Vocabulary Size, Reading Proficiency and Curricular Design: The Case of College Chinese, Russian and Spanish

Jane F. Hacking, Fernando Rubio, and Erwin Tschirner

Abstract A key goal of college foreign language study is L2 literacy development and literary texts from the target culture form the backbone of upper division curricula. Much of the empirical research to date on vocabulary size and reading proficiency has focused on learners of English. This article presents data on the reading proficiency level of 155 college students of Chinese (N = 46), Russian (N = 48) and Spanish (N = 61) and considers these results in terms of these same students' receptive vocabulary knowledge in the language. The study shows very high correlations between reading proficiency and receptive vocabulary size and that, in general, the vocabulary sizes of the college students participating in the study were not sufficient to read at the Advanced level. We suggest that programs and instructors consider a more intentional approach to vocabulary learning across the curriculum.

Keywords Reading · Proficiency · Vocabulary · Postsecondary · Assessment · Chinese · Russian · Spanish

1 Introduction

Undergraduate foreign language programs aim to develop students' L2 proficiency in speaking, reading, listening and writing. Reading proficiency assumes particular importance as it gives students access to literary texts from the target culture that form the backbone of upper division curricula. There is extensive research on

J. F. Hacking (✉)
World Languages and Cultures, University of Utah, Salt Lake City, UT, USA
e-mail: j.hacking@utah.edu

F. Rubio
Second Language Teaching and Research Center, University of Utah, Salt Lake City, UT, USA
e-mail: Fernando.Rubio@utah.edu

E. Tschirner
Herder Institute, University of Leipzig, Leipzig, Germany
e-mail: tschirner@uni-leipzig.de

P. Winke, S. M. Gass (eds.), *Foreign Language Proficiency in Higher Education*, Educational Linguistics 37, https://doi.org/10.1007/978-3-030-01006-5_3

second language reading that explores the role of various factors (e.g., L1 literacy, grammatical and/or vocabulary knowledge) in the development of L2 reading proficiency. Bernhardt (2011) argues for a model of L2 reading proficiency that captures the contribution and interaction of these many variables. A deeper understanding of the role of specific variables, such as vocabulary knowledge, will contribute to such a comprehensive model. This article presents data on the reading proficiency level of 155 college students of Chinese (N = 46), Russian (N = 48) and Spanish (N = 61) and considers these results in terms of these same students' receptive vocabulary knowledge in the language. Much of the empirical research to date on vocabulary size and reading proficiency has focused on learners of English, but a recent study of L2 Russian learners (Hacking & Tschirner, 2017) reported identifiable lexical minima associated with particular levels of L2 Russian reading proficiency. This chapter builds on the Russian data with the addition of data from learners of Chinese and Spanish.

2 Background

2.1 Vocabulary Size and L2 Development

The relationship between a second language (L2) learner's vocabulary size and his or her overall L2 proficiency has been well-established. Milton (2009) provides a comprehensive overview of this general line of research. There are also a number of studies that focus on the relationship between vocabulary knowledge and reading ability in particular (e.g., Nation, 2006; Tschirner, 2004). Staehr (2008) reports research showing that vocabulary size correlated strongly with listening, reading and writing scores, but that, of these, the strongest correlation was with reading scores. An important finding in the reading proficiency and vocabulary knowledge research is that there are lexical thresholds associated with the achievement of specific language goals. Studies have found that, for example, a particular exam score, a certain proficiency rating, or a reading comprehension score requires specific levels of L2 vocabulary knowledge.

Overall, the research suggests that lexical thresholds are tied to expected text coverage, that is, the percentage of lexical items in a text the reader understands. Researchers concur that readers need to understand between 95% and 98% of the tokens (running words) of a text in order to comprehend the text (Carver, 1994; Hirsh & Nation, 1992; Hu & Nation, 2000; Nation, 2006; Schmitt, 2008; Schmitt, Jiang, & Grabe, 2011). Nation (2006) proposes that for English the most frequent 9000 word families provide coverage of 98% of words in a wide range of texts. Similarly, Laufer and Ravenhorst-Kalovski (2010) assert that for English learners, an optimal threshold for reading purposes is knowledge of 8000 words and a minimal one is 4000–5000 words. Clearly then, adequate comprehension of unsimplified texts is not readily accessible for L2 learners until they have acquired substantial

vocabulary knowledge. As Nation (2007, p. 9) noted, "…most text beyond the 3,000-word level of graded readers series is very difficult for foreign language learners. This is because in most novels a very large number of different words occur beyond the learners' current vocabulary knowledge." While Nation focuses on narrative texts, similar figures have been cited for newspaper and other kinds of writing (Schmitt et al., 2011). And while reading itself can be a route to vocabulary acquisition, research shows approximately twelve repetitions of a word in different contexts may be needed for the word to be acquired (Nation, 2014), and that, to achieve twelve repetitions of, e.g., the fourth most frequent 1000 words in English, i.e. the most frequent 3000–4000 words, approximately half a million running words need to be read. We explore the implications of this research in light of results of this study in the discussion section where we consider approaches to vocabulary learning.

The idea of lexical thresholds is found also in descriptions and discussions of various proficiency scales. Descriptors can be broadly worded, for example, "limited range of vocabulary" or "uses an adequate range of vocabulary for the task" (IELTS, n.d.). Or, they can appeal to concrete numbers. Milton (2009), in discussing the CEFR, notes that at earlier stages of development, CEFR descriptors included vocabulary lists. He references lists tied to level B1, observing that different languages had different thresholds according to these lists (German 2400 words, English 2200 words, Italian and French 1800 words, and Spanish 800 words.) For the languages considered here, the national testing instruments of those languages establish several levels of proficiency and in the case of Russian and Chinese, establish lexical minima for each level.

The Russian Ministry of Education and Science has developed the Test of Russian as a Foreign Language (TORFL)/Тест по русскому языку как иностранный (ТРКИ) as part of the CEFR. Some official test specifications can be found on the website of Moscow State University's Training and Testing Language Center for Foreigners: http://russian-test.com/tests/torfl/. Learners may achieve one of six levels: Elementary, Basic, Level 1, Level 2, Level 3, Level 4. Each level outlines a set of competencies and is accompanied by a description of what language at that level enables the learner to do. For example, a learner who scores at the Basic Level, the level required to become a naturalized Russian citizen, is described as being able to "satisfy the most basic communicative needs…in a limited number of predictable situations" (http://russian-test.com/assets/docs/Trebovaniya_-_basic.pdf; translated from the Russian). The official test documentation also specifies a minimum number of vocabulary words required at a given level as shown in Table 1. After Level 1, vocabulary knowledge targets are split into receptive and productive categories. Productive vocabulary knowledge targets are smaller than those for the receptive lexicon. Test documentation states that a learner must have achieved Level 1 to begin a course of study at a Russian institution of higher education, and that Level 2 proficiency is necessary to receive a degree taught in Russian (with the exception of degrees in, for example, philology for which Level 3 proficiency is the stated requirement). Table 1 shows TORFL proficiency levels with minimum vocabulary for each as well as established equivalencies between TORFL levels and those

Table 1 ACTFL, CEFR, correspondences to TORFL and lexical minimums per level as established by the TORFL

ACTFL	CEFR	TORFL	Minimum vocabulary
N	A1	Elementary	780
IM	A2	Basic	1300
IH	B1	Level 1	2300
AM	B2	Level 2	10,000 (6000 active)
AH	C1	Level 3	12,000 (7000 in active)
S	C2	Level 4	20,000 (8000 in active)

Table 2 ACTFL, CEFR, and HSK correspondences as claimed by Mandarin House (ACTFL) and FaCH (CEFR) and their lexical minimums as established by Hanban

ACTFL	CEFR	HSK	Vocabulary
NL		Level 1	150
NM	A1.1	Level 2	300
IL	A1	Level 3	600
IM	A2	Level 4	1200
IH	B1	Level 5	2500
Advanced	B2	Level 6	5000

for ACTFL and CEFR. It is unclear if the suggested correspondences between ACTFL/CEFR and TORFL were based on empirical studies. The correspondences between the CEFR and ACTFL however, seem to be fairly consistent with the official ACTFL CEFR crosswalk (ACTFL, 2016).

The Chinese proficiency test Hanyu Shuiping Kaoshi (HSK) developed by the Office of Chinese Language Council International (Hanban), an organization associated with the Chinese Ministry of Education, specifies six levels of proficiency and provides both correlations to the CEFR scale and brief descriptions of what a learner can be expected to be able to do at each level. For example, "[T]est takers who are able to pass the HSK (Level III) can communicate in Chinese at a basic level in their daily, academic and professional lives. They can manage most communication in Chinese when travelling in China." (http://english.hanban.org/node_8002.htm#no1). While Hanban equates Level III with B1 in the CEFR, the German Association of Teachers of Chinese (FaCH) has resolutely questioned the correspondences established by Hanban and has instead proposed to equate HSK 3 with A1, HSK 4 with A2, and so on (http://www.fachverband-chinesisch.de/sites/default/files/FaCh2010_ErklaerungHSK_en.pdf). Evidence to support this equation comes from the crosswalk published by Mandarin House that uses the Hanban correspondences for the CEFR but very different ones for ACTFL (http://www.mandarinhouse.cn/images/general_chinese_program.pdf). Using the ACTFL CEFR crosswalk established by ACTFL (2016), Table 2 presents the vocabulary size requirements for each HSK level as proposed by Hanban for reading purposes and their putative corresponding ACTFL levels.

Table 3 General and specific notions identified for each level by the Instituto Cervantes's *Plan Curricular*

CEFR	General	Specific
A1	463	1146
A2	608	1584
SUBTOTAL A	1071	2730
B1	1567	3336
B2	2730	5541
SUBTOTAL B	4297	9100
C1	3976	5810
C2	4243	5676
SUBTOTAL C	8219	11,513
TOTAL	13.587	23.343

As Table 2 shows ACTFL IM appears to be associated with the 1000 band, IH with the 2000 band, and *Advanced* with the 5000 band.

The Instituto Cervantes has also developed a set of language proficiency tests for Spanish developed around the CEFR. Although there are no explicit lexical minima associated with particular levels of proficiency, the exams are designed to take into account the vocabulary lists specified for each level in the *Plan Curricular* published by the Instituto Cervantes. Following a notional-functional approach, the *Plan* specifies two lists of *nociones* (notions) per level, one labelled as 'general' and one as 'specific'. These *nociones* are lexical units that include individual words as well as collocations, idiomatic expressions and other phrases. Table 3 provides the overall count that was provided to us by Mr. García-Santa Cecilia, from the Instituto Cervantes (personal communication, June 5th, 2017), expected to be included in the curriculum at each CEFR level of instruction.

The vocabulary sizes expected for each level by the largest national cultural organizations representing the three languages in question vary considerably, and – as far as Russian and Spanish are concerned – are substantially larger than the ones established for English, especially for receptive (reading) purposes, namely, Russian 10,000 and Spanish 9100 (specific vocabulary) for B2 and Russian 20,000 and Spanish 23,000 for C2.

2.2 Vocabulary Learning in L2 Methods Textbooks

It is risky to make generalizations about what goes on in language classrooms without conducting systematic observations, which are beyond the scope of this study. It is the case however, that there is often a gap between empirical research and the implementation of its findings by practitioners. For example, it is probably reasonable to assume that most language instructors do not access primary research to inform their teaching. Therefore, in this section we examine textbooks on L2 teaching methodology because they reflect the profession's favored pedagogies and provide a good indication of the training that prospective teachers receive. Grabe (2009)

asserts that in order to learn vocabulary, students need a combination of "vocabulary instruction, vocabulary-learning strategies, extensive reading and word learning from context, heightened student awareness of new words, and motivation to use and collect words" (p. 283). While L2 teaching pedagogy has emphasized the importance of learning new words in context and the importance of extensive reading for vocabulary acquisition, the specific word-level learning strategies that Grabe mentions are conspicuously absent from the L2 methods textbooks used in most teacher training programs. In fact, many explicitly or implicitly avoid direct vocabulary instruction techniques.

All of the most commonly used textbooks on second language teaching methodology dedicate attention to the development of reading proficiency, sometimes combined with listening as the development of interpretive skills, but the emphasis is typically on the process of reading and a description of the tasks involved in that process (pre-reading, scaffolding, guided practice, etc.). A good example is Omaggio Hadley's (2001) *Teaching Language in Context*, perhaps the most influential and widely used methods textbook in the US for the past two decades. The author dedicates a chapter to developing proficiency in listening and reading and provides a rationale for focusing on the receptive skills as well as a number of techniques and ideas for activities. However, there is no mention of the role of vocabulary knowledge in reading comprehension and no specific suggestions of ways to help learners build their vocabulary. A similar focus can be found in Lee and VanPatten's (2003) *Making Communicative Language Teaching Happen*, another widely used textbook on language teaching methodology. The authors emphasize the role of comprehensible input in the acquisition of new vocabulary and encourage techniques that facilitate making form-meaning connections following Terrell's (1986) notion of binding. They also warn that memorization is "no substitute for meaning-bearing comprehensible input in learning vocabulary" (p. 37). In the same vein, Shrum and Glisan (2010) emphasize the importance of placing new lexical items within a meaningful context and engaging learners in collaborative work with peers to facilitate vocabulary acquisition. They also discuss the limitations of incidental acquisition through reading and acknowledge the need to provide students with opportunities for focused work on vocabulary, but they do not offer any suggestions as to what that focused work should look like.

Brandl (2008) places more emphasis on the importance of lexical acquisition by devoting an entire chapter specifically to the teaching and learning of vocabulary (which is separate from a chapter dedicated to the development of reading skills). In the introduction to the chapter, the author states that, "the learning and acquisition of vocabulary plays one of the most vital roles in becoming proficient in the target language" (p. 75). Brandl provides a variety of suggestions on how to use instructional resources and techniques that can be useful when presenting new vocabulary, with special emphasis on the advantages of using visual support such as realia and multimedia. However, as is the case with the other textbooks reviewed, his emphasis is on the presentation of new lexical items through meaningful, contextualized input.

Research Questions

To provide further evidence of the relationship between reading proficiency and vocabulary size and to make this relationship meaningful within the context of college foreign language study in the U.S., the following research questions were addressed.

1. What reading proficiency levels are attained by the study participants after how many years of college language study for these three languages?
2. How well does vocabulary size – measured as the receptive knowledge of various bands of the most frequent four to five thousand words of a language – predict reading proficiency levels as defined by ACTFL?
3. What ACTFL reading proficiency levels are predicted by what vocabulary sizes?
4. What are the differences, if any, between Chinese, Russian, and Spanish with respect to the relationship between vocabulary size and level of reading proficiency?

3 Methods

3.1 Participants

Participants in this study were college students of Chinese, Russian, and Spanish at a large Western US state university. The ACTFL Reading Proficiency Test (RPT) and the Vocabulary Levels Tests (VLT) were administered to a total of 155 students (Chinese: 46; Russian: 48; Spanish: 61) from the fall semester 2015 to the spring semester 2017. 61 students were female and 94 students were male. 15 students of Russian were enrolled in a second-semester course, 50 students were enrolled in a fourth-semester course (Chinese: 17; Russian: 7; Spanish: 26), 11 students were enrolled in a third-year course (Chinese: 9; Russian: 2), and 79 students were returned missionaries, who had spent between 18 and 24 months in a country where the target language was spoken and who were enrolled in an advanced language course. There was no random selection of students, and no attempt was made to match student characteristics across the three languages. Moreover, the extended immersion of the returned missionary students makes their language learning experience qualitatively and quantitatively different from other participants in the study. We present their data separately below.

3.2 Instruments

The ACTFL Reading Proficiency Test (RPT) is a standardized test for the global assessment of reading ability in a language (ACTFL 2013). The test measures how well a person spontaneously reads texts when presented with texts and tasks as described in the 2012 ACTFL Proficiency Guidelines. The test formats used in this

study consisted of 10–25 texts depending on a participant's proficiency level. There are five sublevels: Intermediate Low (IL), Intermediate Mid (IM), Advanced Low (AL), Advanced Mid (AM), and Superior (S). Each sublevel consists of five texts accompanied by three tasks (items) with four multiple-choice responses, only one of which is correct. Test specifications include genre, content area, rhetorical organization, reader purpose, and vocabulary (cf. ACTFL, 2013). Texts and tasks align at each level, for example, an *Intermediate* task requires understanding information that is contained in one sentence, whereas *Advanced* tasks require the ability to understand information that is spread out over several sentences or paragraphs. Tasks and multiple-choice responses are in the target language.

The RPT is a timed test with a total test time of 25 min per sublevel. Two sublevels are scored together, either the two levels taken or, if more than two levels were taken, the two highest levels that can be scored according to the specific algorithm of the test. Because there are no Novice texts or tasks, the Novice levels are determined according to how close the test-taker is to the Intermediate level. Test takers whose scores are below 33.33% of the maximum Intermediate score possible are rated NL, test takers whose score is between 33.33% and 50% are rated NM, and test takers whose scores are between 50% and 66.66% are rated NH. The test is Internet-administered and computer-scored (ACTFL, 2013).

The Vocabulary Levels Test (VLT) consists of a receptive and a productive test (Institute for Test Research and Test Development, 2013). It is modeled after the English Vocabulary Levels Test pioneered by Paul Nation (Nation, 1990). The VLT measures how many of the most frequent 4000 words of Chinese and 5000 words of Russian and Spanish are known. It consists of four to five bands: the most frequent 1000, 1001–2000, 2001–3000, 3001–4000, and 4001–5000 words. The receptive test, which was used in the present study, consists of ten clusters of six words each for each of these four or five bands. Each band is thus represented by 60 words. These words consist of 30 nouns, 18 verbs, and 12 adjectives and are chosen at random from the 1000 words of a band. Each cluster focuses on one part of speech. Three words of a cluster are targets, which need to be defined by choosing from a list of synonyms and paraphrases. The other three words are distractors.

The definition of receptive mastery of a particular band varies slightly in the literature. The two most common percentages used are 80% (e.g. Xing & Fulcher, 2007) and 85% (e.g., Schmitt, Schmitt, & Clapham 2001).

3.3 Data Coding

The RPTs used in this study were scored using either the Interagency Round Table (IRL) or the ACTFL scale.[1] ILR ratings were recoded into ACTFL ratings on the basis of raw scores according to the ACTFL algorithm. Following Rifkin (2005) and

[1] The RPT can be scored using the ILR or the ACTFL scale according to two different algorithms. The Chinese and Russian RPTs were originally scored using the ILR algorithm because the agency supporting the research requested ILR ratings. The Spanish RPTs were scored using the ACTFL rating.

others, ACTFL RPT results were coded numerically as follows: NL = 1, NM = 2, NH = 3, IL = 4, and so on, up to S = 10. VLT results were analyzed to determine if the words of a particular band (e.g., the most frequent 1000 words) were known. Three mastery criteria were investigated per language: 75% correct, 80% correct, and 85% correct (see below). The highest band at which students attained 75%, 80%, or 85% correct was considered their vocabulary level.

4 Results

4.1 Reading Proficiency

The results of the ACTFL Reading Proficiency Test (RPT) ranged from NL to S. Table 4 shows the distribution of the results by language and class level.

Table 4 shows that reading proficiency levels varied considerably across languages. While the median for fourth-semester Chinese students was 2 (NM), it was 3 (NH) for Russian and 5 (IM) for Spanish. The top 25% of students were at least 3 (NH) in Chinese, 4 (IL) in Russian, and 5 (IM) in Spanish. Returned missionaries had very high reading proficiencies in Russian and Spanish, while they scored below third-year students in Chinese. The median for Russian was 7 (AL) and for Spanish, it was 9 (AH). For Chinese, it was 3 (NH), lower than the 4.5 (IL to IM) for regular third-year students. The top 25% students were 10 (S) in both Russian and Spanish, while they started at 4.75 (IM) in Chinese, which was slightly lower than the third quartile of 5 (also IM) for third-year students. Of the 20 students returning from a mission in a Chinese-speaking country, only three were close to or at the *Advanced* level of reading proficiency (1 each at IH, AL, and AM) and many were NL ($N = 7$). While only one third-year student was AL, only one was NL, and most were IL or IM ($N = 4$). The median for second-semester Russian students was 1 (NL) and the top 25% were 3 (NH) or higher.

Table 4 Reading proficiency by language and class levels

	N	Min	Quartile 1	Median	Quartile 3	Max
Chinese	45					
Fourth semester	17	1	1	2	3	4
Third year	8	1	2.25	4.50	5	7
Returned missionaries	20	1	1	3	4.75	8
Russian	48					
Second semester	15	1	1	1	3	5
Fourth semester	7	1	2	3	4	6
Third year	2	5				6
Returned missionaries	24	1	5	7	10	10
Spanish	59					
Fourth semester	24	1	4	5	5	6
Returned missionaries	35	7	8	9	10	10

4.2 Vocabulary Scores

The maximum time allowed for the VLT was 25 min, approximately 5 min per band. Test takers, however, could spend as much time on any single band as they wanted. To determine the internal consistency of the various VLTs and to provide an overall reliability estimate, Cronbach's alpha was computed with the individual band scores as input. Cronbach's alpha is a measure of internal consistency, and it provides an estimate of the relationship between items. In this case, each band is considered an item. Cronbach's alpha examines how closely related these bands are, and if they can be considered to measure the same construct. In this sense, it can be considered to be a measure of scale reliability. Cronbach's alpha levels above 0.70 are considered to be acceptable levels. Table 5 provides Cronbach's alpha for the three languages. Table 5 also provides the correlations between each mastery criteria: 75%, 80%, and 85% correct and a composite score consisting of the summed individual band scores to determine the mastery criteria that correlates most strongly with the composite score. The maximum composite score for the five bands was 150 for Russian and Spanish and 120 for Chinese (4 bands only).

Table 5 shows that Cronbach's alpha was statistically significant and strong for all three languages, and particularly strong for Russian and Spanish, indicating high internal consistency and reliability of the three vocabulary tests. Table 5 also shows that the mastery criterion correlating best with the composite score was 80% for Chinese and Spanish. Because the mastery criteria 75% and 80% were almost identical in Russian, the criterion 80% was used for all three languages.

Table 6 shows the distribution of vocabulary levels using the 80% mastery criterion by language and class level.

Table 6 shows that second and fourth semester students generally did not yet have receptive mastery of the most frequent 1000 words of their respective languages. There were 2 out of 7 fourth-semester Russian students who had receptive mastery of the most frequent 1000 words, one out of 19 fourth-semester Spanish student who had a vocabulary level of 2000 words, and one fourth-semester Spanish student who had mastered the most frequent 4000 words.[2] Of the two Russian third-year students, only one had mastered the most frequent 1000 words.

Table 5 Cronbach's alpha computed between bands ($p < 0.05$) and Pearson's r correlations between composite vocabulary score and three mastery criteria: 75%, 80%, and 85%

	N	*alpha*	75%	80%	85%
Chinese	46	0.867	0.871*	0.892*	0.886*
Russian	48	0.951	0.960*	0.959*	0.923*
Spanish	54	0.950	0.957*	0.958*	0.934*

*Correlations were significant at $p < 0.01$.

[2] Students with prior exposure to the language are required to take a placement test, but there is no mechanism to exclude a student from enrolling in a level below their placement level. The instructor for this Spanish course confirmed that these two students seemed more advanced than fourth semester.

Table 6 Vocabulary levels by language and class level

	N	Min	Quartile 1	Median	Quartile 3	Max
Chinese	45					
Fourth semester	17	0	0	0	0	0
Third year	9	0	0	0	2000	4000
Returned missionaries	20	0	0	0	0	4000
Russian	48					
Second semester	15	0	0	0	0	0
Fourth semester	7	0	0	0	1000	1000
Third year	2	0				1000
Returned missionaries	24	0	1000	3000	5000	5000
Spanish	59					
Fourth semester	19	0	0	0	0	4000
Returned missionaries	35	3000	4000	4000	5000	5000

Both Russian and Spanish returned missionaries, however, returned with impressive vocabulary levels. The median for Russian was the 3000 band and for Spanish, the 4000 band. The top 25% of students in both Russian and Spanish had mastered the 5000 band. For Chinese, it was different. Sixteen out of 20 did not have receptive mastery of the most frequent 1000 words, 1 had mastered the most frequent 3000 words and 3 the most frequent 4000 words. This is most likely due to the fact that the VLT is a written test, and in the case of Chinese, it requires the ability to read Chinese characters. It is likely that the returned missionaries from China and Taiwan had much higher proficiency levels in speaking than in reading and that the VLT did not capture their oral vocabulary level, but only their written one.

4.3 Reading Proficiency and Vocabulary Levels

In the following, we present crosstabulations of vocabulary levels and reading proficiency and linear regression analyses to predict reading proficiency levels on the basis of vocabulary levels by language.

Table 7 shows that vocabulary levels of less than 1000 were associated with Novice and Intermediate levels and that the 4000 level appeared to be associated with the *Advanced* level (IH readers are almost at the *Advanced* level but inconsistently so).

A linear regression analysis was conducted to predict ACTFL ratings from the vocabulary score. Readers with proficiency levels below NH comprehend words and lists of words only and are unable to comprehend sentences or texts. Because we consider reading ability to involve textual understanding, these levels were excluded. Pearson's correlation between vocabulary score and reading proficiency was 0.843 with $p < .001$ ($N = 23$). The model explained 71.1% of the reading results ($R^2 = 0.711$). The linear regression analysis with reading proficiency as the dependent variable thus yielded a significant and large predictive effect of the

Table 7 Crosstabulation of Vocabulary and Reading Proficiency Levels – Chinese

		Reading proficiency								
		NL	NM	NH	IL	IM	IH	AL	AM	Total
Vocabulary level	Less than 1000	11	7	6	8	3				35
	1000					1				1
	3000					1		1		2
	4000						1	1	1	3
Total		11	7	6	8	5	1	1	1	41

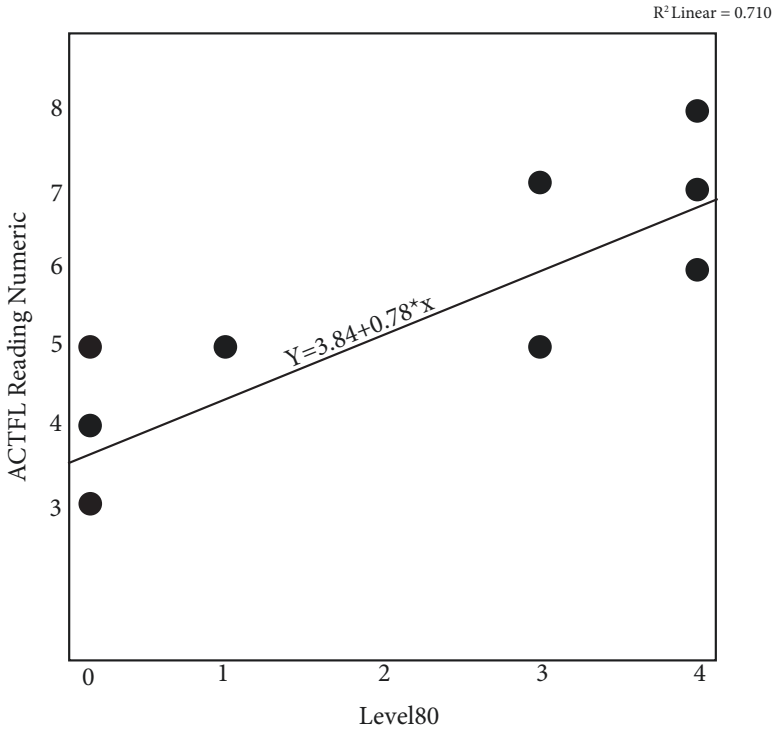

Fig. 1 Vocabulary size predicting reading proficiency levels – Chinese

vocabulary score on the reading proficiency rating: $p < .001$, Intercept (α): 3.84, Slope (β): 0.78. Figure 1 plots vocabulary and reading proficiency levels and includes the results of the regression analysis.

Table 8 (regression analysis) shows that vocabulary sizes of 1000 and 2000 predict the IM level in reading proficiency, while a vocabulary size of 3000 predicts IH and a vocabulary size of 4000 predicts *Advanced* levels of proficiency.

Table 9 shows that vocabulary levels of less than 1000 were associated with *Novice* and *Intermediate* levels. The 1000 level was associated with *Intermediate*,

Table 8 Predicting ACTFL reading proficiency levels on the basis of vocabulary size – Chinese

Vocabulary size	1000	2000	3000	4000
Reading proficiency numeric	4.62	5.40	6.18	6.98
ACTFL reading proficiency	IM	IM	IH	AL

Table 9 Crosstabulation of vocabulary and reading proficiency levels – Russian

		Reading proficiency									
		NL	NM	NH	IL	IM	IH	AL	AM	S	Total
Vocabulary level	Less than 1000	13	2	4	3	3					25
	1000				3		2				5
	2000					2			1		3
	3000					1		2			3
	4000						1	2		2	5
	5000							1	1	5	7
Total		13	2	4	6	6	3	5	2	7	48

the 2000 and 3000 levels with *Intermediate* and *Advanced*, the 4000 level with *Advanced* and *Superior*, and the 5000 level mostly with *Superior*.

A linear regression analysis was conducted to predict ACTFL ratings from the vocabulary score. Reading proficiency levels below NH were excluded. Pearson's correlation between vocabulary score and reading proficiency was .872 with $p < .001$ ($N = 33$). The model explained 76% of the reading results ($R^2 = 0.760$). The linear regression analysis with reading proficiency as the dependent variable thus yielded a significant and large predictive effect of the vocabulary score on the reading proficiency rating: $p < .001$, Intercept (α): 3.81, Slope (β): 1.06. Figure 2 plots vocabulary and reading proficiency levels and includes the results of the regression analysis.

Table 10 shows the results of the regression analysis.

Table 10 shows that vocabulary sizes of 1000 and 2000 predict the IM and IH levels, respectively, in reading proficiency, while vocabulary sizes of 3000, 4000, and 5000 predict AL, AM, and AH, respectively. The regression analysis thus supports the assumptions derived from the crosstabulation of vocabulary and reading levels.

Table 11 shows that that there were no students with reading proficiency levels below Intermediate.[3] Sixteen students who were Intermediate had a vocabulary level below 1000 words. A closer look at the data revealed that many of them had composite vocabulary scores that were similar to composite scores of the 1000 band in Russian (Spanish: Min = 34; Max = 81; Mean = 55; SD = 15.62; Russian: Min = 42; Max = 75; Mean = 59; SD = 12.98).[4] The student who had a vocabulary size of 2000 was one point short of receiving an IM rating. The 3000 and 4000 levels were mostly associated with *Advanced* and the 5000 level mostly with *Superior*.

[3] Two students who took the RPT were rated NL and NH, but neither of them took the VLT.

[4] Cf. Russian statistics for less than 1000: Min = 8; Max = 53; Mean = 33.84; SD = 15.42.

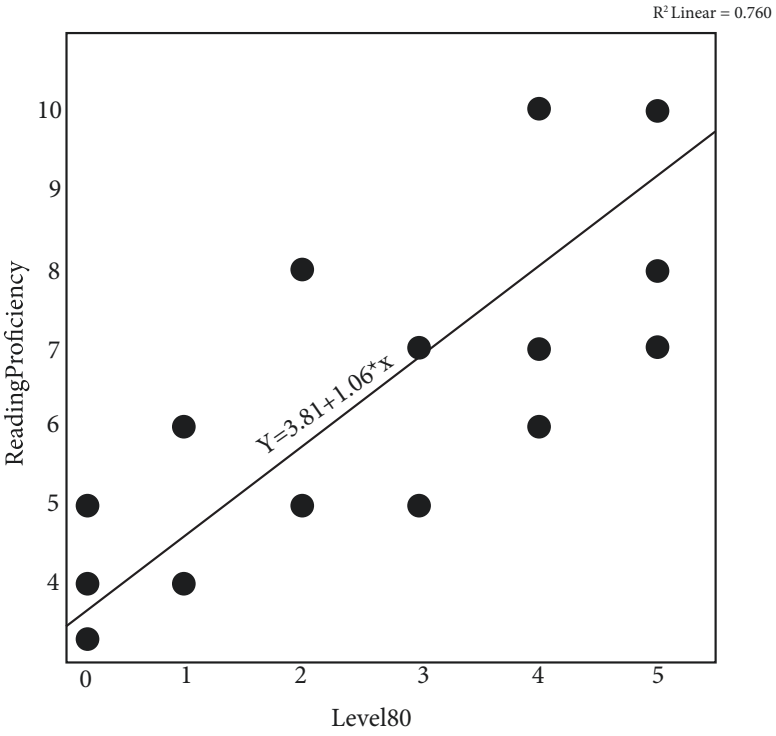

Fig. 2 Vocabulary size predicting reading proficiency levels – Russian

Table 10 Predicting ACTFL reading proficiency levels on the basis of vocabulary size – Russian

Vocabulary level	1000	2000	3000	4000	5000
Reading proficiency numeric	4.87	5.93	6.99	8.05	9.11
ACTFL reading proficiency	IM	IH	AL	AM	AH

Table 11 Crosstabulation of vocabulary and reading proficiency levels – Spanish

		Reading proficiency										
		NL	NM	NH	IL	IM	IH	AL	AM	AH	S	Total
Vocabulary level	Less than 1000				6	8	2					16
	2000				1							1
	3000							2		1	1	4
	4000							5	5	6	3	19
	5000							1	1	5	5	12
Total		0	0	0	7	8	2	8	6	12	9	52

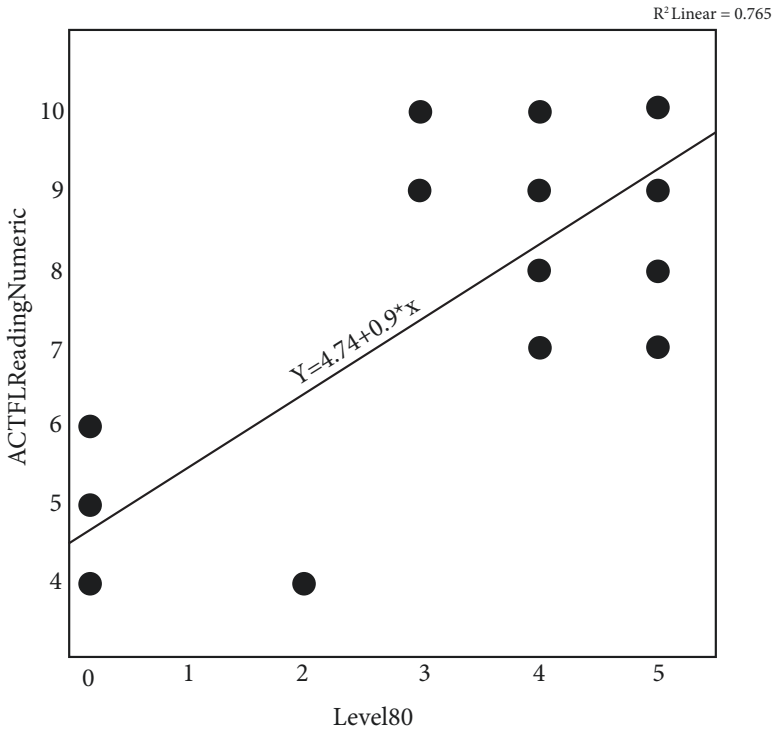

Fig. 3 Vocabulary size predicting reading proficiency levels – Spanish

Table 12 Predicting ACTFL reading proficiency levels on the basis of vocabulary size – Spanish

Vocabulary level	1000	2000	3000	4000	5000
Reading proficiency numeric	5.64	6.54	7.44	8.34	9.24
ACTFL reading proficiency	IH	AL	AL	AM	AH

A linear regression analysis was conducted to predict ACTFL ratings from the vocabulary score. There were no reading proficiency levels below NH. Pearson's correlation between vocabulary score and reading proficiency was .875 with $p < .001$ ($N = 52$). The model explained 76.5% of the reading results ($R^2 = 0.765$). The linear regression analysis with reading proficiency as the dependent variable thus yielded a significant and large predictive effect of the vocabulary score on the reading proficiency rating: $p < .001$, Intercept (α): 4.74, Slope (β): 0.90. Figure 3 plots vocabulary and reading proficiency levels and includes the results of the regression analysis.

Table 12 shows the results of the regression analysis.

Table 12 shows that vocabulary sizes of 1000 and 2000 predict the IH and AL levels, respectively, in reading proficiency, while vocabulary sizes of 3000, 4000, and 5000 predict AL, AM, and AH, respectively. While the reading proficiency levels

of the 1000 and 2000 bands appear to be inflated, probably due to the large number of below 1000 associated with intermediate reading proficiency, the results of the 3000–5000 bands are similar to Russian, and only slightly higher than Chinese. Note also that the 1000 and 2000 bands barely predict IH and AL reading proficiencies, being very close to the midpoint between IM and IH, and IH and AL, respectively.

5 Discussion

The present study showed very high correlations between reading proficiency and receptive vocabulary size (.843-Chinese; .872-Russian; .875-Spanish). Vocabulary size thus accounted for 71% (Chinese), 76% (Russian), and 76.5% (Spanish) of reading proficiency, and vice-versa, reading proficiency explained the same percentages of vocabulary size. These are very large effect sizes. Furthermore, crosstabulations and linear regression analyses showed that vocabulary sizes of the most frequent 1000 and 2000 words were generally associated with the ACTFL *Intermediate* level, while vocabulary sizes of 3000 and 4000 were associated with the ACTFL *Advanced* level. A vocabulary size of 5000 was associated with the ACTFL *Superior* level. Because student vocabulary sizes and reading proficiency levels were unevenly distributed between and within languages, it is even more remarkable that such a clear pattern emerged across three very different languages.

In general, the vocabulary sizes of the college students participating in the study were not very impressive. Even after 2 years of foreign language study, the vocabulary sizes of the students did not include even the most frequent 1000 words of their respective second language, particularly for Chinese and Russian. These low vocabulary sizes seem to be directly related to low reading proficiency ratings. The median after four semesters was NM in Chinese and NH in Russian with the top 25% of students reaching NH and higher in Chinese and IL and higher in Russian. While most Spanish students were solidly *Intermediate* at the end of 2 years, they had not quite mastered the first 1000 band, which may have precluded them from reaching higher proficiency levels.

Immersion learners (returned missionaries) who spent between 18 and 24 months in a country where the target language was spoken fell into two groups. Russian and Spanish students returned with impressive vocabulary sizes and high levels of reading proficiency. The Russian median was 3000 words, while the Spanish median was 4000 words. The top 25% of students of both languages had mastered at least 5000 words. This correlated with high reading proficiency levels. The median for Russian immersion learners was AL, for Spanish immersion learners, it was AH. The top 25% of students in both languages scored at the *Superior* level. For Chinese, it was different. 16 out of 20 of the immersion learners had not mastered the 1000 most frequent words of Chinese. As mentioned previously, this was most likely due to the fact that VLT was a written test, capturing only written vocabulary knowl-

edge. The other four had levels of 3000 and 4000, and possibly higher than 4000.[5] Overall, the median reading proficiency for this group was NH, lower than the median reading proficiency of regular third-year Chinese students, which was 4.50 (IL to IM). Thus, in general, third-year Chinese students were better in both reading proficiency and written vocabulary knowledge. The top 25% of third-year students had a reading vocabulary size of at least 2000 words.

The difference in reading level and vocabulary knowledge between the Russian and Spanish students on the one hand, and the Chinese students on the other, is striking. While the Chinese students had acquired fairly high levels of speaking proficiency while abroad, their literacy did not show similar development. Why this should be the case is beyond the scope of this study, but the explanation most likely lies in the difference in writing systems. The Russian and Spanish learners accessed their L2 through an alphabetic system as in their native language. The Russian learners did have the additional challenge of learning the Cyrillic alphabet, nonetheless, the essential principle of grapheme to sound correspondence remains consistent across the two alphabets. By contrast, the Chinese learners had to shift from their familiar alphabetic orthography to the character based Chinese writing system, the mastery of which requires extensive time and memorization. It is likely that the Chinese learners have greater vocabulary knowledge than was measured. The VLT is a written test. Had they been tested auditorially, they may well have scored better. The traditional third-year students had *Intermediate* levels of reading proficiency as well as higher vocabulary scores, results consistent with classroom learning that focuses on developing knowledge of Chinese characters.

6 Curricular and Pedagogical Implications

Upper division language curricula aim to introduce students to the literature of the target culture. As noted earlier, research has shown that a reader must know 95–98% of a text's vocabulary in order to understand the text. If the participants in this study are typical, and we focus in particular on students who did not have an extended immersion experience, we see that authentic literary texts are beyond the reach of many. Students do not have the necessary vocabulary knowledge to read at the *Advanced* or *Superior* level. The challenge for programs and instructors is how to promote vocabulary learning so that students achieve higher levels of reading proficiency.

Grabe (2009) makes a number of suggestions for vocabulary instruction based on findings from research on vocabulary acquisition. These include, for example, reading aloud to students and drawing their attention to keywords while reading; teaching a limited set of key words for depth, precision and multiple encounters; focusing on word relationships (parts-of-speech variations, word families, synonyms,

[5] The three students who had mastered the most frequent 4000 words had the three highest reading proficiency ratings (IH, AL, and AM), while the student who had 3000 words was rated IM.

antonyms, graded relations); working with dictionary definitions to rewrite more accessible definitions (pp 283–284). A number of other studies have pointed out the benefits of explicit vocabulary instruction, for example the effects of helping students recall and produce newly-learned words (Lee & Muncie, 2006; Lin & Hirsh, 2012; Webb, 2009). Strategies such as these appear to be at odds with the recommendations of the most frequently used L2 methods textbooks. As discussed above, the communicatively oriented ethos that underpins L2 methods textbooks privileges contextualized presentation of vocabulary with a focus on developing reading ability. For example, Lee and VanPatten (2003) caution against memorization in place of meaning-based comprehensible input, and yet, there is recent research to suggest that memorization may deserve reconsideration. A study comparing vocabulary learning via rote memorization on the one hand, and semantic mapping on the other, showed "no significant difference … between the vocabulary mean scores of the two groups on the post-test at the end of the four-month treatment period." (Khoii & Sharififar, 2013, 206). More empirical research on the efficacy of a variety of vocabulary learning strategies is needed.

7 Conclusion

The current study demonstrates that there is a strong correlation between vocabulary size and reading proficiency and it also shows that the level of L2 vocabulary that students acquire during a typical undergraduate program is not substantial. In its 2009 report to the Teagle Foundation, the Modern Language Association (MLA) acknowledged the crucial role of the study of texts for an undergraduate degree in languages and insisted that "the most beneficial among these are literary works, which offer their readers a rich and challenging—and therefore rewarding—object of study" (p. 4). As discussed in the introduction, a reader needs to know between 95% and 98% of the words in a text in order to comprehend it. This implies that access to original literary texts in the target language requires a vocabulary size that is probably beyond the reach of most undergraduate majors. The curricula of undergraduate degrees in languages typically include a number of required upper-level courses built around literary texts. The assumption is that advanced undergraduates should be able to engage in the critical reading of these texts and that exposure to them will improve their vocabulary and their command of the language. However, language learners face a disheartening conundrum: while incidental acquisition of a new lexical item requires multiple encounters with it, words beyond the most common 3000 are so infrequent that ordinary reading is not enough to learn them. Yet, explicit attention to vocabulary acquisition is often only a tangential focus of the introductory curriculum and may be altogether absent from advanced literary courses. If the ability to engage with literary texts is to remain a central goal of the language major, a renewed and revised attention to vocabulary building is necessary.

References

ACTFL. (2012). *ACTFL proficiency guidelines*. Alexandria, VA: ACTFL. Available from https://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012

ACTFL. (2013). *ACTFL reading proficiency test* (*RPT*). *Familiarization manual and ACTFL proficiency guidelines 2012—Reading*. Retrieved June 8, 2016, from http://www.languagetesting.com/wp-content/uploads/2015/02/ACTFL_FamManual_Reading_2015.pdf

ACTFL. (2016). *Assigning CEFR ratings to ACTFL assessments*. Retrieved July 12, 2017, from https://www.actfl.org/sites/default/files/reports/Assigning_CEFR_Ratings_To_ACTFL_Assessments.pdf

Bernhardt, E. (2011). *Understanding advanced second-language reading*. New York, NY: Routledge.

Brandl, K. (2008). *Communicative language teaching in action: Putting principles to work*. Upper Saddle River, NJ: Pearson.

Carver, R. P. (1994). Percentage of unknown vocabulary words in text as a function of the relative difficulty of the text: Implications for instruction. *Journal of Reading Behavior, 26*(4), 413–437. https://doi.org/10.1080/10862969409547861

Grabe, W. (2009). *Reading in a second language: Moving from theory to practice*. New York, NY: Cambridge University Press.

Hacking, J., & Tschirner, E. (2017). Reading proficiency, vocabulary development and curricular design: The case of college Russian. *Foreign Language Annals, 50*(3), 1–19. https://doi.org/10.1111/flan.12282

Hirsh, D., & Nation, I. S. P. (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language, 8*(2), 689–696. Available at: http://nflrc.lll.hawaii.edu/rfl/PastIssues/rfl82hirsh.pdf

Hu, M., & Nation, I. S. P. (2000). Unknown vocabulary density and reading comprehension. *Reading in a Foreign Language, 13*(1), 403–430.

IELTS. (n.d.). *IELTS writing mark schemes*. Retrieved April 20, 2018, from https://www.examenglish.com/IELTS/IELTS_Writing_MarkSchemes.html

Institute for Test Research and Test Development. (2013). *Assessing evidence of validity of the ACTFL reading proficiency test* (*RPT*). Retrieved June 8, 2016, from http://www.languagetesting.com/wp-content/uploads/2013/10/Technical-Report-ACTFL-RPT-for-publication.pdf

Khoii, R., & Sharififar, S. (2013). Memorization versus semantic mapping in L2 vocabulary acquisition. *English Language Teaching Journal, 67*(2), 199–209. https://doi.org/10.1093/elt/ccs101

Laufer, B., & Ravenhorst-Kalovski, G. C. (2010). Lexical threshold revisited: Lexical coverage, learners' vocabulary size and reading comprehension. *Reading in a Foreign Language, 22*(1), 15–30.

Lee, J. F., & Van Patten, B. (2003). *Making communicative language teaching happen*. Boston, MA: McGraw-Hill.

Lee, S. H., & Muncie, J. (2006). From receptive to productive: Improving ESL learners' use of vocabulary in a postreading composition task. *TESOL Quarterly, 40*(2), 295–320. https://doi.org/10.2307/40264524

Lin, C.-C., & Hirsh, D. (2012). Manipulating instructional method: The effect on productive vocabulary use. In D. Hirsh (Ed.), *Current perspectives in second language vocabulary research* (pp. 117–148). New York, NY: Peter Lang.

Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol, UK: Multilingual Matters.

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. New York, NY: Heinle and Heinle.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening. *Canadian Modern Language Review, 63*(1), 59–82. https://doi.org/10.3138/cmlr.63.1.59

Nation, I. S. P. (2007). Fundamental issues in modelling and assessing vocabulary knowledge. In H. Daller, J. Milton, & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 35–43). Cambridge: Cambridge University Press.

Nation, I. S. P. (2014). How much input do you need to learn the most frequent 9,000 words? *Reading in a Foreign Language, 26*(2), 1–16.
Omaggio-Hadley, A. (2001). *Teaching language in context*. Boston, MA: Heinle.
Rifkin, B. (2005). A ceiling effect in traditional classroom foreign language instruction: Data from Russian. *Modern Language Journal, 89*(1), 3–18. https://doi.org/10.1111/j.0026-7902.2005.00262.x
Schmitt, N. (2008). Review article. Instructed second language vocabulary learning. *Language Teaching Research, 12*(3), 329–363. https://doi.org/10.1177/1362168808089921
Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal, 95*(1), 26–43. https://doi.org/10.1111/j.1540-4781.2011.01146.x
Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing, 18*(1), 55–88. https://doi.org/10.1191/026553201668475857
Shrum, J. L., & Glisan, E. W. (2010). *Teacher's handbook: Contextualized language instruction*. Boston, MA: Heinle.
Staehr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal, 36*(2), 139–152. https://doi.org/10.1080/09571730802389975
Terrell, T. (1986). Acquisition in the natural approach: The binding/access framework. *The Modern Language Journal, 70*(3), 213–227. https://doi.org/10.1111/j.1540-4781.1986.tb05266.x
Tschirner, E. (2004). Breadth of vocabulary and advanced English study: An empirical investigation. *Electronic Journal of Foreign Language Teaching, 1*, 26–38. Available at: http://www.itt-leipzig.de/static/literatur/Tschirner_2004_Breadth_of_Vocabulary.pdf
Webb, S. (2009). The effects of pre-learning vocabulary on reading comprehension and writing. *Canadian Modern Language Review, 65*(3), 441–470. https://doi.org/10.3138/cmlr.65.3.441
Xing, P., & Fulcher, G. (2007). Reliability assessment for two versions of vocabulary levels tests. *System, 35*(2), 181–191. https://doi.org/10.1016/j.system.2006.12.009

Jane F. Hacking is Associate Professor of Russian and Linguistics at the University of Utah, where she Co-Directs the Second Language Teaching and Research Center (L2TReC). Her research focuses on L2 phonology and the overall development of L2 proficiency. She received the 2017 award for Outstanding Contribution to the Profession from the American Association of Teachers of Slavic and East European Languages.

Fernando Rubio is Professor of Spanish Linguistics at the University of Utah, where he also serves as the Co-Director of the Second Language Teaching and Research Center. His research interests include second language acquisition, language teaching methodology, and proficiency assessment. From 2014 to 2018 he served as the PI in the Language Flagship Proficiency Initiative grant at the University of Utah.

Erwin Tschirner is Gerhard Helbig Professor of German as a Foreign Language at the University of Leipzig, Germany. His main research areas are language testing and second language acquisition, primarily the acquisition of vocabulary and grammar and reading and listening proficiency.