tesol QUARTERLY
tesol international association

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Appendix A. Explicit instruction (EI) sheet.
Appendix B. Text reconstruction (TR) test.
Appendix C. Criteria for scoring the past counterfactual conditional.

Vocabulary Coverage of Reading Tests: Gaps Between Teaching and Testing

TAN JIN
Sun Yat-sen University
Guangzhou, China

YUNTING LI
Northeastern University
Shenyang, China

BAICHUAN LI
YouMi Mobile
Guangzhou, China

doi: 10.1002/tesq.324

In exploring the relationship between vocabulary and reading, a consensus that vocabulary is one of the most essential factors contributing to the comprehension of texts (Laufer, 1992; Nation, 2006) has been reached. The vocabulary in most reading studies has traditionally been measured and analyzed by counting the quantity of word families to reflect the vocabulary needed (i.e., vocabulary demands) for reading comprehension (see Webb & Macalister, 2013). A word family, including a root form, its inflections, and its derivatives, has been widely employed as an estimator quantifying vocabulary demands (Schmitt & Schmitt, 2014). The vocabulary demands were measured through an investigation of vocabulary coverage of texts in a number

of studies on comprehension of texts (Schmitt & Schmitt, 2014). In a pioneering study, Laufer (1989) proposed a threshold of 95% vocabulary coverage for reading comprehension. Subsequently, compared to the 95% threshold, Hu and Nation (2000) found that 98% vocabulary coverage was a more reasonable figure for learners' unassisted comprehension of fiction texts. More recently, Laufer and Ravenhorst-Kalovski (2010) have suggested two thresholds of vocabulary coverage for comprehension of texts—an optimal one, 98%, and a minimal one, 95%.

Using 98% vocabulary coverage as a target and baseline, vocabulary demands of English reading for various purposes have been examined. As a pioneer, Nation (2006) reported on the trials of fourteen 1,000 word family lists from the British National Corpus (BNC) measuring vocabulary demands of unassisted comprehension, with the result indicating that 8,000 to 9,000 word families were needed for the comprehension of written texts. Specifically, to clarify vocabulary demands of texts written for language learners, Webb and Macalister (2013) conducted a corpus-driven analysis to calculate the number of word families, and 3,000 word families were deemed necessary for reaching 98% coverage of graded readers. More recently, in investigating vocabulary demands of texts employed for testing reading comprehension, Webb and Paribakht (2015) analyzed 38 texts from the reading section of CanTEST, a Canadian university admissions English test. Results showed that, on average, 6,000 and 14,000 word families were needed to reach 95% and 98% coverage, respectively. As mentioned, attempts have been made to investigate vocabulary demands of teaching or testing reading based on the 95% or 98% vocabulary coverage, but it is still an underexplored area in examining the vocabulary coverage of reading tests by the vocabulary lists used for teaching reading in order to investigate the vocabulary gaps existing between teaching and testing (cf. L. J. Zhang & Anual, 2008). This study is thereby designed to address this concern.

THE CONTEXT

In China, the vocabulary lists provided in official documents specify the vocabulary items to be covered in classroom teaching (Wang, 2013). Three national English curriculum standards are in use for guiding English language teaching in China: *The English Curriculum Standards for Compulsory Education* (2011 version, Ministry of Education, 2012), *The English Curriculum Standards for High School* (Ministry of Education, 2003), and *The College English Curriculum Requirements* (Higher Education Department of Ministry of Education, 2007). The current study focuses on the compulsory education stage, where—in line with the huge population of basic English learners, according to the

Ministry of Education (2014)—approximately 138 million Chinese students were learning in schools at the compulsory education stage in 2013. At the compulsory education stage, vocabulary demands have been specified in *the vocabulary list* provided by the *English Curriculum Standards for Compulsory Education* (2011 version, Ministry of Education, 2012; hereafter the curriculum vocabulary list). More directly, all teaching materials, including textbooks, must pass an examination ahead of use by the committee commissioned by the Ministry of Education (see Ministry of Education, 1995), with a vocabulary check ensuring that vocabulary used in the teaching materials is covered by the curriculum vocabulary list. As such, the vocabulary demands of teaching basic English reading have also been within the curriculum vocabulary list.

With respect to testing basic English at the compulsory education stage, high-stakes English tests have been conducted in senior high school entrance examinations to assess the English proficiency of students (Chen, 2011). Senior high school entrance English tests routinely consist of listening, reading, and writing sections, with the speaking component as optional depending on the testing policies of local governments (see Zhongkao Website, 2015). For testing basic English reading, four to five texts, followed by a number of multiple choice items, are used to test comprehension (Liang, 2009) along similar lines to the TOEFL iBT in its reading section (Educational Testing Service, 2005). The texts included in the reading section of senior high school entrance English tests pass through a lengthy process that includes text selection and revision by experts and experienced item writers to ensure that texts are suitable for students at the particular level of English reading proficiency during the compulsory education stage (see Xu, 2013). The current study will use the curriculum vocabulary list to analyze the vocabulary coverage of the reading texts from senior high school entrance English tests with a focus on detecting vocabulary gaps between teaching and testing reading.

RESEARCH QUESTIONS

Specifically, two coordinated questions guided the study in the Chinese context:

1. Do vocabulary gaps exist between teaching and testing basic English reading?
2. If there are vocabulary gaps, what is the nature of the gaps (i.e., are the gaps big or small)?

METHOD

The Curriculum Vocabulary List

The curriculum vocabulary list provided by the *English Curriculum Standards for Compulsory Education* in 2011 (see Ministry of Education, 2012, pp. 59–91) includes 1,507 vocabulary items, setting up vocabulary requirements for teaching basic English reading.

Texts From High School Entrance English Tests

Test papers of senior high school entrance English tests in China are in most cases routinely released and published subsequent to the completion of senior high school entrance examinations; *A Nationwide Collection of Real Test Papers from Senior High School Entrance English Tests* is, for example, published annually by the Tianli Cultural Media Group (see Tianli, 2015). In this connection, the published test papers, in which each reading section includes four to five texts, provide rich text data for the current study. In order to elicit representative texts, we collected 859 texts from test papers of senior high school entrance English tests during the period 2003–2013, covering 26 provinces and municipalities (e.g., Beijing city and Jiangsu province) from four economic regions in China (i.e., East China, Central China, Northeast China, and Western China; see National Bureau of Statistics of China, 2011). As shown in Table 1, among the 859 texts the texts from the East China region account for the largest portion (64%; 554 texts) and Northeast China accounts for the smallest (7%; 57 texts). The distribution of texts from the four economic regions was overall representative of the educational development of basic English teaching in China. The text data for the study comprised all 859 texts.

TABLE 1
The Reading Texts From Senior High School Entrance English Tests

Economic regions	Provinces and municipalities	Number of texts (%)
East China	Beijing, Tianjin, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong and Guangdong	554 (64)
Central China	Shanxi, Anhui, Jiangxi, Henan, Hubei and Hunan	121 (14)
Northeast China	Heilongjiang, Jilin and Liaoning	57 (7)
Western China	Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Shaanxi, Gansu and Xinjiang	87 (10)
Not identified	N/A	40 (5)
	Total	859 (100)

Procedure and Results

In the curriculum vocabulary list, 1,507 vocabulary items do not appear as headwords of word families, leading to difficulty in counting the number of word families. To convert the vocabulary items into headwords of word families, we adopted the BNC/COCA word family lists (Nation, 2014). For example, two vocabulary items, *ability* and *able*, are subsumed into a headword representing their word family, *able*. To facilitate the conversion procedure, an automatic transformer, Word Family Converter (Jin & Li, 2015), was employed. As a result, the 1,507 vocabulary items were converted to 1,357 word families, which forms List A in the current study.

Regarding the 859 texts from high school entrance English tests, words in each text were first subsumed into headwords representing their word families using the automatic transformer. In analyzing the vocabulary coverage of List A for the 859 texts, we developed another list, List B, including marginal words (e.g., *en*, *oh*, and *la*) and proper nouns (e.g., *Nanjing* and *Guangzhou*), because marginal words and proper nouns create minimal burdens in the comprehension of texts (see Webb & Macalister, 2013). List B was originally created based on the corpus of the 859 texts as well as other corpora including reading texts from high-stakes English tests in China such as the National Matriculation English Test (NMET). We identified approximately 3,800 words as marginal words and proper nouns, using two trained independent coders with an agreement proportion of 97%. Table 2 reports the values of minimum (Min), maximum (Max), mean (*M*), and standard deviation (*SD*) for the vocabulary coverage using List A and List B. As shown in Table 2, in total the vocabulary coverage has a mean of 92.82% for all 859 texts. As for the four economic regions, the texts from Western China region obtain the highest mean (i.e., 94.14%) while the texts from the other three economic regions have comparatively lower means, all closer to 93%.

TABLE 2
The Vocabulary Coverage of the Reading Texts

Economic regions	*n*	Min (%)	Max (%)	Mean (%)	*SD* (%)
East China	554	74.68	100.00	92.69	4.12
Central China	121	75.44	100.00	92.52	4.24
Northeast China	57	77.59	98.13	92.52	3.34
Western China	87	82.76	100.00	94.14	3.83
Not identified	40	82.91	100.00	93.09	3.96
Total	859	74.68	100.00	92.82	4.07

DISCUSSION

This article reports on a vocabulary coverage study of reading tests in order to explore vocabulary gaps between teaching and testing basic English reading in China. For conducting the vocabulary coverage study, we adopted the curriculum vocabulary list used for teaching to analyze the vocabulary coverage of 859 reading texts from high school entrance English tests. To deal with the first research question, as the curriculum vocabulary list has a vocabulary coverage mean of 92.82% for all of the reading texts (see Table 2) it could thus be concluded that, on the whole, vocabulary gaps do exist between teaching and testing basic English reading. The vocabulary discrepancy between teaching and testing may be explained from two aspects. On the one hand, the curriculum vocabulary list in China can only set up a comparatively lower vocabulary standard in the central curriculum due to the huge population of students and in the face of uneven regional education development (Liu, 2013). On the other hand, however, to discriminate between test-takers for admission purposes, the high school entrance English tests have inevitably had to raise the vocabulary demands of reading comprehension (Sun, 2009). As a consequence, the discrepancy between these two situations of the lower vocabulary standard in teaching and the higher vocabulary demands for testing results, inevitably perhaps, in the existence of vocabulary gaps.

To answer the second research question, we further examined the nature of the existing vocabulary gaps. As reviewed, the optimal and minimal targets of vocabulary coverage for comprehension of reading texts are 98% and 95%, respectively (Laufer and Ravenhorst-Kalovski, 2010; Nation, 2006; see also Schmitt & Schmitt, 2014). The vocabulary coverage mean for all of the reading texts (i.e., 92.82%)—even for the highest vocabulary coverage mean (i.e., 94.14%) for the texts from the Western China region—is lower than the minimal target of vocabulary coverage (i.e., 95%). As such, the existing vocabulary gaps could probably affect the comprehension of reading texts. In this connection, the vocabulary gaps might be considered to be big. It is also worth noting, as demonstrated in Table 2, that there are variations among the vocabulary coverage mean in the four economic regions. With the comparatively higher means of vocabulary coverage, more vocabulary could be needed for comprehension of texts from the East, Central, and Northeast regions than those from the Western region. One possible explanation is that the progressive local education development allows students to accumulate more vocabularies through extensive readings beyond the compulsory teaching textbooks, for example, the East China region (see N. N. Zhang, 2013).

IMPLICATIONS AND CONCLUSION

The current study has strong implications for teaching vocabulary and testing development in China. Due to the existence of vocabulary gaps, there certainly is room for improvement in vocabulary for teaching English reading. The curriculum vocabulary list sets the vocabulary standard in the central curriculum, but it is recommended that teachers incorporate additional vocabulary teaching into their school-based curriculum development based on the vocabulary profiles of learners (see EDB Website, 2015). In the meantime, given the complexity and variability of English language teaching in China, the vocabulary taught in some educationally developed regions may be larger than those stipulated by the curriculum vocabulary list. As such, equating the curriculum vocabulary list with vocabulary covered in actual teaching should be considered in future. Moreover, as big vocabulary gaps could probably affect the comprehension of texts in testing reading, it would be a great challenge for item writers to create or edit texts for setting appropriate vocabulary gaps in the production of texts to be included in reading tests. In this connection, specific guidelines should be developed together with certain vocabulary lists to control the composition of vocabulary gaps so that the quality of texts used in reading tests is maintained. Although the study was conducted solely in China, it advances our understanding and contributes empirical evidence to the vocabulary coverage issue on testing reading in general. As 95% vocabulary coverage is widely accepted as the minimal target for comprehension of reading texts (see Schmitt & Schmitt, 2014), the vocabulary gaps detected between teaching and testing in the current study have been deemed as big. However, for testing reading, it might not be necessary to reach the minimal reading level for all students; it might be sufficient to assess their reading proficiency and to allow for shortfalls in some situations. Perhaps due to this concern, doubt has been cast on the use of 95% vocabulary coverage as the minimal target in reading tests (see Green & Hawkey, 2012; Webb & Paribakht, 2015). For example, in Green and Hawkey (2012), the BNC 2,000 word families and the Academic Word List (Coxhead, 2000) were not expected to reach 90% vocabulary coverage of the four reading texts for IELTS (Nation & Waring, 1997). Likewise, in examining the vocabulary coverage of 38 reading texts from CanTEST, Webb and Paribakht (2015) found that 3,000 word families were sufficient to reach 95% vocabulary coverage of a text, but 14,000 word families were required to reach 95% vocabulary coverage of another text. The findings of these two studies seem to support the use of a lower vocabulary coverage target for testing reading. On this point, the results of the

current study, a vocabulary coverage mean of 92.82% based on a large sample of 859 reading texts, lends support to the adoption of a lower vocabulary coverage target for testing reading.

As for the vocabulary coverage of reading tests, few attempts have been documented in employing the vocabulary lists used for teaching to analyze the reading texts from high-stakes tests. This paper has therefore contributed an empirical investigation in the context of China. Using the curriculum vocabulary list to analyze the 859 reading texts from high school entrance English tests, the vocabulary coverage mean of 92.82% has indicated vocabulary gaps existing between teaching and testing reading. Moreover, the vocabulary gaps have been seen as big, considering 95% vocabulary coverage as the minimal target for comprehension of reading texts. On the whole, the vocabulary gaps, which could affect the comprehension of reading texts, have enhanced awareness of teaching vocabulary and testing development in China. More importantly, the findings of the study also link to international literature on the topic of vocabulary coverage showing that a lower vocabulary coverage target might deserve second thought from the perspective of testing reading. In view of China's huge population of students as well as its uneven regional education development, a larger sample size of text data from more regions in China is needed in future studies to reflect the variability and complexity of English teaching and testing there. Future qualitative studies should examining the extent and importance of the vocabulary gaps in affecting students' comprehension of the reading texts. For example, some vocabularies not covered by the curriculum vocabulary list are easy to guess using the context or some clues involving English interpretations and Chinese translations.

ACKNOWLEDGMENTS

This research was supported by the National Social Science Fund of China (13AYY005). We are indebted to the editor of *TESOL Quarterly* and the anonymous reviewers. Their insightful comments and suggestions on the draft were invaluable in helping us to introduce clarity and focus.

THE AUTHORS

Tan Jin is a research associate professor at Sun Yat-sen University in China. His research interests include language testing, data mining, and mobile learning. He created the first confidence scoring method for performance assessments, and is recently constructing corpora to investigate distinguishing features across proficiency levels.

Yunting Li is an MA student at Northeastern University in China. She is currently working on her master's thesis, which focuses on the vocabulary demands of

English reading. Her research interests include language testing and corpus linguistics.

Baichuan Li leads a data mining group at YouMi Mobile in China. His main research interests are in natural language processing and data mining. During his research visits with Google (2011) and Microsoft (2012), he created a text mining system for processing language data in industry contexts.

REFERENCES

Chen, T. (2011). The analysis and enlightenment of high school entrance examination. *Journal of Yunnan Education*, (9), 4–7.

Coxhead, A. (2000). An academic word list. *TESOL Quarterly*, *34*, 213–238. doi:10.2307/3587951

EDB Website. (2015). School-based curriculum development in the primary schools. Retrieved from http://www.edb.gov.hk/en/edu-system/primary-secondary/applicable-to-primary-secondary/sbss/school-based-curriculum-primary/index.html

Educational Testing Service. (2005). *TOEFL iBT tips: How to prepare for the next generation TOEFL test and communicate with confidence*. Princeton, NJ: Educational Testing Service.

Green, A., & Hawkey, R. (2012). Re-fitting for a different purpose: A case study of item writer practices in adapting source texts for a test of academic reading. *Language Testing*, *29*, 109–129. doi:10.1177/0265532211413445

Higher Education Department of Ministry of Education. (2007). *College English curriculum requirements*. Shanghai, China: Shanghai Foreign Language Education Press.

Hu, M. H., & Nation, P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language*, *13*, 403–430.

Jin, T., & Li, B. (2015). *Word family converter [computer software]*. Guangzhou: Sun Yat-sen University.

Laufer, B. (1989). What percentage of text lexis is essential for comprehension? In C. Lauren & M. Nordman (Eds.), *Special language: From humans thinking to thinking machines* (pp. 316–323). Clevedon, England: Multilingual Matters.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In H. Bejoint & P. Arnaud (Eds.), *Vocabulary and applied linguistics* (pp. 126–132). Basingstoke, England: Macmillan.

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical text coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language*, *22*, 15–30.

Liang, H. M. (2009). The characteristics of high school entrance English tests and the strategies of answering the test items. *Educational Practice and Research*, *10*, 27–31.

Liu, Y. (2013). The practice of vocabulary teaching in junior high school. *Inspiration and Wisdom: Education*, *11*, 47–47.

Ministry of Education, China. (1995). The stipulates of the compilation, review and selection of textbooks for primary and secondary schools. *Subject Education*, *10*, 1–3.

Ministry of Education, China. (2003). *English curriculum standards for high school*. Beijing, China: People's Publication Press.

Ministry of Education, China. (2012). *English curriculum standards for compulsory education* (2011 version). Beijing, China: Beijing Normal University Press.
Ministry of Education, China. (2014). *2013 national educational development statistics bulletin I.* Retrieved from http://www.moe.edu.cn/publicfiles/business/htmlfiles/moe/moe_633/201407/171144.html
Nation, P. (2006). How large a vocabulary is needed for reading and listening? *Canadian Modern Language Review, 63*, 59–82. doi:10.3138/cmlr.63.1.59
Nation, P. (2014). The BNC/COCA word family lists (17 September 2012). Retrieved from http://www.victoria.ac.nz/lals/about/staff/publications/paul-nation/Information-on-the-BNC_COCA-word-family-lists.pdf
Nation, P., & Waring, R. (1997). Vocabulary size, text coverage, and word lists. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition, pedagogy* (pp. 6–19). Cambridge, England: Cambridge University Press.
National Bureau of Statistics of China. (2011). Four economic regions in China. Retrieved from http://www.stats.gov.cn/ztjc/zthd/sjtjr/dejtjkfr/tjkp/201106/t20110613_71947.htm
Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching, 47*, 484–503. doi:10.1017/S0261444812000018
Sun, Y. Y. (2009). How to effectively teach vocabulary in English reading teaching in junior high school. *Foreign Language Teaching in Schools, 32*, 31–34.
Tianli. (2015). The website for senior high school entrance tests. Retrieved from http://zhongkao.tl100.com/
Wang, Q. (2013). A good comprehension and implementation of English curriculum standards for compulsory education (2011 version, Ministry of Education, 2012). *Jiangsu Education, 5*, 7–9.
Webb, S., & Macalister, J. (2013). Is text written for children useful for L2 extensive reading? *TESOL Quarterly, 47*, 300–322. doi:10.1002/tesq.70
Webb, S., & Paribakht, T. S. (2015). What is the relationship between the lexical profile of test items and performance on a standardized English proficiency test? *English for Specific Purposes, 38*, 34–43. doi:10.1016/j.esp.2014.11.001
Xu, J. Y. (2013). The analysis of the validity of tests for reading comprehension. *Educational Measurement and Evaluation, 2*, 44–48.
Zhang, L. J., & Anual, S. (2008). The role of vocabulary in reading comprehension: The case of secondary school students learning English in Singapore. *RELC Journal, 39*, 52–77. doi:10.1177/0033688208091140
Zhang, N. N. (2013). An empirical study of incidental vocabulary acquisition by English learners. *Foreign Language Education, 34*, 73–77.
Zhongkao Website. (2015). Senior high school entrance English tests. Retrieved from http://www.zhongkao.com/