are modeled using so-called random intercepts. The influence of fixed-effect predictors may vary per subject as well. For instance, while most children will improve from the first to the second test round, the improvement may be smaller for one child than for another. This variability is captured by including so-called random slopes (in this case, a by-subject random slope for the influence of testing round). By including both random intercepts and slopes, type-I errors are prevented (Baayen 2008). We assessed whether random slopes and intercepts were necessary by comparing the Akaike Information Criterion (AIC; Akaike 1974) of the two models. A lower AIC indicates that the more complex model is warranted due to a better fit (offset against the additional complexity). We only included (fixed or random-effect) predictors when the AIC reduction was at least 2.

The dependent variable in our model was the test score of the participants. As each participant contributed at most eight test scores (four tests and two testing rounds), the number of observations lies around 280 (not all participants participated in all testing rounds). More specifically, we fitted a single model including all test scores of the participants. The test scores for each test were standardized (i.e., transformed to have a mean of 0 and standard deviation of 1) to ensure comparability. Two predictors were used to indicate the test type and the testing round. Note that even though not all children participated in both testing rounds, their data is still used in the model to more reliably assess the influence of various subject-related predictors such as age of onset.

In our exploratory analysis, we fitted a model step-by-step, removing predictors that did not contribute significantly to the model fit. We assessed whether interactions were significant, and specifically focused on interactions including the test type (to statistically assess if there were differences in how the various predictors affected the test scores on the four tests). After obtaining a final model, we ascertained that the residuals of our model followed a normal distribution.

4. Results

4.1 Descriptive statistics and t-tests

The descriptive statistics of the internal factors, external factors and the English scores of all 43 children are summarized in Appendix II. Out of the whole group, 31 children participated in both test rounds. A few of them had one or two missing scores due to missing test records. One child missed the English vocabulary tests and nonverbal intelligence test in the second round due to illness. The descriptive statistics of this group is shown in Table 1. In terms of internal factors, the paired *t*-tests revealed that all of them (i.e., including nonverbal intelligence, phonological short-term memory, and the four aspects of Chinese vocabulary knowledge) improved significantly over the course of seven months (i.e., the time between the two test rounds). In terms of external factors, children's English input quantity and quality was quite limited. In addition to two hours of English class at Happy English, children only received approximately one and a half hours of English exposure at home from fewer than two English sources. Most of the parents choose either English movies or a CD (included with the textbook) as the material to use at home. This type of media was used less than four times per week. The places where English was used regularly were likewise quite limited. Most children only used English at Happy English or when reviewing the English words at home. Although most parents had a bachelor degree, many of them were not confident enough in using their English. The majority believed that they could only communicate with simple words and basic sentences, and therefore barely used any English with their children at home. They worried that their English accent would mislead the children.

[Table 1 near here]

4.2 The development of English vocabulary in seven months

English vocabulary skill developed significantly (see Table 2). Paired *t*-tests based on children who took both rounds of English tests demonstrated that all four aspects of English vocabulary knowledge improved significantly over a period of seven months, though the increase is small.

[Table 2 near here]

4.3 Prediction of English vocabulary performance by internal and external factors

Table 3 shows our best-fitting mixed-effects regression model obtained in an exploratory analysis. The whole model (including a random intercept for word and a by-subject random slope for the influence of the Chinese productive vocabulary) explained 65% of the variance of the test scores (across all four tests, and two testing rounds), out of which 40.1% was

attributable to the fixed-effects only. The interpretation of the model is as follows. First, there is a significant interaction between age of English onset and the use of English. Age of English onset is not significant for the average use of English for the participants in our study (i.e. line 2 of the table). For the average age of onset (the centered value equals 0), the effect of the use of English positively and significantly affects English performance (across all four tests; line 3). The fourth line of the table indicates that for a higher age of onset, the effect of the total use of English is even more beneficial (i.e. resulting in higher test scores). Therefore, the beneficial effect of practicing English was greater for the older children in our group (ranging from 31 to 68 months; see Table 1). Lines 5 to 8 show that the total amount of English input only significantly influences English syntagmatic knowledge. In addition, while the *t*-tests in section 4.2 above revealed that all aspects of English vocabulary knowledge increased significantly over time when not taking into account the influence of any predictors, the mixed-effects regression model (lines 9 to 12) shows that when taking into account all of the significant predictors mentioned above, the improvement due to time (i.e., the first versus the second round of testing, seven months later) is only significant for English comprehension. Of course, the passing of time is also reflected in the improvement in Chinese proficiency across the two testing rounds (see Table 1). Line 13 to 16 of Table 3 indeed show that the Chinese paradigmatic knowledge significantly influences English paradigmatic and syntagmatic performance. There is a close-to significant (p = .08) influence on English productive vocabulary, but no influence on English comprehension (for which the testing round is a significant predictor).

[Table 3 near here]

4.4 Power analysis

Given our relatively small sample size, and large number of predictors (18), we conducted a power analysis using the R package. However, as our data points are grouped per subject, a power analysis appropriate for linear regression will underestimate the sample size needed for our design. To correct the sample size (for a given power) we have multiplied this with the so-called design effect, which indicates how the longitudinal design influences the standard errors (Snijders 2005). By multiplying the required sample size with the design effect, we obtain a better estimate of required sample size for a given power. Since our sample size was already known (281 values, associated with 43 subjects), we determined the power for small, medium and large effects according to the values suggested in the documentation of the pwr

package (Champely 2015). The results indicated only a power of 0.085 (i.e. 8.5% chance to detect a true effect) for detecting small effects, 47% chance for detecting medium effects and 90.8% chance to detect large effects. In sum, our study is only reliably able to detect large effects.

5. Discussion

The current study intended to explore two questions. First, we investigated how the FL vocabulary knowledge of very young FL learners developed over a relatively short period of time and, second, if such development is significant, what internal and external factors contribute to the growth. Paired *t*-tests revealed that children's four aspects of English vocabulary knowledge – i.e., English productive vocabulary size, English receptive vocabulary size, English paradigmatic knowledge and English syntagmatic knowledge – have all developed significantly over seven months. The focus of the discussion then shifts to the factors that could significantly predict the growth.

5.1 Age of English onset and the use of English

Age of English onset itself did not significantly predict performance on any aspect of English vocabulary knowledge. However, its interaction with English use had a positive impact on all four aspects of English vocabulary knowledge. For children with an average age of English onset, the use of English has been found to significantly influence the development of all four aspects of English vocabulary. The amount of English use was estimated from the number of settings where the participants could practice English regularly. This result is in line with the previous findings on child SL learners in both naturalistic settings (Bohman et al. 2010; Paradis 2011) and instructional settings (Muñoz 2011). The interaction between age of English onset and English use further showed that older children experienced a greater beneficial effect of English practice for improving their L2 vocabulary knowledge. They might be more cognitively mature than younger children and therefore better capable of extracting information during L2 use (Miralpeix 2006).

In the present study, all of the three functions of L2 output hypothesized by Swain (2007) – the noticing function, the hypothesis-testing function, and the metalinguistic function – are likely to contribute to children's EFL development during the onset period of learning English in the current study. Language use as part of human psychological processes originates from collective behaviors (e.g., discourse) and facilitates the internalization of language knowledge

as part of people's mental activity (Stetsenko and Arievitch 1997; Swain 2007). Collaborative conversations (e.g., talking to a native English speaker on vacation abroad) enabled the child EFL learners to participate in problem solving and engaged a knowledge building process, and, in the end, helped them to build up knowledge about the language. During such conversations, older children might be quicker to sense their language limitations and raise questions to mediate language problems. Moreover, they might be better at guessing meanings, forming responses to the questions and searching for answers together with the other interlocutors. These communication skills, which are probably related to a higher cognitive maturity, might facilitate the speed of building up L2 knowledge.

5.2 Total amount of English input at school

The total amount of English input in the instructional settings significantly predicts English syntagmatic knowledge, but not the other types of knowledge. This seems contradictory to the previous findings on the predictive power of input quantity for breadth of vocabulary knowledge (e.g., Unsworth et al. 2014; Paradis 2011). However, given the nature of the input and the length of English exposure the children had in the current study, the finding is probably realistic (Riegel 1968). Child EFL learners in this study learned simple words and songs in class, and most of the content was related to the attributes of an object, such as colors, shapes and size. The description tasks adopted to test the children's syntagmatic knowledge could sensitively capture the differences caused by class instructions. The other types of English vocabulary knowledge, such as productive vocabulary size, might take a longer time and more intensive in-class exposure for the beginners to build up. The result reminds us of the importance of choosing the type of tests when exploring child EFL learners' early FL development. The standard tests, such as PPVT and EOWPV, could be more appropriate to use with children ESL learners in the naturalistic settings or with child EFL learners with a longer learning history in class. For the beginners in the instructional settings, depth of vocabulary tasks, which are more tailored to their learning materials, could be more effective in capturing children's variation.

5.3 Chinese paradigmatic knowledge

Chinese paradigmatic knowledge was found to be positively related to the children's depth of English vocabulary knowledge. Paradigmatic knowledge is closely related to children's ability to conceptualize and categorize certain objects. Once children have a good organization of vocabulary knowledge in their L1, they might transfer the concepts of the L1

words to the L2 and pay more attention to the new phonological forms. With the help of concept transfer, child EFL learners with better Chinese paradigmatic knowledge could show an advantage at the speed of L2 vocabulary acquisition. A period of seven months already enabled this factor to show effects on English depth vocabulary knowledge during children's onset of English learning. This result could be a piece of the jigsaw puzzle that Snow and Kim (2007) proposed. They wondered whether for very young child L2 learners who are developing both languages, the relationship between the L1 and the L2 vocabulary knowledge is positive or negative and to what extent this relationship is shaped by children's age and environmental factors. The current study found that for child FL learners in the instructional settings, at least at the beginning years of L2 learning, a positive relationship might be found between the depth of vocabulary knowledge of the two languages (Riegel 1968). The paradigmatic knowledge in the L1 could promote the development of the L2 vocabulary depth knowledge, even when the exposure to the L2 is limited, L2 practice is rare and the children are very young.

It is worth noting that, apart from the significant internal and external factors mentioned above, time still positively influenced the development of English receptive vocabulary size. This might be related to factors that have not been covered in the present study, for instance, teachers' English language proficiency (Unsworth et al. 2014). Studies in the future might take this factor into consideration.

5.4 Insignificant factors

External factors related to the quality of English at school (native English ratio) and the quantity and quality of English at home (e.g., weekly English input at home and mother's English proficiency level) have not significantly predicted any aspect of English vocabulary knowledge. This seems to contradict some findings in similar contexts. For instance, Sun et al. (2015) found that both school input and home environment significantly predicted receptive and productive vocabulary knowledge. However, in the current study, only the development of syntagmatic knowledge has been found to be significantly predicted by school input. The differences with the previous findings might be due to the small sample size of the current study (e.g., 43 participants in the current study vs. 71 in Sun et al.'s study). It might also be related to children's limited length of English learning in the current study (e.g., 38 hours at first round and 79 hours at the second round vs. 125 hours in Sun et al.'s study). The limited amount of English input in general might restrain the effects of quantity on vocabulary

development. In terms of input quality at home, children's language proficiency might not match the level of the English materials. For instance, despite many hours of movie input, some children in the current study still performed below the average. According to the parental interview, these children might show an interest in the English movies at the beginning, but they would soon get frustrated by the difficult language. Merely watching the vivid sceneries could contribute little to children's English development. Parents also played a marginal role in terms of English exposure and use at home. Some parents would like to practice English with their children; however, they were too busy to use English regularly with their children. More parents were reluctant to use English with their children out of concern over their own basic English skills. They worried that their own English accent would affect their children's pronunciation.

Internal factors, such as nonverbal intelligence and phonological short-term memory, also did not significantly predict the development of English vocabulary knowledge. One reason could be the small participant population and the limited times of testing. Another reason could be children's specific environment and limited L2 proficiency in the current study. According to Sun et al. (2015), the influence of internal factors might be restrained in the EFL settings, where both language input and output are scarce. As the total amount of input increases with time, these internal factors might gain in significance.

6. Limitations

The present study has several limitations. First of all, teachers' English proficiency was not included as a predictor. Previous studies indicate that teachers' qualification and language use could influence EFL children's language acquisition (Unsworth et al. 2014; Bowers and Vasilyeva 2011). The different amounts of vocabulary richness and grammatical complexity provided by the teachers might affect the language development in general. Due to the frequent change of English teachers in the private language institute, the authors were unable to test all teachers for their English proficiency, not to mention record them on video and analyze their class instruction. Studies in the future could focus on the very beginning of the project and only track the children with the same teachers over a longer period of time to overcome the measurement difficulties resulting from the variety of teachers, a typical characteristic of these settings.

Secondly, more participants should be recruited and be tested for more rounds to verify the findings. The current study only included 43 participants and a larger sample size is

needed in future research, which would allow us to more reliably detect small to medium effects. More rounds of tests on the same population would provide us with detailed information about the consistency and potentially dynamic role of internal and external factors on learning outcomes at different stages of FL learning.

Thirdly, data collection regarding the home English environment could be improved. A language diary approach (De Houwer and Bornstein 2003) could be used to track children's daily schedule for several weeks. These documents would comprehensively demonstrate children's language input in different situations and therefore provide us with more insight into the impact of various input factors over language development.

Next, in order to avoid multicollinearity, age at time of testing was not included in the final dataset because of its significant correlation with age of onset ($r = .93$) and total amount of school input ($r = .38$). This might affect the interpretability of age of onset.

Moreover, due to the lack of common Chinese vocabulary tests, the authors had to translate the English tests (e.g., PPVT) to Chinese. The majority of the English words and sentences used in the tests have good equivalents in Chinese. However, some words, such as squash and knight, are not used as frequently in Chinese as in English, therefore, these words might affect children's vocabulary performance.

Finally, the choice of measurements on children's verbal short term memory and analytical reasoning might have been more appropriate for the population. For instance, Ravens, which was used to measure children's analytical reasoning ability in the current study, might be replaced by the Wechsler Non Verbal Scale of Ability (Wechsler and Naglieri 2006) since the latter is specifically developed for children from culturally and linguistically diverse backgrounds other than English speaking countries and for those with relatively low language abilities.

7. **Conclusion and implications**

The current paper investigated child EFL learners' English vocabulary development in China and explored the significant predictors for such development. English vocabulary knowledge has been operationalized as receptive vocabulary size, productive vocabulary size, paradigmatic vocabulary knowledge, and syntagmatic vocabulary knowledge. The first two categories fall into realm of breadth of vocabulary knowledge, which are considered crucial and have been widely studies in child L2 development. The other two categories, which

belong to the depth of vocabulary knowledge, however, haven't drawn much attention until recent years. They have been found to effectively facilitate vocabulary growth and reading comprehension (August et al. 2005; Proctor et al. 2009). The current study is one of the few studies that examined both breadth and depth of vocabulary knowledge and therefore provided a more comprehensive picture of children's early foreign language vocabulary development.

Over seven months, both the breadth and the semantic depth of English vocabulary knowledge increased slightly but significantly. Children with later onset age profited more from English practice to develop their English vocabulary knowledge, and this is the case for all four aspects of English vocabulary knowledge. It is possible that children with better cognitive maturity could more easily notice their language limits during L2 practice and elicit more constructive feedback in conversations. Total amount of English input at school was found to positively influence the growth of English syntagmatic knowledge. Word description tasks, which have been used to test syntagmatic knowledge, could accurately reflect the contents of classroom instruction. Chinese paradigmatic knowledge significantly predicted English depth vocabulary knowledge. The transfer of word concept from Chinese to English might promote children's English vocabulary building process.

The results indicate the crucial role that external factors play in child FL acquisition in instructional settings. Both L2 input and output could significantly affect the acquisition of vocabulary, one of the most important language aspects. Parents should create more opportunities for their children to practice the L2. Outside school, children might use the L2 regularly at home, at the English corner at the university campus and during vacation. Practice could not only make language features more transparent for acquisition but also enable the children to realize their language limits. However, it is worth noting that such benefits of language practice could be more enjoyed by child FL learners with an older age of onset. Passionate parents believing in "the younger, the better" for SL learning should be informed that this belief might not be true in FL settings. The L2 input at home should also be paid attention to. Parents should select the materials that match their children's proficiency level. Once the input is too complicated or too long, children might get frustrated or get bored and soon lose the interest in these materials. Rhymes, songs, little stories, activities with a ritualistic character, were found appropriate input for child L2 acquisition (Sokolov and Snow 1994). These contents could be displayed through interactive media, such as on a computer or iPad. Some free apps on that relate to child FL learning, such as LatelyLily, Vehicle Book and

You-Things, were reported to be favored by the children in the current study. Future research could pay more attention to child FL learners' use of such media and track their impact on L2 vocabulary development.

References

Abrahamsson, Niclas, and Kenneth Hyltenstam. 2008. The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition* 30 (4): 481-509.

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

Akaike, H. 1974. A new look at the statistical identification model. *IEEE Transactions on Automatic Control* 19(6): 716–723.

Anglin, J. M. 1985. The child's expressible knowledge of word concepts. In *Children's language.*, ed. K. E. Nelson. Vol. 5, 77-127. Hillsdale, NJ: Erlbaum.

Anglin, J. M. 1993. Vocabulary development - a morphological analysis. *Monographs of the Society for Research in Child Development* 58 (10): 1-166.

Asher, J. 1996. *Learning another language through actions*. 5th ed. Los Gatos, CA: Sky Oaks Productions.

August, D., M. Carlo, C. Dressler, and C. Snow. 2005. The critical role of vocabulary development for English language learners. *Learning Disabilities Research & Practice* 20 : 50-7.

Baayen, R. H. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Blom, Elma, Johanne Paradis, and Tamara Sorenson Duncan. 2012. Effects of input properties, vocabulary size, and L1 on the development of third person singular 's in child L2 English. *Language Learning* 62 (3): 965-94.

Bohman, T. M., L. M. Bedore, E. D. Pena, A. Mendez-Perez, and R. B. Gillam. 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

Bowers, Edmond, and Marina Vasilyeva. 2011. The relation between teacher input and lexical growth of preschoolers. *Applied Psycholinguistics* 32 (1) (JAN 2011): 221-41.

Brownell, R. 2010. *Expressive one-word picture vocabulary test* . San Antonio: Pearson.

Butler, Y. G. 2013. Parental factors and early English education as a foreign language: A case study in mainland china.. *Research Papers in Education*.

Champely, S. 2015. *Pwr: Basic functions for power analysis. R package version 1.1.3.* https://cran.r-project.org/web/packages/pwr/index.html

Chondrogianni, V., and T. Marinis. 2011. Differential effects of internal and external factors on the development of vocabulary, morphology and complex syntax in successive bilingual children. *Linguistic Approaches to Bilingualism* 1 (3): 223-48.

Clark, E.V. 2003. *First language acquisition.* Cambridge: Cambridge University Press.

Copland, Fiona, and Sue Garton. 2014. Key themes and future directions in teaching English to young learners: Introduction to the special issue. *ELT Journal* 68 (3) (July 01): 223-30.

Cruse, D. A. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

Cummins, J. 1979. Linguistic interdependence and the educational development of bilingual children *Review of Educational Research* 49 : 222-51.

Cummins, J. 1984. *Bilingualism and special education*. Clevedon, UK: Multilingual Matters.

David, A., and L. Wei. 2008. Individual differences in the lexical development of French-English bilingual children. *International Journal of Bilingual Education and Bilingualism* 11 : 598-618.

Deane, P., R. R. Lawless, C. Li, J. Sabatini, I. I. Bejar, and T., O'Reilly. 2014. Creating vocabulary item types that measure students' depth of semantic knowledge. *ETS Research Report Series* 2014(1): 1-19. doi:10.1002/ets2.12001

De Houwer, A., and M. Bornstein. 2003. Balancing on the tightrope: Language use patterns in bilingual families with young children. Paper presented at Presented at the Fourth International Symposium on Bilingualism, Tempe, Arizona, April 30 to May 3.

Dörnyei, Z., and P. Skehan. 2003. Individual differences in second language learning. In *The handbook of second language acquisition.*, eds. C. Doughty, M. Long, 589-630. Oxford: Blackwell.

Dunn, L. M., and D. M. Dunn. 2007. *Peabody picture vocabulary test*. 4th ed. Minneapolis: Minneapolis: Pearson.

García Mayo, M. P., & M. L. García Lecumberri. 2003. *Age and the acquisition of English as a foreign language: Theoretical issues and field work.* Clevedon: Multilingual Matters.

Gathercole, V. C. M., and E. M. Thomas. 2005. Minority language survival: Input factors influencing the acquisition of welsh. Paper presented at *Proceedings of the 4th international symposium on bilingualism* Somerville, MA: Cascadilla.

Genesee, Fred, and Else Hamayan. 1980. Individual differences in second language learning. *Applied Psycholinguistics* 1 (1): 95-110.

Golberg, Heather, Johanne Paradis, and Martha Crago. 2008. Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics* 29 (1): 41-65.

Gutierrez-clellen, V. F., and J. Kreiter. 2003. Understanding child bilingual acquisition using parent and teacher reports. *Applied Psycholinguistics* 24: 267-88.

Jaeger, T. F. 2008. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language* 59 (4): 434-46.

Jia, G. and D, Aaronson. 2003. A longitudinal study of Chinese children and adolescents learning English in the united states. *Applied Psycholinguistics* 24(1): 131-161. doi:10.1017/S0142716403000079

Krueger, Charlene, and Lili Tian. 2004. A comparison of the general linear mixed model and repeated measures ANOVA using a dataset with multiple missing data points. *Biological Research for Nursing* 6 (2): 151-7.

Leseman, P. M. 2000. Bilingual vocabulary development of Turkish preschoolers in the Netherlands. *Journal of Multilingual and Multicultural Development*(21): 93-112.

Li, W. 2013. When should a child start to learn English? the approaches to get interested in a foreign language. *Sanlian Life Weekly*(44): 46-82.

Lindsey, KA, FR Manis, and CE Bailey. 2003. Prediction of first-grade reading in Spanish-speaking English-language learners. *Journal of Educational Psychology* 95 (3): 482-94.

Meara, P. 1996. The dimensions of lexical competence. In *Performance and competence in second language acquisition.*, eds. G. Brown, K. Malmkjær and J. Williams, 35-53. Cambridge: Cambridge University Press.

Miralpeix, I. 2006. Age and vocabulary acquisition in EFL. In *Age and Foreign Language Learning Rate: The BAF Project*., eds. C. Muñoz, 89-106. Clevedon: Multilingual Matters.

Muñoz, C. 2008. Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics* 29 (4): 578-96.

Muñoz, C. 2011. Input and long-term effects of starting age in foreign language learning. *International Review of Applied Linguistics in Language Teaching* 49 (2): 113-33.

Munoz, C. 2014. Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics* 35 (4): 463-82.

Nagy, W. E. & J. A. Scott. 2000. Vocabulary Processes. In *Handbook of Reading Research, Volume III.,* eds. M. Kamil, P. Mosenthal, P. D. Pearson, and R. Barr, 269-284. Muhwah, N.J.: Lawrence Erlbaum Associates.

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

Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.

Ordoñez, C. L., M. S. Carlo, C. E. Snow, and B. McLaughlin. 2002. Depth and breadth of vocabulary in two languages: Which vocabulary skills transfer? *Journal of Educational Psychology* 94 (4): 719-28.

Paradis, J. 2009. *Oral language development in French and English and the role of home input factors.* (Report for the Conseil scolaire centre-nord[north-central school board]). Edmonton, Alberta.

Paradis, J. 2011. Individual differences in child English second language acquisition: Comparing child-internal and child-external factors. *Linguistic Approaches to Bilingualism* 1: 213-37.

Place, Silvia, and Erika Hoff. 2011. Properties of dual language exposure that influence 2-year-olds' bilingual proficiency. *Child Development* 82 (6): 1834-49.

Proctor, C. P., P. Uccelli, B. Dalton, and C. E. Snow. 2009. Understanding depth of vocabulary online with bilingual and monolingual children. *Reading & Writing Quarterly* 25(4): 311-33.

Rashidi, N., and N. Khosravi. 2010. Assessing the role of depth and breadth of vocabulary knowledge in reading comprehension of Iranian EFL learners. *Journal of Pan-Pacific Association of Applied Linguistics* 14 (1): 81-108.

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

Read, J. 2000. Assessing vocabulary. Cambridge, UK: Cambridge University Press.

Read, J. 2004. Plumbing the depths: How should the construct of vocabulary knowledge be defined? In *Vocabulary in a second language: Selection, acquisition and testing.* eds. P. Bogaards, B. Laufer, 209–227. Amsterdam: John Benjamins.

Richards, J. C. 1976. The role of vocabulary teaching. *TESOL Quarterly* 10(1): 77-89.

Riegel, K. F. 1968. Some theoretical considerations of bilingual development. *Psychological Bulletin* 70(6): 647-670. doi:10.1037/h0026821

Rolstad, K., K. Mahoney, and G. V. Glass. 2005. The big picture: A meta-analysis of program effectiveness research on English language learners. *Educational Policy* 19 (4): 572-94.

Schmitt, N. 2014. Size and depth of vocabulary knowledge: What the research shows. *Language Learning* 64(4): 913-951. doi:10.1111/lang.12077

Schmitt, N., and P. Meara. 1997. Researching vocabulary through word knowledge framework: Word associations and verbal suffixes. *Studies in Second Language Acquisition*(19): 17-36.

Schwartz, M. 2014. The impact of the first language first model on vocabulary development among preschool bilingual children. *Reading and Writing* 27(4): 709-732. doi:10.1007/s11145-013-9463-2

Schwartz, Mila, Victor Moin, and Mark Leikin. 2012. Lexical knowledge development in the first and second languages among language-minority children: The role of bilingual versus monolingual preschool education. *International Journal of Bilingual Education and Bilingualism* 15(5) (2012): 549-71.

Scott, J. A., M. Hoover, S. L. Flinspach, and J. L. Vevea. 2008. A multiple-level vocabulary assessment tool: Measuring word knowledge based on grade level materials. In *57th annual yearbook of the National Reading Conference*, eds. Y. Kim, V. J. Risko, D. L. Compton, D. K. Dickinson, M. K. Hundles, R. T. Jimenes, K. M. Leandor, D. W. Rowe, 325–340. Oak Creek, WI: National Reading Conference.

Skehan, P. 1986. Cluster analysis and the identification of learner types. In *Experimental approaches to second language learning.*, ed. V. Cook, 81-94. Oxford: Pergamon Press.

Snedeker, Jesse, Joy Geren, and Carissa L. Shafto. 2012. Disentangling the effects of cognitive development and linguistic expertise: A longitudinal study of the acquisition of english in internationally-adopted children. Cognitive Psychology 65 (1) (AUG 2012): 39-76.

Snedeker, Jesse, Joy Geren, and Carissa L. Shafto. 2007. Different Paths: Changes in Second-Language Acquisition between Three and Five Years of Age. Unpublished material. https://software.rc.fas.harvard.edu/lds/wp-content/uploads/2012/04/Snedeker_Geren_Shafto_Paths.pdf.

Snijders, T. A. 2005. Power and sample size in multilevel linear models. *Encyclopedia of statistics in behavioral science.*

Snow, C. E. 1990. The development of definition skill. *Journal of Child Language* 17: 697-710.

Snow, Catherine E., and Young-Suk Kim. 2007. Large problem spaces: The challenge of vocabulary for English language learners. In Vocabulary Acquisition: Implications for Reading Comprehension, eds. R. K. Wagner, A. E. Muse and K. R. Tannenbaum, 123-139. New York, NY, US: Guilford Press.

Sokolov, Jeffrey, and Catherine Snow. 1994. The changing role of negative evidence in theories of language development. In Input and interaction in language acquisition, ed. Clare Gallaway and Brian J.Richards., 38-55. Cambridge: Cambridge University Press, http://dx.doi.org/10.1017/CBO9780511620690.004 .

Sparks, R., and L. Ganschow. 1991. Foreign language learning difficulties: Affective or native language aptitude differences? *Modern Language Journal*(75): 3-16.

Sparks, R. L., and L. Ganschow. 2001. Aptitude for learning a foreign language. *Annual Review of Applied Linguistics*(21): 90-111.

Stetsenko, A., and I. Arievitch. 1997. Constructing and deconstructing the self: Comparing post-Vygotskian and discourse-based versions of social constructivism. *Mind, Culture, and Activity*(4): 159-72.

Swain, M. 2007. *The output hypothesis: Its history and its future.* Beijing, China: Beijing Foreign Studies University, .

Sun, H., R. Steinkrauss, J. Tendeiro, and K. de Bot. 2015. Individual differences in very young children's English acquisition in China: Internal and external factors. *Bilingualism: Language and Cognition.* doi: http://dx.doi.org/10.1017/S1366728915000243

Thordardottir, Elin. 2011. The relationship between bilingual exposure and vocabulary development. *International Journal of Bilingualism* 15 (4): 426-45.

Thorn, A. S. C., and S. E. Gathercole. 1999. Language-specific knowledge and short-term memory in bilingual and non-bilingual children. *Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology* 52 (2): 303-24.

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

Unsworth, S., L. Persson, T. Prins, and K. de Bot. 2014. An investigation of factors affecting early foreign language learning in the Netherlands. *Applied Linguistics 1-23.*

Unsworth, Sharon. 2013. Assessing the role of current and CUMULATIVE exposure in simultaneous bilingual acquisition: The case of Dutch gender. *Bilingualism-Language and Cognition* 16 (1): 86-110.

Verhallen, M., and R. Schoonen. 1993. Lexical knowledge of monolingual and bilingual children. *Applied Linguistics* 14 (4): 344-6.

Verhoeven, L. 1994. Transfer in bilingual development - the linguistic interdependence hypothesis revisited. *Language Learning* 44(3): 381-415. doi:10.1111/j.1467-1770.1994.tb01112.x

Vermeer, A. 2001. Breadth and depth of vocabulary in relation to L-1/L-2 acquisition and frequency of input. *Applied Psycholinguistics* 22 (2): 217-34.

Wagner, R., J. Torgesen, and C. Rashotte. 1999. *Comprehensive test of phonological processing*. Austin, TX: Pro-Ed.

Wechsler, D., and J.A. Naglieri. 2006. *Wechsler Nonverbal Scale of Ability (WNV)*. San Antonio, TX: Harcourt Assessment.

Willett, J. 1994. Measuring change more effectively by modeling individual growth over time. In *The international encyclopedia of education.*, eds. T. Husen, T. Postlehwaite. Oxford (UK) ed. Pergamon.

Wolter, B. 2001. Comparing the L1 and L2 mental lexicons: A depth of individual word knowledge model . *Studies in Second Language Acquisition* (23): 41-70.

Appendix I

Language Exposure Questionnaire

Part 1 Background

Child's name	Test number	Gender
Place of birth	Date of birth	Date of testing
Mother's name	Mother's occupation	Mother's highest education
Father's name	Father's occupation	Father's highest education
Sibling's name	Sibling's age	

Part 2 English learning and usage

1. From whom and where did your child first come into contact with English?
2. From when did your child receive consistent and significant exposure to English?
3. Think about the English exposure of your kid each normal week at home (and at kindergarten if applicable)

Weekly exposure (in minutes)	English TV programs made in China	English TV programs made in English speaking countries	English videos(e.g. movie " Three pigs")	English audios (e.g. Action Verses)	English electronic devices (e.g. smart-pen reader)	English books	English games	Kindergarten/ School
Mon.								
Tue.								
Wed.								
Thur.								
Fri.								

Sat.								
Sun.								
Total (in minutes)								
Materials								

4. How long does your child use English with you in a normal week (in minutes) at home?

Mother: ____ Father: ____ Other relatives____

5. Think about the oral English proficiency of the following members at home

Mother: (Poor 1 – Fair 2 – Average 3 – Good 4 – Excellen5)

Father: (Poor 1 – Fair 2 – Average 3 – Good 4 – Excellen5)

Other relatives : (Poor 1 – Fair 2 – Average 3 – Good 4 – Excellen5)

6. Has any Chinese teacher taught your child English besides Happy English and kindergarten/ school? If so, how long was it? (in minutes)

7. Has any native English speaker taught your child English besides Happy English and kindergarten/ school? If so, how long was it? (in minutes)

8. How many times does your child review English words learnt at Happy English in general?

9. Where does your child use English regularly?

a. Happy English b.at home c. on vacation

d. At kindergarten/school e. English stage performance

f. Other places, such as ___

10. What's your biggest concern so far to help your child learn English?

a. Lack of appropriate materials

b. Lack of good teaching approach and skills

c. Lack of confidence in my oral English

d. Lack of time due to being busy

e. I don't want to give my child too much burden

f. I believe that good English acquisition relies on language aptitude

g. I believe that good English acquisition relies on English teachers

h. If you have other concerns, please write them down:_____

Appendix II

Means and standard deviations of variables for all the 43 children

	Time 1			Time 2		
Internal factors	**Mean(SD)1**	**N1**	**Range1**	**Mean(SD)2**	**N2**	**Range2**
AoO	46.86(10.43)	43	30-68			
NonInte	15.51(4.94)	39	6-25	19.83(5.47)	35	9-33
ShoMem	13.00(4.17)	39	2-21	15.11(3.45)	36	7-22
Chi.Pro	56.68(15.82)	38	29-87	70.00(14.35)	36	42-99
Chi.Com	93.47(35.65)	38	39-156	116.42(37.94)	36	57-186
Chi.Para	28.18(10.41)	39	12-49	41.31(16.63)	35	15-83
Chi.Syn	28.28(10.37)	39	8-46	41.33(12.48)	36	12-63
External factors	**Mean(SD)1**	**N1**	**Range1**	**Mean(SD)2**	**N2**	**Range2**
SchInput	37.63(12.20)	43	13-66	83.13(27.36)	36	43.33-160
HomInput	1.58(1.32)	39	0-5.33	1.65(1.82)	31	0-9.5
HomFreq	3.77(2.69)	39	0-9	3.72(4.03)	32	0-18
HomSource	1.87(1.34)	39	0-6	1.50(1.03)	36	0-4
Native	0.43(0.17)	43	0.14-0.81	0.47(0.10)	34	0.25-0.7
EngUse	2.26(1.02)	39	1-5	2.47(1.11)	36	1-5
LoE	6.10 (1.49)	31	4-11			
MotEng	2.53(0.93)	43	1-5			
FatEng	2.50(0.97)	42	1-4			
MotEdu	2.84(1.02)	43	1-5			
FatEdu	2.88(1.18)	43	0-5			
Outcome variables	**Mean(SD)1**	**N1**	**Range1**	**Mean(SD)2**	**N2**	**Range2**
Eng.Pro	7.50(3.31)	38	1-14	11.83(4.03)	35	4-23
Eng.Com	14.08(6.16)	38	5-31	21.44(7.96)	34	10-43
Eng.Para	9.13(4.91)	38	0-21	13.34(7.20)	32	1-32
Eng.Syn	6.61(4.07)	38	0-17	12.23(6.21)	31	0-24

Note. **AoO**=age of onset in months; **NonInte**= nonverbal intelligence; **ShoMem**= phonological short-term memory; **Chi.Pro**= Chinese productive vocabulary size; **Chi.Com**= Chinese receptive vocabulary size; **Chi.Para**= Chinese paradigmatic knowledge; **Chi.Syn**= Chinese syntagmatic knowledge; **SchInput** = total amount of English input at Happy English and bilingual kindergartens in hours; **HomInput**=weekly English input quantity at home in hours; **HomFreq**=times of using English media at home per week; **HomSource**=number of different English media format used at home per week; **Native**= the percentage of native English input out of the total English input at school; **EngUse**=number of places of using English in total; **LoE** = Length of exposure to English in months; **MotEng** = mother's self-rated proficiency in English; **FatEng** = father's self-rated proficiency in English; **MotEdu** = mother's highest educational level; **FatEdu** = father's highest educational level; **Eng.Pro**= English productive vocabulary size; **Eng.Com**= English receptive vocabulary size; **Eng.Para**= English paradigmatic knowledge; **Eng.Syn**= English syntagmatic knowledge

Tables

Table 1. *Means and standard deviations of internal and external factors for the 31 children who took two rounds of tests*

	Time 1			Time 2			
Internal factors	**Mean (SD)1**	**N1**	**Range1**	**Mean (SD)2**	**N2**	**Range2**	***t***
AoO	46.68 (10.23)	31	31-68				
NonInte	15.42 (4.53)	31	6-23	20.17 (5.66)	30	9-33	7.19***
ShoMem	12.84 (4.24)	31	2-20	15.23 (3.32)	31	8-22	3.00**
Chi.Pro	56.13 (16.86)	31	29-87	71.16 (13.14)	31	42-93	9.6***
Chi.Com	94.52 (37.43)	31	39-156	119.61 (36.91)	31	66-186	4.80***
Chi.Para	27.90 (10.62)	31	12-49	42.35 (16.92)	31	15-83	7.31***
Chi.Syn	28.90 (10.77)	31	8-46	42.26 (12.17)	31	12-63	5.87***
External factors	**Mean (SD)1**	**N1**	**Range1**	**Mean (SD)2**	**N2**	**Range2**	***t***
SchInput	38.36 (10.96)	31	20.67-66	78.77 (21.45)	30	43.33-129	-0.58
HomInput	1.54 (1.30)	31	0-5	1.52 (1.89)	27	0-10	-0.53
HomFreq	3.77 (2.66)	31	0-9	3.57 (4.10)	28	0-18	0.35
HomSource	1.81 (1.25)	31	0-6	1.39 (1.02)	31	0-4	1.89
Native	.43 (.17)	31	0.14-0.73	.49 (.09)	29	0.3-0.7	-1.96
EngUse	2.29 (1.04)	31	1-5	2.32 (1.11)	31	1-5	-0.16
LoE	6.10 (1.49)	31	4-11				
MotEng	2.42 (.77)	31	1-4				
FatEng	2.52 (.96)	31	1-4				
MotEdu	2.81 (.91)	31	1-5				
FatEdu	2.94 (1.18)	31	0-5				

Note. AoO: age of English onset in months; **NonInte**: nonverbal intelligence as a measure of analytic reasoning; **ShoMem**: phonological short-term memory based on digit span and non-word repetition; **Chi.Pro**: Chinese productive vocabulary size based on Chinese EOWPV test; **Chi.Com**: Chinese receptive vocabulary size based on translated version of PPVT test; **Chi.Para**: Chinese paradigmatic knowledge based on the word description task and verbal fluency task; **Chi.Syn**: Chinese syntagmatic knowledge based on the word description task; **SchInput**: total amount of English input at Happy English and bilingual kindergartens in hours; **HomInput**: weekly English input quantity at home in hours; **HomFreq**: frequency of using English media at home per week; **HomSource**: number of different English media formats used at home per week; **Native**: the percentage of native English input out of the total English input at school; **EngUse**: number of places of using English in total; **LoE**: length of exposure to English in months; **MotEng**: mother's self-rated proficiency in English on a 5-point scale; **FatEng**: father's self-rated proficiency in English on a 5-point scale; **MotEdu**: mother's highest educational level on a 6-point (0-5) scale; **FatEdu**: father's highest educational level on a 6-point (0-5) scale; **p<.005, ***p<.001.

Table 2. *English outcomes for the 31 children in the two testing rounds seven months apart.*

	Time 1			Time 2			
Outcome variables	**Mean (SD)1**	**N1**	**Range1**	**Mean (SD)2**	**N2**	**Range2**	***t***
Eng.Pro	7.65 (3.54)	31	1-14	11.53 (3.75)	30	4-20	5.17***
Eng.Com	14.74 (6.08)	31	7-31	20.55 (7.11)	29	10-35	4.12***
Eng.Para	9.58 (5.01)	31	0-21	13.67 (6.43)	27	1-32	2.63*
Eng.Syn	7.13 (4.18)	31	0-17	11.88 (5.74)	26	0-24	3,54**

Note. Eng.Pro: English productive vocabulary size based on EOWPV test; **Eng.Com**: English receptive vocabulary size based on PPVT test; **Eng.Para**: English paradigmatic knowledge based on the word description task and verbal fluency task; **Eng.Syn**: English syntagmatic knowledge based on the word description task; *p<.05, **p<.005, ***p<.001.

Table 3. *Fixed-effects part of the model*

Fixed effects	Estimate	Std. Error	*t value*
1. (Intercept)	-1.32	0.17	-7.85
2. AoO	0.01	0.01	1.06
3. EngUse	**0.12**	**0.06**	**2.05**
4. AoO:EngUse	**0.02**	**0.01**	**2.25**
5. SchInput:Eng.Para	0.00	0.00	-0.30
6. SchInput:Eng.Syn	**0.01**	**0.01**	**2.92**
7. SchInput:Eng.Pro	0.01	0.00	1.42
8. SchInput:Eng.Com	0.00	0.01	-0.86
9. Time:Eng.Para	0.22	0.27	0.80
10. Time:Eng.Syn	-0.28	0.28	-1.02
11. Time:Eng.Pro	0.29	0.26	1.13
12. Time:Eng.Com	**1.17**	**0.26**	**4.52**
13. Chi.Para:Eng.Para	**0.03**	**0.01**	**3.58**
14. Chi.Para:Eng.Syn	**0.03**	**0.01**	**2.85**
15. Chi.Para:Eng.Pro	0.02	0.01	1.78
16. Chi.Para:Eng.Com	-0.01	0.01	-0.74

Note. Time:Eng.Para: the impact of testing round on English paradigmatic knowledge (*z*-transformed); **Time:Eng.Syn**: the impact of testing round on English syntagmatic knowledge (*z*-transformed); **Time:Eng.Pro**: the impact of round on English productive vocabulary size (*z*-transformed); **Time:Eng.Com**: the impact of round on English receptive vocabulary size (*z*-transformed); **Chi.Para:** the impact of Chinese paradigmatic knowledge; **SchInput:** the impact of total amount of English input; **AoO**: the impact of age of onset (centered); **EngUse**: the impact of output (centered); **AoO:EngUse**: the impact of the interaction of age of onset and English use; $|t|>2$, $p < .05$; $|t| > 1.65$, $p < 0.1$