Abstract

This study examined the impact of access to word pronunciations on assessing Chinese learners' vocabulary knowledge. Chinese heritage learners (HLs) and foreign language learners (FLs) studying in American universities were administered a computer-based test in which they were to first select a picture that represented the meaning of a target word presented in characters, and then answer the same item presented in both characters and pinyin. Pinyin access substantially increased the test reliability for both groups of learners. The no-pinyin and pinyin conditions were less distinct for FLs than HLs. In the no-pinyin condition, the groups showed no significant score difference, whereas in the pinyin condition, HLs significantly outperformed FLs. HLs' proportion of cases who successfully corrected the original choices was also notably higher. These findings suggested that pinyin availability differentially impacted the two groups. Implications for vocabulary knowledge assessment for different types of Chinese learners are discussed.

Keywords: Foreign Language; Heritage Language; Chinese; Vocabulary Knowledge; Pinyin

Introduction

If a Chinese language teacher is asked what s/he usually teaches about a word, the answer would likely include the word's pronunciation, recognizing and writing its component characters, and the meaning of the word. This answer is aligned with a principle that usually characterizes Chinese language curriculum and instruction, that is, the integration of 音 yīn (sound), 形 xíng (orthography), and 义 yì (meaning) for word teaching and learning (Li, D. 2003; Li, R. and Wu 2005). It is also entailed in what vocabulary researchers often define for what it means to know a word, such as Nation (2001), who contends that vocabulary knowledge covers the form (sound and orthography), meaning, and use of words in receptive and productive forms and in oral/aural and written modalities. Yet, the answer's highlight on the integration of *yīn*, *xíng*, and *yì* does not appear to be as straightforward for assessing vocabulary knowledge as for teaching it in Chinese.

In the literature on vocabulary assessment, there appeared to be 'an unspoken assumption that if a word is known, then it is likely to be known in both written and aural forms' (Milton 2009: 93). This assumption seems to characterize classroom-based formal instruction in a foreign language context where instruction has simultaneous focuses on sound (e.g., using International Phonetic Alphabet for learning English words), orthography (e.g., spelling), and meaning (e.g., using learners' first language or L1 translations) through print materials presented in textbooks. Possibly as a result of this assumption, vocabulary knowledge is often assessed in the written form with little attention to and discussion on the distinction between phonological and orthographic vocabulary or oral and written vocabulary (Milton 2009). For example, commonly-used tests for assessing L2 vocabulary (breadth) knowledge like the Vocabulary Levels Test (Nation 1990; Schmitt, Schmitt, and Clapham 2001) and Yes/No tests (e.g., Harrington and Carey 2009; Stubbe 2012) ask learners to demonstrate their knowledge of a word by matching its

written form with a given meaning definition or simply indicating whether they 'know' the word. Consequently, these tests address 'only the learner's knowledge of orthographic form rather than a broader construct … which includes the phonological form' (Milton and Hopkins 2006: 130).

The lack of attention to a distinction between oral and written modality in vocabulary assessment would be unjustifiable, however, when attention is given to how language is often learned in different ways, that is, variation in learning contexts or learner backgrounds/profiles (Milton 2009). In a monolingual context, children have acquired a lot of word meanings through substantial oral language experiences in early childhood; yet, without instruction on how to read those words (e.g., using the alphabetic principle for phonological recoding in English), they would not be able to access the meanings of those words in print. Thus, oral versus written vocabulary is usually an important distinction; and young children's vocabulary knowledge is often assessed in oral/aural form through an interview or a test which asks them to select a picture to represent the meaning of a word heard (i.e., the Peabody Picture Vocabulary Test or PPVT; Dunn and Dunn 2007).

In the domain of learning an additional language, a distinction often made between second and foreign language context also helps explain the importance of distinguishing between oral and written vocabulary (Milton 2009). The former is often understood as a context where the target language is the societal language, such as learning English in the United States and learning Chinese in China, and the latter a context where learning is typically through textbook-based instruction in a classroom setting. In a second language context, learners could have substantial experiences of being exposed to and using the target language in aural/oral form for communicative purposes, whereas the opportunity (and need) is rare if not non-existent in a foreign language context. As a result, some words could have been acquired primarily in

oral/aural form and accessed as such in a second language context as opposed to a foreign language context. Consequently, a written test of vocabulary knowledge presumably would not capture the full vocabulary repertoire of learners whose lexical/language acquisition happens primarily in the former context; and bias in vocabulary knowledge assessment might be induced due to variation in learning contexts or learner profiles.

The aforementioned issue seems to be particularly relevant and salient for non-alphabetic languages like Chinese (and Japanese), which uses a logographic writing system. Alphabetic languages, such as English and French, follow the rule of phoneme-to-letter correspondence and allow for the use of alphabetic principle to decode print words and access their meanings (Rayner et al. 2002).[1] On the other hand, Chinese characters are square-shape symbols composed of strokes and stroke patterns (DeFrancis 1984; Taylor and Taylor 2014). The most common 3500 characters yield about 70,000 single-character and multi-character (mostly two-character) words in Chinese (Zhang, K. 1997). Thus, to access the meaning of a written Chinese word, each component character of the word needs to be recognized without the kind of immediate phonological clues available in words in alphabetic languages.[2] Thus, testing Chinese vocabulary knowledge in the written form (i.e., characters) without considering variation in learning contexts or learner backgrounds/profiles would be limiting.

Previous studies have shown a very high correlation between being able to name a word and knowing its meaning among foreign language learners of Chinese (i.e., FLs). Everson (1998), for example, asked first-year university learners in the United States who had no background in Chinese to name 48 words that they had learned and were presented in characters on the computer. They were then asked to write the English meanings of those words presented on a piece of paper. The number of words correctly named and defined appeared comparable; the

naming and meaning scores also showed a very high correlation of about .96. In addition, it was also estimated that when participants were able to pronounce a word correctly, there was a mean probability of 90.7% that they were able to 'identify' or give the meaning of the word. Similar findings were later reported in Jiang (2003) and Zhao (2003), too. The very close relationship between knowledge of word pronunciations and meanings among FLs, from an assessment perspective, seems to indicate that whether or not there is access to character/word sounds may not matter for a vocabulary test presented in characters for those learners.

On the other hand, sound access may be an important consideration for those learners whose lexical acquisition involves significant exposure to and use of Chinese in the oral/aural modality. A notable group of those learners is heritage language learners (HLs) (He 2008). In contrast to FLs, HLs grow up in a Chinese-speaking family and usually have much more experiences in the spoken (through communication with family members and members of extended families) than the written modality (e.g., formal literacy instruction in a weekend community school). Li, D. (2003) asked non-HLs and HLs to write pinyin and make multi-character words for given characters and write characters for given pinyin syllables. Learners' responses were then used to estimate the number of characters for which they knew the sound, meaning, and orthography, respectively. It was found that HLs demonstrated a notably larger gap between knowledge of character sounds and character forms than did non-HLs. This finding seems to suggest that a written Chinese vocabulary test without sound access may underestimate HLs' vocabulary knowledge more substantially than FLs'.

Research questions

Little if any research has examined the issue of character/word pronunciations from the perspective of assessing the vocabulary knowledge of Chinese learners from different

backgrounds. This study, thus, aimed to address this gap. Specifically, it addressed the following two questions:

1. Does access to word pronunciations have an impact on the assessment of vocabulary knowledge for Chinese learners?

2. Does the impact if any differ between HLs and FLs?

Method

Participants

The participants were 53 undergraduate students studying Chinese in American universities. They included 31 HLs (19 females and 12 males with an average of 19.81 years) and 22 FLs (13 females and 9 males with an average age of 20.41years). All of them indicated that they had learned pinyin systematically and used it regularly to pronounce characters/words in Chinese. Among the HLs, all reported a Chinese-speaking mother, and 19 (about 61.3%) reported that both parents spoke Chinese. All but one indicated English as their strongest language; about half ($N = 16$) also indicated some proficiency in one or more languages other than English and Chinese, such as French, German, Spanish, and Korean. The only exception was a student who indicated Indonesian as his/her strongest language. At the time of this study, most of them ($N = 23$; about 74.2%) were enrolled in an intermediate Chinese course (200 level), seven (about 30.4%) in an advanced course (300 level), and 1 in a 400 level course entitled Topics in Chinese Culture.[3]

Among the non-HLs / FLs, 18 (about 81.8%) indicated English as their strongest language; two had Korean as their strongest language, and two others indicated Cambodian and Spanish as their strongest language, respectively. At the time of the study, 13 were enrolled in an intermediate Chinese course, 8 in an advanced course, and 1 enrolled in a 400 level course

(course title unknown). None of them indicated any long-term (defined as half a year and longer) experience of studying abroad in a Chinese-speaking country.

As show in Table 2 (see the Results section), the learners' self-assessments of their Chinese skills on a 5-point scale (1 for very low and 5 for near native) indicated that the two groups did not differ significantly in their overall proficiency in Chinese, $t(51) = .687, p = .495$. Among the four language skills, only in listening were the HLs stronger than the FLs, $t(51) = 2.790, p = .007$.

Instruments

The participants worked on the computer and first completed the background questionnaire which elicited the aforementioned background information. They then worked on a radical choice task and a picture selection that measured their orthographic processing skill and vocabulary breadth knowledge task, respectively.

The radical choice task asked the learners to indicate an orthographic component (i.e., semantic radical), from four choices, that represents the meaning of its host character. For example, 脚 (foot) is followed by 去, 月, 卩, and 却, with 月 being the answer, as it indicates characters related to body parts. One point was awarded for each correct choice, with a total score of 10.

The picture selection task included five single-character and 20 two-character words of various frequency levels based on the Modern Chinese Frequency Dictionary (Beijing Language Institute 1986). Each word was followed by four pictures. The participants were asked to circle one picture that best represented the meaning of a target word. Each word had two appearances. In its first appearance, the word was presented in character only (i.e., character-only or no-pinyin condition); after learners selected a picture for the word, they moved on to the second appearance

of the same word which was presented in both character and pinyin (syllable as well as tone) (i.e., pinyin condition) (see Appendix for an example). The picture selection task was set up in such a way on the computer that the learners had to indicate an answer for the first appearance of a target word before they could move on to the word's second appearance where they could either retain their original choice or revise it (with access to the pronunciation of the word through the pinyin provided). The learners could not go backward to revise any answer. One point was awarded for each correct picture selection, with the total score for each appearance being 25.

Results

Test reliability

We first examined how pinyin availability may have an impact on the internal consistency of the vocabulary knowledge task for the two groups of learners. It was found that in the character-only condition (i.e., first appearance), the reliability tended to be low for both the HLs (Cronbach's α = .592) and the FLs (α = .566), although HLs' appeared to be slightly higher than the FLs'. In the condition where pinyin was presented (i.e., second appearance), the reliability substantially increased for both groups, α = .782 and = .793 for the HLs and the FLs, respectively. Thus, from the perspective of test reliability, having pinyin presented together with characters seems to be a preferred design for assessing Chinese vocabulary knowledge, disregarding learners' backgrounds.

Correlations between tasks

As show in Table 1, for both groups of learners, the correlation between the performance in the non-pinyin and the pinyin condition was significant; however, the correlation for the HLs (r = .673, p < .001) was notably lower than that for the FLs (r = .881, p < .001), suggesting that the response pattern for the two conditions was more convergent for the FLs than for the HLs. In

other words, the availability of pinyin made the assessment of Chinese vocabulary knowledge more distinct from when target words were only presented in characters for the HLs than the FLs. The stronger distinction between the two conditions for the HLs also was also confirmed by the finding that their vocabulary knowledge was significantly correlated with orthographic processing skill in the no-pinyin condition ($r = .467, p < 01$) as opposed to the pinyin condition.

Insert Table 1 about here

However, such a pattern did not surface for the FLs, for whom neither correlation was significant, $r = .077$ ($p = .732$) and $r = .131$ ($p = .562$) for the non-pinyin and the pinyin condition, respectively. The lack of a significant correlation, particularly in the non-pinyin condition, seemed to be a surprise, as a close relationship between orthographic processing and word recognition has been identified in the literature among foreign language learners of Chinese (e.g., Ke 1998). Instead of arguing that for the FLs in this study, orthographic processing was not involved in recognizing the characters/words in the picture selection task, it is speculated that the result might be related to the relatively low reliability of the radical choice task for the FLs (Cronbach's $\alpha = .575$, in comparison to .734 for the HLs).

Vocabulary score comparisons

Table 2 shows the means and standard deviations of the learners' performance on all the tasks (and their self-assessments of language skills). A two-way mixed ANOVA with pinyin availability and group membership as independent variables was performed on the learners' vocabulary knowledge scores. A significant main effect of pinyin availability was found, $F(1, 51)$ = 44.000, $p < .001$, partial $\eta^2 = .463$, indicating that averaged across groups of learners, the

9

presence of pinyin (M = 14.85, SE = .64) resulted in significantly better vocabulary scores than when it was not available (M = 12.09, SE = .49). No significant main effect of group, however, was found, $F(1, 51)$ = .634, p = .429, partial η^2 = .012, which suggests no significant difference between the two groups disregarding the factor of pinyin availability, M = 13.89, SE = .68 and M = 13.05, SE = .81 for the HLs and the FLs, respectively. On the other hand, the interaction effect was found to be significant, $F(1, 51)$ = 11.282, p < .001, partial η^2 = .181. Specifically, when there was no pinyin, the FLs' scores (M = 12.36, SE = .75) were slightly higher than the HLs' (M = 11.81, SE = .63), but the difference did not achieve significance; on the other hand, when pinyin was available, the HLs' scores (M = 15.97, SE = .82) were significantly higher than the FLs' (M = 13.73, SE = .822) (see Figure 1). This interaction effect indicated that the presence of pinyin boosted the vocabulary scores more substantially for the HLs than for the FLs.

Insert Table 2 about here

Insert Figure 1 about here

Proportion of participants revising choices

To obtain additional evidence for the impact of pinyin availability on the assessment of vocabulary knowledge, item analysis was conducted with a focus on the proportion of learners in each group who successfully revised their original choice to get to the answer for each vocabulary item. For the 25 items, an average of 6.7 HLs (range 1-16) (6.7 out of 31 or about 21.6%) revised their scores successfully when pinyin was provided, which was notably higher

than an average of 2.7 FLs (range 0-7) (about 2.7 out of 22 or about 12.3%) who successfully

corrected their choices. There was also no significant correlation between the two groups on the

proportion of participants with successful revision (for the 25 vocabulary items), Spearman's ρ

= .147, p = .484. Taken together, these findings of item analysis confirmed that pinyin

availability affected the HLs and the FLs differentially for the assessment of their vocabulary

knowledge.

Discussion

Impact of pinyin

To answer the first research question, pinyin availability did have an impact on the assessment of

vocabulary knowledge for Chinese learners. The presence of pinyin (together with characters)

not only substantially increased the test reliability but also resulted in significantly better

vocabulary test scores, which was true for both the HLs and the FLs. In both groups, almost all

vocabulary items witnessed successful correction of original choices. These findings suggest that

giving learners' access to the pronunciation of characters/words through the provision of pinyin

may be a better test design for assessing Chinese learners' vocabulary knowledge, disregarding

learners' backgrounds. Such a design, which implies integrated assessment (i.e., sound,

orthography, and meaning), would also seem to make vocabulary assessment better aligned with

Chinese curricula where there are typically concurrent focuses on all three aspects of a word. In

other words, having characters presented together with pinyin may better discriminate the

vocabulary knowledge of Chinese learners and improve the content and construct validity of a

written vocabulary knowledge test.

Differential impact on HLs and FLs

To answer the second research question, it was found that the impact of pinyin also differed between the HLs and the FLs. To begin with, there was a strong correlation between the non-pinyin and pinyin conditions for the FLs ($r = .881$), which tended to corroborate previous research findings about very strong correlations between Chinese learners' ability to pronounce, or write pinyin for, given characters/words and their knowledge of character/word meanings (e.g., Everson 1998; Jiang 2003; Zhao 2003). On the other hand, the correlation between the two conditions was only moderate in size ($r = .673$) for the HLs. These findings suggested that whether or not pinyin or pronunciations were available made much less difference to a written Chinese vocabulary test for FLs than for HLs.

Over and beyond the correlations between the two conditions themselves, how the vocabulary knowledge demonstrated in the two conditions correlated with orthographic processing skill also varied between the two groups. Specifically, for the HLs, but not the FLs, a significant correlation was found when the target words were presented in characters only as opposed to when pinyin was also presented. This seems to suggest that HLs were more actively involved in using phonological information than orthographic information for accessing the meanings of target words when pinyin was available.

Pinyin presence was also found to bring significantly more benefits to the HLs than the FLs. This was reflected not only in the significant interaction effect revealed by the ANOVA of the learners' vocabulary scores, but also the substantially larger proportion of the HLs than the FLs who successfully corrected their choices with the presence of pinyin. Considering that the HLs and the FLs were enrolled in similar levels of Chinese courses at their respective universities and gave similar self-ratings of their Chinese proficiency, these findings appeared to be attributable to the variation in the two groups' lexical/language experiences in the oral/aural

modality. Specifically, the HLs had substantial aural/oral experiences with Chinese, and likely more so than their experiences with print Chinese. Presumably their receptive, oral vocabulary would be larger than their written vocabulary. Thus, when pinyin was present to provide phonological information of the target words, they would be able to access the meanings of those words that would otherwise not be accessed due to the difficulty to recognize the characters. In contrast, the FLs learned Chinese primarily through textbook-based, classroom instruction in a context where there was little opportunity or need to use the target language in aural/oral form outside of the classroom. Consequently, the concurrent instructional focuses on character/word sound and orthography (and meaning) for the FLs did not seem to result in as great an imbalance between oral and written vocabulary as among their HL counterparts; and adding pinyin, while being helpful (with respect to test reliability and score, as discussed earlier), did not result in as significant a benefit for them as for the HLs.

Implications

As Milton (2009) pointed out, L2 vocabulary knowledge tests are largely designed as written tests and administered as such with little attention to the distinction between the written and aural modalities due possibly to an assumption that if one can recognize a print word (with its meaning accessed), s/he would know the sound of the word. This assumption was perhaps a result of the traditional focus of language teaching (and assessment) for foreign language learners for whom word sound and orthography are often acquired simultaneously through classroom instruction. At the beginning of the paper, it was also added that the assumption might also be a result of the assessment practice that had been heavily established on the basis of testing practices for alphabetic languages; and an argument was thus made for the importance of considering sound access for assessing vocabulary knowledge in non-alphabetic languages like Chinese.

The findings of this study, as discussed above, clearly supported that argument. Given that it may be logistically difficult to have word pronunciations recorded and played to learners, which was perhaps another reason why vocabulary tests are often administered in written form (Milton 2009), adding pinyin would be a very convenient and viable choice for a more accurate estimate of Chinese learners' vocabulary knowledge. The suggestion also seems to be very important to L2 Chinese researchers who for various research purposes may need to incorporate a measure to assess learners' vocabulary knowledge (e.g., Cai 2015; Zhang, H. 2016). In such cases, presenting pinyin together with characters, as opposed to characters only, seems particularly necessary in that it could make the measure more validly discriminate learners' vocabulary knowledge. In a situation where there is a need to document individual differences in vocabulary knowledge and use them to predict language skills development, such as in the case of Qian (1999) for predicting reading comprehension (for English learners), the predictive effect might be underestimated if the test is administered in print with target vocabulary items presented only in characters.

Limitations and future research

A few limitations of this study are noted. To begin with, the impact of sound/pinyin access on vocabulary assessment was explored and some suggestions proposed for assessment practices based on the examination of HLs in comparison to FLs. While the findings should be relevant to other types of Chinese learners, such as those who learn in a context where there is substantial opportunity and need for using Chinese in aural/oral form (e.g., learning Chinese primarily through studying abroad in China), direct research on those learners is needed in the future. Another issue that was not addressed but warrants attention in future research pertains to word frequency. Milton (2009) noted that target word frequency could moderate the relative strength

of (English) learners' orthographic and phonological vocabulary at different stages of learning. Specifically, 'frequent vocabulary is proportionately more frequent in speech than in writing, given the development of the phonological side of the lexicon an advantage at the outset of learning. Beyond this stage, less frequent vocabulary is much more accessible in written form, given the orthographic side of the lexicon an advantage at the higher levels of knowledge' (Milton 2009: 99). Thus, there is a possibility that the (differential levels of) benefit of the availability of pinyin for HLs and FLs might disappear when target words are very low in frequency and the learners are highly proficient. In this study, the picture selection task only had a small number of words (25 in total) and the sample size in each group was also small, which did not allow for a reliable examination of possible moderation induced by word frequency and learner proficiency.

In addition, Milton and colleagues (Milton 2009; Milton and Hopkins 2006) focused on 'true' aural vocabulary with recordings of target words played to learners, whereas the focus of the present study was on pinyin, a written phonetic coding system to pronounce Chinese characters/words. Thus, disregarding the convenience of sound access through pinyin, it remains to be tested whether audio recordings and pinyin would make any difference in assessing the vocabulary knowledge of different types of Chinese learners. Logically, because of HLs' much more substantial aural/oral experiences with Chinese, audio-recordings would seem to provide those learners with a greater advantage than FLs, and might further enhance the reliability of and performance on the test in HLs. Finally, a related issue is that all students in the present study learned pinyin systematically in the Chinese program(s) they had been or were enrolled, and it was thus assumed that nobody would have any difficulty in accessing the pronunciations of characters/words through pinyin, which might not necessarily be the case. What's more, the

findings, which were based on pinyin, also remain to be confirmed or disconfirmed for other coding systems that help learners pronounce Chinese characters (e.g., zhuyin fuhao).

Conclusions

This study examined the implications of access to the pronunciations of target words for assessing Chinese vocabulary knowledge for different groups of learners. It was found that the presence of pinyin (together with characters) substantially increased the test reliability for both the HLs and the FLs. Whether or not pinyin was available had a less distinct effect for the FLs than for the HLs. Specifically, when pinyin was not present, there was no significant difference between the two groups of learners, whereas when pinyin was presented together with characters, the HLs' performance was significantly better than the FLs'. In addition, there was a notably larger proportion of the HLs who successfully corrected their choices with the phonological information from pinyin. Based on these findings, it was suggested that Chinese vocabulary knowledge tests, if for logistic considerations needing to be administered in written form, consider having pinyin presented together with characters. It is hoped that the findings and suggestions could shed light on vocabulary assessment for Chinese learners (and L2 vocabulary assessment in general), and help L2 Chinese researchers make better decisions when there is a need to incorporate a measure to assess learners' vocabulary knowledge.

Notes

1. This might also be a reason why little attention has been given to the distinction between written and oral vocabulary in the vocabulary assessment literature in that there might be an assumption that (adult) second language (as opposed to foreign language) learners can access the sound (and thus the meaning) of a word 'easily' through phonological recoding. Thus, a written test would thus not tend to underestimate learners' meaning repertoire.

2. There are at least three methods to help Chinese learners (and Chinese-speaking children) to learn the sound of a character. The first method, or a 'look-and-say method', is characterized by learners simply 'imitating' the sound given by the teacher. The other two methods involve the use of a Romanization system, such as zhuyin fuhao (or bopomofo) and hanyu pinyin (hereafter, pinyin). The former is commonly used in Taiwan, and the latter in mainland China. Pinyin is also widely used in Chinese language curriculums and teaching outside of China. For example, in the United States, commonly used Chinese textbook series for university learners, such as Integrated Chinese (Liu et al. 2011) and Chinese Link (Wu et al. 2010), all begin with familiarizing learners with the pinyin system, and subsequently, pinyin is provided, together with English translations, for new characters/words on vocabulary lists.

3. University-based Chinese language programs, like programs of other foreign languages, in the United States usually define the levels of their course offerings in numbers to show a course sequence. Typically, 100-level (e.g., 101 and 102) courses are for beginning-level learners (i.e., Elementary Chinese); 200-level courses are for intermediate-level learners (i.e., Intermediate Chinese); 300-level courses are for advanced-level learners (i.e., Advanced Chinese); and 400-level courses are usually special topics courses on Chinese language, culture, and society.

References

Beijing Language Institute. 1986. *Modern Chinese Frequency Dictionary*. Beijing: Beijing Language Institute Press.

Cai, W. 2015. Investigating the roles of vocabulary knowledge and word recognition speed in Chinese language listening. *Chinese as Second Language Research* 4, no. 1: 47-65.

DeFrancis, J. 1984. *The Chinese Language: Fact and Fantasy*. Honolulu, HI: University of Hawaii Press.

Dunn, L.M. and D.M. Dunn, D. M. 2007. *Peabody Picture Vocabulary Test* (4th ed.). Minneapolis, MN: NCS Pearson.

Everson, M.E. 1998. Word recognition among learners of Chinese as a Foreign Language: investigating the relationships between naming and knowing. *The Modern Language Journal* 82, no. 2: 194-204.

Harrington, M. and M. Carey. 2009. The on-line yes/no test as a placement tool. *System* 37: 614-626.

He, A.W. 2008. Chinese as a heritage language: an introduction. In *Chinese as a Heritage Language: Fostering Rooted World Citizenry*, ed. A.W. He and Y. Xiao, 1-12. Honolulu, HI: University of Hawaii Press.

Jiang, X. 2003. Relationship between the knowledge of the sounds and meanings of Chinese characters among foreign students with different native language backgrounds (不同母语背景的外国学生汉字知音和知义之间关系的研究). *Language Teaching and Linguistics Studies* (语言教学与研究) no. 6: 51-57.

Ke, C. 1998. Effects of language background on the learning of Chinese characters among foreign language students. *Foreign Language Annals* 31: 91-100.

Li, D. 2003. Intermediate and advanced Chinese learners' character knowledge (中高级留学生识字量抽样测试报告). *Journal of the College of Chinese Language and Culture of Jinan University* (暨南大学华文学院学报) no. 2: 12-19.

Li, R. and M. Wu. 2005. Two principles for teaching Chinese vocabulary to speakers of other languages (略论对外汉语词汇教学的两个原则). *Language Teaching and Linguistic Studies* (语言教学与研究) no. 2: 41-47.

Liu, Y. et al. 2011. *Integrated Chinese* (中文听说读写) (3rd ed.). Boston, MA: Cheng and Tsui Company.

Wu, S.-M. et al. 2010. *Chinese Link* (中文天地). Upper Saddle River, NJ: Pearson Prentice Hall.

Milton, J. 2009. *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.

Milton, J. and N. Hopkins. 2006. Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners. *The Canadian Modern Language Review* 63: 127-147.

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.

Qian, D.D. 1999. Assessing the roles of depth and breadth of vocabulary knowledge in reading comprehension. *Canadian Modern Language Review* 56: 282-307.

Rayner, K. et al. 2002. How psychological science informs the teaching of reading. *Psychological Science in the Public Interest* 2, no. 2: 31-74.

Schmitt, N., D. Schmitt and C. Clapham. 2001. Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing* 18: 55-88.

Stubbe, R. 2012. Do pseudowords false alarm rates and overestimation rates in Yes/No vocabulary tests change with Japanese university students' English ability levels. *Language Testing* 29: 471-488.

Taylor, I. and M.M. Taylor. 2014. *Writing and Literacy in Chinese, Korean and Japanese.* Philadelphia, PA: John Benjamins.

Zhang, H. 2016. Does morphology play an important role in L2 Chinese vocabulary acquisition. *Foreign Language Annals* 49: 384-402.

Zhang, K. 1997. Statistical analysis of commonly-used characters in yield words (汉语构词基本字的统计分析). *Language Teaching and Linguistic Studies* (语言教学与研究) no. 1: 42-51.

Zhao, G. 2003. Beginning Chinese learners' character knowledge (初级阶段欧美留学生识字量与字的构词数). *Applied Linguistics* (语言文字应用) no. 3: 106-112.

Table 1. *Correlations between tasks*

	Tasks	1	2	3
1	Vocabulary No-Pinyin	-	.881***	.077
2	Vocabulary Pinyin	.673***	-	.131
3	Orthographic Processing	.467**	.331	-

Note. Correlations for HLs below diagonal for FLs above diagonal.

** $p < .01$ *** $p < .001$

Table 2. *Means and standard deviations of scores*

	HLs		FLs	
	M	SD	M	SD
Vocabulary No-Pinyin	11.81	3.58	12.36	3.39
Vocabulary Pinyin	15.97	4.45	13.73	4.74
Orthographic Processing	4.26	.73	4.23	.97
Self-Assessment				
Speaking	3.10	1.01	2.77	.81
Listening	3.65	.80	3.00	.87
Reading	2.87	.85	3.14	.64
Writing	2.68	.83	2.73	.70
Overall	3.07	.63	2.96	.49

Figure 1. Pinyin X group ANOVA

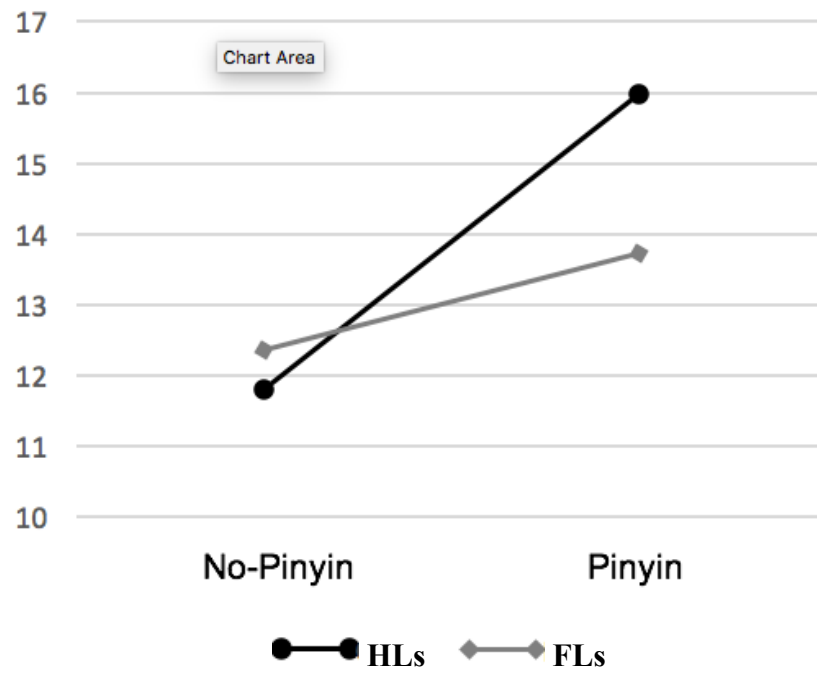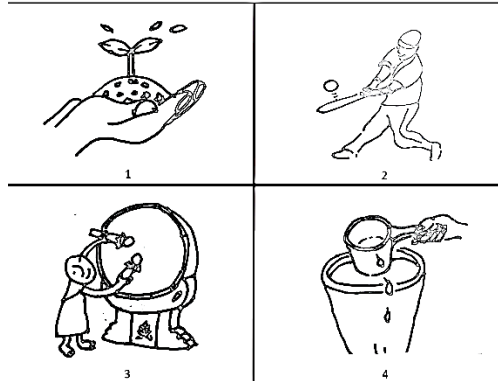# Appendix. Examples of the picture selection task

捧 pěng

轮椅 lúnyǐ