Investigating the Role of Vocabulary Size in Second Language Speaking Ability

Takumi Uchihara (University of Western Ontario)
Jon Clenton (Hiroshima University)

Abstract

The current study investigates the extent to which receptive vocabulary size test scores can predict second language (L2) speaking ability. Forty-six international students with an advanced level of L2 proficiency completed a receptive vocabulary task (Yes/No test; Meara & Miralpeix, 2017) and a spontaneous speaking task (oral picture narrative). Elicited speech samples were submitted to expert rating based on speakers’ vocabulary features as well as lexical sophistication measures. Results indicate that vocabulary size was significantly associated with vocabulary rating. However, learners with large vocabulary sizes did not necessarily produce lexically sophisticated L2 words during speech. A closer examination of the data reveals complexities regarding the relationship between vocabulary knowledge and speaking. Based on these findings, we explore implications for L2 vocabulary assessment in classroom teaching contexts and provide important suggestions for future research on the vocabulary-and-speaking link.

Keywords: vocabulary size, receptive vocabulary, L2 speaking, L2 vocabulary
Investigating the Role of Vocabulary Size in Second Language Speaking Ability

Introduction

Vocabulary is widely-researched in second language (L2) teaching and learning, regarded as fundamental to overall language proficiency and development (Nation, 2013). For L2 instruction, less time-consuming and user-friendly vocabulary testing methodologies fulfil an important role for diagnostic, placement, and screening purposes, resulting in improved pedagogy practices (Alderson, 2005; Nation & Chung, 2009). The usefulness of testing vocabulary is based on the assumption that vocabulary knowledge serves as an indicator of language proficiency (Milton, 2013; Nation, 2013), as supported by empirical work designed to investigate potential relationships between vocabulary measures and second language proficiency (e.g., IELTS; Milton, Wade, & Hopkins, 2010), as well as between vocabulary knowledge and each of the four language skills (e.g., reading, Laufer & Aviad-Leviztky, 2017; listening, Noreillie, Kestemont, Heylen, Desmet, & Peters, 2018; writing, Schoonen, van Gelderen, Stoel, Hulstijn, & de Glopper, 2011; speaking, Koizumi & In’nami, 2013). However, fewer studies explore the relationship between vocabulary and speaking (e.g., De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012; Koizumi & In’nami, 2013) compared to the relationship between vocabulary and the other language skills (e.g., Laufer & Aviad-Leviztyk, 2017; Noreillie et al., 2018; Qian, 2002; Schoonen et al., 2011; Stæhr, 2008, 2009; Wang & Treffers-Daller, 2017). The aim of the current study is to gain further insight into the relationship between vocabulary knowledge and L2 speaking proficiency. Based on our findings, we offer pedagogical implications for vocabulary assessment in L2 classroom settings.

Vocabulary Size Measures

Defining the construct of word knowledge is notoriously challenging due to its multidimensional nature. Researchers conceptualize and describe word knowledge in different ways, adopting componential (Nation, 2013), developmental (Henriksen, 1999), and holistic (Meara, 2009) approaches. A common way of capturing the lexical construct across researchers is to measure vocabulary size (i.e., the number of known words in terms of form-meaning connections) at “receptive” and “productive” dimensions. Although a developmental trajectory of these types of knowledge remains open to question (Meara, 1997; Melka, 1982), it seems ecologically valid and practical for language teachers to view receptive knowledge as the knowledge necessary for reading and listening, with productive knowledge as the knowledge required for writing and speaking (Schmitt, 2010). This view of a receptive-and-productive distinction has to date informed vocabulary test development, evidenced by a large number of receptive and productive vocabulary measures currently available (see Beglar & Nation, 2014; Meara & Miralpeix, 2017; Read, 2000; Schmitt, 2010; Webb &
Vocabulary Knowledge and Second Language Speaking Proficiency

Extant research emphasizes vocabulary knowledge and other language skills: reading (e.g., Laufer & Aviad-Levitzky, 2017) and listening (e.g., Noreillie et al., 2018), highlighting the lack of research designed to explore the relationship between vocabulary knowledge and productive language skills such as speaking (Koizumi, 2013). Research suggests that relatively strong associations are consistently found between receptive vocabulary size and reading skills ($r = .52$ to $.91$; Laufer & Aviad-Levitzky, 2017; Milton et al., 2010; Miralpeix & Muñoz, 2018; Qian, 2002; Stæhr, 2008; $r = .79$ from meta-analysis conducted by Jeon & Yamashita, 2014) as well as between size and listening skills ($r = .42$ to $.70$; Milton et al.,
An emerging number of studies suggest an important relationship between vocabulary size and L2 speaking proficiency. De Jong and colleagues (2012) used controlled productive vocabulary measures (sentence completion) focusing on upper-intermediate to advanced L2 learners and found that two variables—productive vocabulary knowledge and pronunciation accuracy (intonation)—were strong predictors of general speaking proficiency. Koizumi and In’nami (2013) identified the predictive role of productive vocabulary knowledge (measured via L1-L2 translation) in various aspects of L2 oral ability (fluency, accuracy, syntactic complexity) with novice and lower-intermediate learners studying English as a foreign language (EFL). Uchihara and Saito (2016) found that EFL learners’ productive vocabulary knowledge (measured via a free word association task) significantly related to perceived oral fluency.

Unlike studies using productive vocabulary measures, studies using receptive vocabulary measures suggest a more complex picture of the vocabulary-and-speaking link. De Jong and Mora (2017) measured the receptive vocabulary size of EFL learners using Yes/No tests \((M_{size} = 6,144 \text{ words, \ Range = 3,350 to 8,200})\). These results suggest that vocabulary size is associated with speed fluency \((r = -.311)\) but not with other fluency measures (breakdown fluency). Milton et al. (2010) used aural and written versions of Yes/No tests (measuring up to 5,000 frequency level) to assess receptive vocabulary size of ESL learners \((M_{size} = 2,844 \text{ words})\) and compared the vocabulary scores to the IELTS tasks designed to elicit the four language skills. Their results show that a written form of a receptive vocabulary measure does not relate to speaking, but they did find a significant correlation with an aural variant of the same vocabulary measure \((r = .71)\). A recent study (Miralpeix & Muñoz, 2018) targeted advanced EFL learners studying in English-medium instruction \((M_{size} = 5,127 \text{ words, \ Range = 2,500 to 7,200})\) and examined the relationships between Yes/No test scores and various aspects of oral ability (fluency, pronunciation, lexicogrammar). Interestingly, a correlation with lexicogrammar rating was found to be medium in effect size \((r = .413)\), slightly smaller than that with oral fluency \((r = .485)\).

Based on these papers, we might argue that vocabulary size should be associated with lexical use in production because larger L2 lexicons allow for the selection and retrieval of conceptually and contextually appropriate words in communication. One possible reason for the lack of a strong link between vocabulary size and lexical production might relate to the view that vocabulary size is not considered as important for spoken discourse as for written discourse. Because of the diminished value of low-frequency words in spoken discourse compared with written discourse, learning new words beyond a certain level of mastery (the 2,000 to 3,000 most frequent word families) might be less beneficial for improvement of oral performance (Milton, 2009, 2013). A further complication pertaining to the relationship between
vocabulary and speaking is that productive vocabulary use is moderated by the individual. When learners fail to produce lexically rich texts (e.g., infrequent words), we might simply attribute such phenomena to inadequate vocabulary knowledge, but this might be confounded by factors such as motivation and willingness to respond (Nation & Webb, 2011), or to avoidance strategies (Skehan, 2009).

Motivation for the Current Study

Although the primary role of vocabulary knowledge in L2 speaking for low-proficiency L2 learners has been investigated and supported empirically (Koizumi & In’nami, 2013), the extent to which it applies to advanced L2 learners remains unclear. As highlighted above, advanced EFL learners with large vocabulary sizes ($M > 5,000$ words) might perform L2 speech production relatively well (Miralpeix & Muñoz, 2018) or might not be orally proficient at all or, if any, proficient in a limited aspect (De Jong & Mora, 2017). Unlike earlier studies targeting EFL learners, one study (Milton et al., 2010) examined the vocabulary-and-speaking relationship with learners studying English as a second language (ESL), finding a medium but non-significant correlation ($r = .35, p > .05$) between written vocabulary size and L2 speaking. Since ESL learners likely receive considerable L2 input outside the classroom, participants in their study might be regarded proficient as determined by the authors reporting their ability as ranging “between intermediate and relatively advanced” (p. 88). With this assumption, the findings appear to indicate a non-linear relationship between vocabulary size and L2 speaking ability for high proficiency learners. However, several limitations with the study needs to be noted before drawing any strong conclusions. Primarily, Milton and colleagues measured a limited range of vocabulary size (up to 5,000 words) with a small sample size ($N = 30$). Also, participants in their study may not have benefited from input-rich naturalistic environments as they were tested before starting their university degrees. Lastly, their study, in addition to the majority of studies of this kind (De Jong et al., 2012; Harrington & Carey, 2009; Miralpeix & Muñoz, 2018; Uchihara & Saito, 2016) tend to rely on human rating for speech measures, apart those that measure oral fluency (De Jong & Mora, 2017; Hilton, 2008; Koizumi & In’nami, 2013). Considering confounding influences from individual variability across human rater's (Li & Lorenzo-Dus, 2014), objective measures involving quantifying and counting given linguistic features serve as an important addition to assess oral proficiency.

The current study is therefore designed to address the aforementioned research gaps to explore the vocabulary-and-speaking relationship by targeting advanced ESL adult learners with considerable L2 experience (length of residence > 7 months). The wide range of receptive vocabulary size was measured (up to 10,000 words), and objective measures along with human rating were employed to assess L2 oral proficiency. The study was guided by the following question: To what extent is advanced ESL learners’ vocabulary size associated
Method

Participants

The participants were forty-six international students (28 females and 18 males, aged between 18 and 51 years; $M = 29.11$) from 15 different countries – Japan (14), China (7), Kazakhstan (5), Saudi Arabia (4), Turkey (3), Taiwan (2), Thailand (2), and the remainder (9). The majority of the participants were postgraduate students (2 undergraduates and 2 doctoral students were included in the current study) at a UK university. In terms of L2 proficiency, we considered the participants to be advanced L2 learners of English (B2 to CI levels) because they had lived in the UK for at least 7 months at the time of testing (mean length of residence = 19.6 months) and 26 participants had achieved an IELTS score of at least the 6.5 required for university admission. The remaining 20 participants who did not reach the admission criteria were required to take pre-sessional EAP courses prior to the beginning of the academic year, so all had achieved an equivalent score of at least IELTS 6.5 based on the in-house university entrance examination (TEEP, 2014).

Vocabulary Measure

To measure receptive vocabulary knowledge, we used the Yes/No format task based on the Eurocentres Vocabulary Size Test (EVST) developed by Meara and Jones (1988). The rationale for choosing the Yes/No format is that the test has been extensively validated and established as a reliable vocabulary measure particularly for higher proficiency learners (Meara & Miralpeix, 2017). Also, the simplicity of the task format attracts test users in various academic contexts including university foundation programs (Roche et al., 2016) and undergraduate university programs (Harrington & Roche, 2014). Further, using the Yes/No test enables us to compare previous studies reporting correlations between Yes/No test scores and L2 speaking proficiency (De Jong & Mora, 2017; Milton et al., 2010; Miralpeix & Muñoz, 2018).

Yes/No Test. We employed V _YesNo, an updated version of the EVST (Meara & Jones, 1988), accessible at the Lognostics Tools website (http://www.lognostics.co.uk/tools/). This test is based on the EVST using the same format and target items, but different in two ways (Meara & Miralpeix, 2017): (a) the number of imaginary words is equal to that of real words and (b) V _YesNo is not computer adaptive (i.e., the EVST used to begin with the most frequent words and gets progressively more difficult with less frequent words). The V _YesNo uses 200 target items (100 real and 100 imaginary words) and measures knowledge of words up to a ceiling of 10,000 words from the Thorndike and Lorge (1944) list. The imaginary words are included to correct the final score for guessing. The number of incorrectly identified words (YES responses to imaginary words) in relation to that of correctly identified words (YES responses to real words) determines the extent to which the
L2 VOCABULARY AND SPEECH

final score is adjusted downwards (see Meara & Miralpeix, 2017, p. 120 for the equation to calculate adjusted scores). In the current study, participants saw a target item on a computer screen and were asked to click on “YES” when they know the meaning of the word and “NEXT” when not knowing or not sure about the meaning. The participants were also forewarned that the test items include imaginary words and should not answer YES to them.

L2 Oral Measures

We adopted both subjective and objective measures to assess L2 oral proficiency based on speakers’ lexical use in spontaneous production. As for subjective assessment, human rating was adopted as it is widely used to assess spoken lexical proficiency in the area of standardized L2 proficiency testing (e.g., TOEFL, IELTS) as well as L2 speech research (Saito, 2018; Saito, Trofimovich, & Isaacs, 2017; Saito, Webb, Trofimovich, & Isaacs, 2016). Regarding objective assessment, we measured lexical sophistication based on word frequency as an index (Read, 2000). Word frequency in a spoken or written text has long been considered as a reliable indicator of learners’ lexical proficiency (e.g., Crossley, Salsbury, McNamara, & Jarvis, 2011; Laufer & Nation, 1995). The importance of including both subjective and objective measures has been emphasized in L2 speech research in order to capture a fuller picture of learners’ L2 oral ability (Pallotti, 2009).

Speech material. A picture narrative task was adopted to elicit speech samples. Picture narrative tasks have been extensively used in L2 speech research and found suitable for eliciting and assessing lexical features in spoken texts (Saito, 2018; Saito et al., 2016, 2017; Skehan, 2009) from advanced L2 learners (IELTS 7.0; Li & Lorenzo-Dus, 2014). Comparing the effects of several tasks on lexical use in L2 oral production, Skehan (2009) suggested the utility of narrative task as it can “push second language speakers … into using less frequent lexis, presumably as they are responding to the events within the narrative” (p. 517). We selected four cartoon strips (Society for Testing English Proficiency, 2015, p. 120) featuring a story of protests against the construction of a mall, including words and sentences to facilitate the narrative (indicated by double quotations) as follows: (a) A couple find two construction workers placing a signpost noting “new shopping mall coming soon,” the woman says “we need to do something,” and both appear to consider taking action against the construction of the mall, (b) “few days later” the couple start a campaign, the man holds a placard reading “protect nature,” and they ask local people to sign a petition against the upcoming construction, (c) “three months later” a notice of “cancelled” is placed on the signpost noting the construction, and (d) “a month later” the couple look upset as they find a local newspaper headline reporting that “unemployment is on the rise.”

All descriptions were recorded individually in a quiet office, and stored on a digital voice recorder. The length of the speech samples ranged between one and four minutes and the average number of words in production ranged from 81 to 319 words, which was
L2 VOCABULARY AND SPEECH

comparable to earlier L2 speech studies assessing lexical performance (e.g., Range = 61 to 268 words in Saito, 2018). Speech samples were elicited with no explicit time limit set in advance.

**Human rating.** IELTS speaking band descriptors were adopted to rate the speech samples for two main reasons: (a) IELTS speaking descriptors are widely used and found to be valid in various studies (e.g., Read, 2005) and (b) the raters of the current study were considered familiar with the IELTS speaking descriptors. IELTS speaking descriptors consist of four linguistic components (fluency, vocabulary, grammar, and pronunciation) (available from https://takeielts.britishcouncil.org/sites/default/files/IELTS_Speaking_band_descriptors.pdf). In the current study, the vocabulary component (i.e., Lexical Resource) was used to assess L2 oral proficiency.

Three L1 (English) speech raters were recruited (2 females, 1 male) at a UK university. All raters were part of a university English language proficiency assessment team, and considered by the researchers to have sufficient rating experience of a variety of L2 learner speech, as reported in background questionnaires conducted prior to the rating sessions. None of the raters reported any hearing difficulties. To rate all speech samples, the raters were first given a few minutes to study IELTS speaking descriptors and then rated a training set of four speech samples to familiarize themselves with what was required of their judgements. Subsequently, the raters listened to speech samples from each participant in a random order. While listening to one sample at a time from the beginning to the end, each of the raters assigned a rating score (from 1 to 9) based on the speaking descriptors.

**Lexical sophistication.** Elicited speech samples were analyzed in terms of lexical frequency (Read, 2000). Prior to a series of lexical analyses, the speech samples were transcribed and cleaned with all orthographic markings of pausing (e.g., uh, um, oh, ehh) removed and obvious pronunciation errors fixed (e.g., the story for the stoly). We adopted two approaches to assessing lexical sophistication: band-based and count-based word frequency measurement (Crossley, Cobb, & McNamara, 2013). According to the band-based approach, L2 texts elicited through the narrative task were profiled according to frequency levels (Laufer & Nation, 1995). Given that the most frequent first and second 1,000 (K1 and K2) words are considered as high frequency items (Nation, 2013), the third or above 1,000 frequency (K3+) words were rated as sophisticated items (i.e., low frequency words). The proportion of K3+ in the texts were calculated using Vocab Profile in the Compleat Lexical Tutor (available at https://www.lextutor.ca/vp/comp/). In addition to lexical frequency profiling analysis (Laufer & Nation, 1995), we also adopted a count-based method, which gives a corpus-based frequency count to each individual item used in production and calculates an average frequency score per text (Crossley et al., 2013). With the Tool for the Automatic Analysis of Lexical Sophistication (TAALES; Kyle & Crossley, 2015), lexical
L2 VOCABULARY AND SPEECH

frequency scores were calculated as the sum of log-transformed frequency scores (in reference to the Corpus of Contemporary American English) divided by the number of word tokens receiving frequency scores. Lower frequency scores indicate that L2 learners likely include less frequent words in their texts, suggesting that they have more advanced L2 lexical knowledge (e.g., Crossley et al., 2011).

Procedure

We conducted testing sessions individually, with one sitting per participant with one researcher (the first author). The participants were first required to complete consent forms and language background questionnaires. Participants then completed the speaking (picture narrative) task and the Yes/No test. For the speaking task, they were instructed to describe a four-strip picture narrative immediately after studying the pictures to familiarize themselves with the story for one minute. They were allowed to look at the pictures while describing it and given as much time as they wanted for the task. Lastly, the computerized version of the Yes/No test was completed in approximately 10 minutes. Subsequent to administration of the three tests to all participants, the speech samples were evaluated by the raters, based on IELTS speaking descriptors. Through an individual appointment made by the researcher with each of the three raters, they were asked to rate speech samples in the researcher’s presence. To avoid fatigue effects, raters evaluated 46 samples in two or three separate meetings. Each rater took approximately three hours in total for rating all samples.

Results

Vocabulary and Speaking Measures

The descriptive statistics of the Yes/No test scores show that participants were of relatively high L2 proficiency with a receptive vocabulary size of approximately 6,000 words on average ($M = 6,058.1, SD = 1,247.2$). The data also reveal a considerable variation in vocabulary size across participants ($Range = 3,100$ to $8,770$). The participants in the study were considered to have larger receptive vocabulary sizes than those with 5,000 words ($M = 5,127$) in Miralpeix and Muñoz (2018) and comparable to those with 6,000 words ($M = 6,144$) in De Jong and Mora (2017).

Regarding the vocabulary rating scores, participants were overall considered lexically proficient ($M = 5.9, SD = 0.8$) according to Band 6 in the IELTS descriptor, indicating the speaker “has a wide enough vocabulary to discuss topics” and “generally paraphrases successfully” despite some “inappropriacies.” A wide variability across speakers ($Range = 4.3$ to $7.3$) suggests that some can only “talk about familiar topics” and make “frequent errors in word choice” (Band 4) and others can use “vocabulary resource flexibly” and “some less common and idiomatic vocabulary” with “awareness of style and collocation” (Band 7). A Cronbach reliability analysis indicates that inter-rater agreement was acceptable ($\alpha = .77$).
L2 VOCABULARY AND SPEECH

exceeding a minimum benchmark value in L2 research (α > .70; Larson-Hall, 2010). Pearson correlations among the three spoken lexical measures, vocabulary rating, average frequency (M = 3.1, SD = 0.1), 3K+ proportion (M = 5.3, SD = 2.2), show a significant correlation between average frequency and 3K+ proportion (r = -.471, p = .001), whereas no significant correlation was found between rating and average frequency (r = .172, p = .253) or rating and 3K+ proportion (r = .153, p = .311). The absence of the significant link between rating and sophistication measures suggests that speakers who were judged to be lexical proficient by human raters did not necessarily produce lexically sophisticated L2 words in spontaneous speech.

In order to further inspect the range of lexical items elicited through the picture narrative task used in this study, the result of lexical profiling analysis is presented in Table 1. The result shows high frequency words accounted for the majority of the words used in oral narratives (95% in total). The remaining 5% were considered low frequency words. Compared to the proportion of 3K+ words elicited through writing tasks in Laufer and Nation (1995) (i.e., 16% and 20%), the smaller percentage of 3K+ words were elicited in the current study.

Table 1.

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>K1 words</td>
<td>90.9</td>
<td>2.8</td>
<td>82.7</td>
<td>96.0</td>
</tr>
<tr>
<td>K2 words</td>
<td>3.8</td>
<td>2.0</td>
<td>0.6</td>
<td>9.4</td>
</tr>
<tr>
<td>K3+ words</td>
<td>5.3</td>
<td>2.2</td>
<td>1.7</td>
<td>12.3</td>
</tr>
</tbody>
</table>

Note. K1 = the first thousand frequency words; K2 = the second thousand frequency words; K3+ = the third thousand frequency or above words.

Relationship between Receptive Vocabulary Size and L2 Speaking

Pearson correlation analyses were conducted to examine the relationship between receptive vocabulary size (Yes/No test) and L2 speaking proficiency (vocabulary rating, average frequency, 3K+ proportion). The magnitude of correlation was interpreted according to Plonsky and Oswald’s (2014) field-specific criteria of effect sizes (small = .25, medium = .40, large = .60). As summarized in Table 2, receptive vocabulary size (Yes/No test) scores significantly correlated with vocabulary rating (r = .552, a large effect). Although neither of the lexical sophistication measures were significantly correlated with learners’ receptive vocabulary size, the correlation between size and 3K+ proportion approached statistical significance with a small effect size (r = .274, p = .066).
L2 VOCABULARY AND SPEECH

Table 2. *Pearson Correlations between Receptive Vocabulary Size and Spoken Lexical Measures*

<table>
<thead>
<tr>
<th></th>
<th>Yes/No test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
</tr>
<tr>
<td>Vocabulary rating</td>
<td>.552**</td>
</tr>
<tr>
<td>Average frequency</td>
<td>.173</td>
</tr>
<tr>
<td>K3+ proportion</td>
<td>.274</td>
</tr>
</tbody>
</table>

*Note.* ** indicates $p < .01$. K3+ = the third thousand frequency or above words.

Discussion

The current study was designed to investigate the relationship between receptive vocabulary size and L2 speaking proficiency with 46 advanced learners who had been immersed in an input-rich naturalistic setting for at least seven months. Receptive vocabulary size measured via a Yes/No test significantly related to the vocabulary rating based on the participants’ spontaneous L2 speech. This result suggests that learners with larger receptive vocabulary sizes are more likely to be lexically proficient in speaking (De Jong et al., 2012; Koizumi & In’nami, 2013). However, the absence of meaningful associations between the Yes/No test scores and lexical sophistication measures suggests a complex landscape with regard to the relationship between vocabulary size and use, merit ing further exploration and discussion.

With regard to the association between receptive vocabulary size and human rating, according to Plonsky and Oswald’s (2014) benchmarks, there was a large correlation between size and spoken lexical use ($r = .552$). The magnitude of the size-and-speaking link, however, appears much smaller than the size of correlations found in receptive language skills (e.g., reading, $r = .79$ on average; Jeon & Yamashita, 2014). This finding is consistent with the view that vocabulary size beyond the 2,000-frequency level might not be as important for L2 speaking improvement as for reading or writing. Corpus-based research evidence suggests that the most frequent 2,000 words account for around 90% of the spoken corpora from general conversation (Milton, 2009). The significant role high frequency words play in spoken texts was also confirmed in our data (K1 and K2 words together accounted for around 95% of our participants words). This was especially the case for participants who had already mastered the most frequent 2,000 words (*Range* = 3,100 to 8,770 words). Between-speakers variability in the size beyond 3,000 words might not provide a strong indication of their L2 speaking proficiency. Compared with Koizumi and In’nami (2013) finding that vocabulary size alone explained up to 60% of the variance in speaking proficiency, the current study failed to find a comparable predictive power of vocabulary size for L2 speaking (i.e., 29%). The marked difference between these two studies was that Koizumi and In’nami targeted
L2 VOCABULARY AND SPEECH

novice EFL learners, whereas the current study tested advanced ESL learners. These findings support the notion of the diminishing power of vocabulary knowledge to predict L2 speaking when vocabulary size grows beyond a certain point around the 2,000-frequency level (Milton, 2009). However, the correlation found in the current study ($r = .552$) is larger than that found in Miralpeix and Muñoz (2018) who used a similar oral measurement (human rating based on lexicogrammatical features, $r = .413$). Interestingly, such a difference emerged despite the fact that our participants are likely more proficient ($M_{size} = 6,058$ words) than their participants ($M_{size} = 5,127$ words). A potential reason might relate to the difference in rating rubrics used in the two studies. In Miralpeix and Muñoz’s study, their raters judged oral proficiency quality based on a rubric featuring lexis and grammar combined, whereas our rubric focused exclusively on lexis. A further reason might relate to the task formats we used to elicit L2 speech (picture narratives vs. interviews asking open questions). Since their participants were asked to answer questions on familiar topics that had already been discussed in the classroom, they could have avoided errors by relying on readily available knowledge of topic-related words and phrases without the need to use new words (Skehan, 2009).

Regarding the relationship between receptive vocabulary size and objective lexical measures, we did not find significant correlations between Yes/No test scores and lexical sophistication measures, and a small effect ($r = .274$, $p = .066$), if any, was found in relation to the proportion of low frequency words (3K+). These findings indicate that learners with larger vocabulary sizes do not necessarily produce rare and advanced L2 words in oral narratives. These little-to-no correlations might have resulted from the diminishing role of low frequency words beyond the most frequent 2,000 words in spoken discourses (Milton, 2009). Conversational communication does not normally require speakers to use sophisticated words, as one of the most important goals of communication is to be comprehensible to listeners (Saito et al., 2016). Knowledge of high frequency words may be adequate for speakers to express themselves. Even when required to produce infrequent words, they could paraphrase difficult words using frequent and easy words (but in formal or academic spoken contexts, a larger vocabulary size may be necessary to be functional; Dang & Webb, 2014). Lexical profiling analysis in the current study supports this view, as indicated by the 95% coverage of 1K and 2K words in the elicited speech. Given the participants’ relatively high mean vocabulary rating scores ($M = 5.9$; “a wide enough vocabulary to discuss topics at length” and “generally paraphrases successfully” according to IELTS Band 6), they were judged to be lexically proficient with the predominant use of high frequency words (Kyle & Crossley, 2015).

Notably, no significant correlations were found between vocabulary rating and sophistication measures, indicating production of low frequency words did not positively affect how raters perceived L2 speech. As suggested by lexical profiling analysis, a limited
L2 VOCABULARY AND SPEECH

range of low frequency words elicited through the narrative task might be responsible for the absence of significant correlations. In addition, as highlighted above, oral output of low frequency words might not contribute to general speaking proficiency particularly for advanced L2 speakers. Saito et al. (2016) found that appropriate use of L2 words during oral narrative distinguished between groups of speakers with higher proficiency levels (intermediate vs. advanced), whereas sophistication was not sensitive enough to capture differences at higher proficiencies (beginner vs. advanced). It is notable that in an IELTS lexical descriptor, terms related to appropriateness and accuracy appear more frequently in the upper bands (“precision” in Band 9, “precise meaning” and “inaccuracies” in Band 8, “inappropriate choices” in Band 7, “inappropriacies” in Band 6). Perhaps the relative contribution of appropriateness and sophistication according to different proficiency levels is agreed in speaking assessment.

On a final note, the findings of the current study overall confirm the relationship between vocabulary knowledge and L2 speaking proficiency, indicating the possibility of using vocabulary measures when gauging information related to learner speaking ability. However, we also need to understand that the relationship does not seem as simple as first expected. A closer examination of our data suggests that the relationship between receptive vocabulary size and L2 oral proficiency reveals a potential difficulty in extrapolating learners’ speaking ability from vocabulary test scores. Regarding the relationship between Yes/No test scores and spoken vocabulary rating, Figure 1 demonstrates that three participants with smaller than a vocabulary size of 4,000 words (3,100, 3,483, 3,770, compared with \( M = 6,058.1 \)) were judged as more lexically proficient than fourteen participants with vocabulary size scores higher than 4,000 words (Range = 4,628 to 6,795). Such a complex relationship between vocabulary knowledge and spoken lexical use could be explained by individual differences (Nation & Webb, 2011) in addition to insensitivity of growing vocabulary size as a predictor of oral ability (Milton, 2009). Learners with large vocabulary sizes do not necessarily produce sophisticated L2 words simply because they may not be willing to say the words or decide to take avoiding strategies and express themselves using alternative easy words.
Conclusion, Implications, and Future Research

Earlier research related to receptive vocabulary size tests suggest test scores provide an indication of general L2 proficiency when evaluating reading, listening, writing, and speaking (Harrington & Carey, 2009; Harrington & Roche, 2014; Roche & Harrington, 2013; Roche et al., 2016). In order to move forward our understanding of the role that vocabulary plays in L2 proficiency, the current study explored the relationship between vocabulary size and speaking with advanced L2 learners in an input-rich L2 learning context. A relatively large correlation was found between receptive vocabulary size and oral lexical proficiency based on the human rating scores. However, the L2 learners with large vocabulary sizes did not necessarily produce lexically sophisticated words in speech. Overall, the findings tentatively suggest that vocabulary size predicts L2 oral ability for advanced learners, albeit to a lesser extent than when compared to studies with low proficiency learners (Koizumi & In’nami, 2013).

With evidence in addition to the current data, we suggest that Yes/No tests potentially serve as a practical assessment instrument in L2 teaching and learning contexts. L2 teaching practitioners including course instructors and program organizers can use the tests when assigning class levels to learners in EAP courses or short-term intensive L2 learning programs. The clear advantage of using a Yes/No test is that the test administration is time-efficient (i.e., around 10 minutes) and user-friendly (the computerized version available at http://www.lognostics.co.uk/tools/), contrasting with considerable efforts and costs L2 skill assessment should take (e.g., rating 46 speech samples took approximately three hours in the current study). We also need to consider that Yes/No test scores indicate learners’ L2 proficiency with varying accuracy depending upon language skills in question. For instance,
L2 VOCABULARY AND SPEECH

Yes/No test scores might be more accurately predictive of writing than either listening or speaking skills (Harrington & Carey, 2009; Miralpeix & Muñoz, 2018). The test score is less likely to inform learners’ pronunciation accuracy rather than their fluency or lexical use (Miralpeix & Muñoz, 2018). Considering these caveats, Yes/No tests will be of practical use for language teachers and practitioners when assessing learners’ vocabulary knowledge.

Finally, we should highlight several limitations with the current study when more robust studies are designed to scrutinize the vocabulary-and-speaking relationship. First, the vocabulary size which we compared with speaking was based on a “receptive” vocabulary task (i.e., recognition task) rather than a “productive” vocabulary task. Vocabulary knowledge accessible at a recognition level might not be readily available for L2 production. For example, learners who can distinguish different words during listening may not be able to pronounce them accurately. Thus, it is important to measure productive vocabulary knowledge for the relationship with productive language skills to be made clearer. In fact, empirical studies that find vocabulary knowledge as a strong predictor of L2 speaking proficiency used productive vocabulary tests (i.e., sentence completion in De Jong et al., 2012; L1-L2 translation in Koizumi & In’nami, 2013). Despite challenges to defining productive vocabulary construct (Fitzpatrick & Clenton, 2017), a number of productive vocabulary measures exist based on a variety of task formats: a sentence completion task (Productive Vocabulary Levels Test; Laufer & Nation, 1999), a free production task (Lexical Frequency Profile; Laufer & Nation, 1995), a word association task (Meara & Fitzpatrick, 2000), a gap-filling task (Fitzpatrick & Clenton, 2017), and other new productive measures (e.g., $V_{\text{Size}}$; Gesa, 2015 cited in Meara & Miralpeix, 2017). Future studies need to measure receptive and productive vocabulary knowledge and explore the relative weights of the two types of measures predicting L2 speaking proficiency.

Second, the task format was not matched in mode (written, spoken) between vocabulary and skill measures. To test vocabulary knowledge, we used a Yes/No check task in which L2 words were elicited in written form, whereas L2 production was elicited in spoken form. The majority of earlier studies measured vocabulary in written form regardless of target L2 skills (e.g., De Jong et al., 2012; Koizumi & In’nami, 2013; Miralpeix & Muñoz, 2018; Stehr, 2008, 2009; Wang & Treffers-Daller, 2017). As one of few exceptions, a study conducted by Milton and colleagues (2010) confirmed the significant effects of mode difference in comparing vocabulary size and four language skills. Another recent study (Cheng & Matthews, 2018) suggests that the constructs of written and spoken vocabulary knowledge are related, yet independent of each other. Thus, written and spoken modes of knowledge in combination with a receptive and productive distinction are in need of systematic investigations in future studies. Although limited, some spoken vocabulary measures exist: Aural Lex (Milton & Hopkins, 2006), Listening Vocabulary Levels Test (McLean, Kramer, & Beglar, 2015; Uchihara & Harada, in press), and a partial dictation test.
L2 VOCABULARY AND SPEECH (Chen & Matthews, 2018). Using these tests for future research might further our understanding of the role vocabulary plays in relation to L2 proficiency.

Finally, a restricted range of low frequency words elicited through a picture narrative task (i.e., about 5%) might be inadequate for sophistication measures to fully represent a speaker’s true lexical resources. Although a picture narrative task was considered suitable for eliciting lexically rich language (Skehan, 2009), L2 speech elicited in the current study was based on a single task (i.e., story-narrative) and limited speech samples (1-4 minutes). Since the quality of L2 production can be largely subject to task types (In’nami & Koizumi, 2016), it is crucial to use a large repertoire of production tasks with different communicative purposes in order to capture a detailed picture of L2 oral proficiency (De Jong et al., 2012). It is also worth exploring whether the role of vocabulary size is moderated by different tasks with varying degrees of complexity in different modes of communication (i.e., monologue vs. dialogue; Tavakoli, 2018).
References


