Exploring the relationship between productive vocabulary knowledge and second language oral ability

Takumi Uchihara[a*] and Kazuya Saito[b]

[a]*Faculty of Education, Western University, Canada;* [b]*Department of Applied Linguistics and Communication, Birkbeck, University of London, UK*

Abstract

The current study investigated the extent to which L2 learners' productive vocabulary knowledge could predict multiple dimensions of spontaneous speech production. A total of 39 EFL participants with varying L2 proficiency levels first completed a productive vocabulary knowledge task (Lex30). Their spontaneous speech, elicited via a series of picture description task, was then assessed for comprehensibility (i.e., ease of understanding), accentedness (i.e., linguistic nativelikeness), and fluency (i.e., speech rate). The findings showed that the productive vocabulary scores significantly correlated with L2 fluency, but not with comprehensibility or accentedness. Such results might indicate that more proficient L2 learners, as indicated by their productive vocabulary scores, might be able to speak spontaneously without too many pauses and repetitions, and at a faster tempo. Finally, future research directions will be discussed with a focus on the relationships between vocabulary knowledge and speaking.

Key words: productive vocabulary knowledge; L2 vocabulary; L2 speaking; comprehensibility; fluency

Introduction

Measuring Productive Vocabulary Knowledge

Vocabulary knowledge has gained recent prominence in second language (L2) education, as research seeks to examine the relationship between vocabulary size measures and general language proficiency. This has much practical appeal since scores from vocabulary size measures might act as a proxy for general language proficiency (Meara, 1996). Vocabulary size measures tend to elicit vocabulary knowledge via receptive, or passive, vocabulary knowledge tests (e.g., Vocabulary Levels Test; Nation, 1983, 1990; the Eurocentres Vocabulary Size Test; Meara & Jones, 1990). Accordingly, researchers (e.g. Milton, Wade, & Hopkins, 2010; Milton, 2010, 2013) have generally adopted a wide range of receptive vocabulary tests (e.g., standard yes/no vocabulary size tests), in order to compare test scores to overall proficiency benchmarks (e.g., CEFR levels; Milton 2010), placement tests (e.g., Harrington & Carey, 2009), and standardized proficiency examinations (e.g., IELTS; Milton et al., 2010). Additionally, other research has examined the predictive power of L2 vocabulary knowledge for the four main language skills (i.e., reading, writing, listening, and, to a lesser extent, speaking) (e.g. Baba, 2009; Farvardin & Koosha, 2011; Laufer & Levitzky-Aviad, 2015; Meara & Jones, 1990; Meara & Milton, 2003; Milton, 2009, 2010, 2013; Milton et al., 2010; Qian, 1999, 2002, Schoonen et al., 2003; and, Stæhr, 2008, 2009). This brief summary highlights that a majority of the existing studies likely elicit vocabulary knowledge receptively, and compare such knowledge with reading, writing, and listening performance.

Relatively few studies have compared productive vocabulary knowledge with oral ability (e.g. De Jong, Steinel, Florijn, Schoonen, & Hulstijn, 2012; Koizumi & In'nami, 2013). De Jong et al. (2012) compared various vocabulary measures (e.g. productive vocabulary as well as grammatical knowledge, pronunciation quality, lexical retrieval speed, etc.) and suggested that the elicited productive vocabulary (by Laufer and Nation's Productive Levels Test) is a likely predictor of L2 oral ability. Koizumi and In'nami (2013) also found that productive vocabulary knowledge (as measured by L1-L2 translation tests) "substantially explain(s) L2 speaking proficiency" (p. 910).

A variety of vocabulary researchers have highlighted the relative complexity of the productive vocabulary construct (e.g. Fitzpatrick, 2007, pp. 127–131; Meara & Fitzpatrick, 2000; Nation 2013; Schmitt, 2010). Much of this discussion concerns the variety of different tasks, each pertaining to the elicitation of productive vocabulary. Fitzpatrick and Clenton (2010, p. 539) suggest that particular productive vocabulary tests might be problematic because of what they elicit. For instance, they elicit receptive vocabulary knowledge in sentence completion tasks or, in other tests elicit superfluous language, and do not encourage subjects to demonstrate their full range of "productive" vocabulary knowledge.

In the current paper, we used Lex30 because the productive vocabulary items produced in response to the task are reported as being "broadly speaking' 'representative of … [each subject's] productive mental lexicon" (Fitzpatrick & Clenton, 2010, p. 548) (see Meara & Fitzpatrick, 2000, pp. 20-22 for a discussion of how Lex30 relates to other productive vocabulary measures). Lex30 elicits productive vocabulary knowledge with word associations, and the test was originally elaborated and validated in Meara and Fitzpatrick (2000). Theoretically, the maximum a subject can score on Lex30 is 120 (calculated by tallying 4 responses to each of the 30 cues). In reality, most L2 subjects tend to raw score between 10 and 40 (Fitzpatrick & Clenton, 2010; Meara & Fitzpatrick, 2000; Fitzpatrick & Meara, 2004; Fitzpatrick, 2007; and, Walters, 2012). An alternative percentage score is calculated by looking at the number of infrequent items produced in relation to individual corpora, multiplied by 100. Percentage scoring has been widely used (Fitzpatrick & Clenton, 2010; Fitzpatrick & Meara, 2004; Fitzpatrick, 2007) because it minimizes the influence of varying corpus size derived from the Lex30 task. Thus, the current study adopts the percentage scoring system following previous research (e.g., Fitzpatrick & Meara, 2004).

Measuring L2 Oral Ability

In the field of L2 acquisition, many researchers have persisted with the tendency to draw on controlled speech measures (e.g., reading aloud sentences and paragraphs) in order to examine how L2 learners produce target phonological, lexical and grammatical structures (Piske, MacKay, & Flege, 2001). However, such measures can easily allow adult L2 learners to carefully monitor their correct production (Doughty, 2003). Thus, many researchers have emphasized the importance of adopting free speech measures (e.g., picture narratives), whereby L2 learners are guided to use language for meaning by paying equal attention to phonological, temporal, lexical and grammatical use of language (Spada & Tomita, 2010) under time pressure (Ellis, 2005).

Another crucial aspect of L2 speech assessment is concerned with the multifaceted characteristics of L2 speech (Trofimovich & Isaacs, 2012; Saito, Trofimovich, & Isaacs, 2015, 2016). Although previous studies have tended to use L2 learners' general proficiency scores (e.g., IELTS) as a single index for their oral proficiency, especially in the field of L2 testing (e.g., Milton et al., 2010), there is a tradition in the L2 phonetics literature that L2 speech be analysed

from multiple angles. For example, some researchers use novice native speakers' scalar judgements of L2 speech based on not only accentedness (i.e., linguistic nativelikeness), but also comprehensibility (i.e., how easy to understand) (Derwing & Munro, 2009). Using these two different dimensions is crucial, as they are inter-related but essentially different constructs (e.g., heavily accented speech can be highly comprehensible) (Jenkins, 2000; Levis, 2005). In addition, other researchers ask linguistically-trained judges to evaluate specific aspects of language, such as segmental (Piske et al., 2001), prosodic (Field, 2005), and temporal (Bosker, Pinget, Quené, Sanders, & De Jong, 2013) qualities of L2 speech (see also Saito, Trofimovich, & Isaacs, 2015, 2016).

A number of researchers have extensively examined what kinds of learner variables—the amount and quality of L2 experience, educational backgrounds, motivation, and aptitude—interact to predict various dimensions of L2 speech. For example, adult L2 learners likely attain improved comprehensibility (but not necessarily reduced foreign accentedness) in accordance with increased experience (typically measured via length of residence profile), as long as they have high willingness to communicate with native and other non-native speakers (Derwing & Munro, 2013; Saito, 2015b). Whereas such experience effects can be clearly observed in the domains of suprasegmentals (Trofimovich & Baker, 2006) and fluency (Mora & Valls-Ferrer, 2012), it may require a great amount of experience (e.g., 5-10 years of immersion) to refine the sophisticated use of language (i.e., segmentals, lexical richness, grammatical complexity) (Saito, 2015a; Baker, 2010).

Recently, certain researchers have begun to provide some supporting evidence that L2 learners' vocabulary knowledge can be a relatively strong predictor of their L2 speech learning. According to Best and Tyler's (2007) Perceptual Assimilation Model-L2, L2 learners are induced to enhance their robust linguistic representations in order to use their growing vocabulary repertoires accurately and fluently. While L2 learners' segmental perception appears to be tied to their receptive vocabulary knowledge (Bundgaard-Nielsen, Best, & Tyler, 2011), less is known about the relationship between productive vocabulary knowledge and L2 production skills; this is an intriguing research direction for speech research.

Current Study

As reviewed above, previous studies have provided some evidence in support for the relationship between L2 productive vocabulary knowledge and oral ability. Yet, whereas the validity of their productive vocabulary measures (e.g., Koizumi & In'nami, 2013 for L1-L2 translations) has remained unclear (Fitzpatrick & Clenton, 2010), the linguistic qualities of L2 speech have been analysed only based on participants' general proficiency test scores (e.g., Milton, 2010 for IELTS). The current study was designed to expand the existing literature on this topic by making two methodological changes. First, we adopted Lex30 to measure L2 productive vocabulary knowledge, as the test was validated and suggested to tap into the productive mental lexicon (Fitzpatrick & Clenton, 2010). Second, following the standards of recent L2 speech research, we elicited spontaneous speech samples via a timed picture description task (Saito, 2015a, 2015b; Saito et al., 2016), and assessed two global dimensions of L2 oral proficiency—foreign accentedness (linguistic nativelikness) and comprehensibility (ease of understanding)—and one specific aspect of L2 speech—optimal speech rate (Derwing & Munro, 2009).

Participants

The participants were 39 first year Japanese students (21 female, 18 male) at a Japanese university. Although all of the students had learned English for six years from Grade 7, their

general English proficiency widely varied. According to TOEIC (consisting of listening and reading comprehension questions), their average score was 659 out of 990 (*SD* = 105.9, range = 495-865).

Productive Vocabulary Test (Lex30)

The current study followed the same procedure as earlier Lex30 studies (see Meara & Fitzpatrick, 2000). Each set of responses, representing individual participant corpora were then processed in the same way as Meara and Fitzpatrick (2000). Misspellings were corrected, and responses were lemmatized according to Bauer and Nation's (1993) criteria. Individual corpora were then typed into a text file and pasted into the JACET8000 (JACET, 2003) website (available at http://www.tcp-ip.or.jp/~shim/j8web/j8web.cgi). The automatic scorer awards a point for each infrequent item (i.e. items beyond the first 1000 most frequently occurring words). In terms of the scoring method, we used percentage as well as raw scores.

L2 Oral Ability Measures

The current study adopted a timed picture description task which was elaborated and designed to elicit a certain length of spontaneous speech from not only advanced and intermediate, but also beginner L2 learners (Saito, 2015a, 2015b; Saito et al., 2016). The elicited speech samples were rated by native speakers according to two global rubrics (comprehensibility, accentedness), and then submitted to fluency analyses.

Speech materials. The participants were asked to explain a series of seven pictures. While each picture had three key words as hints to prevent L2 talkers from producing long and frequent pauses, they were given only 5 seconds as preparation prior to each picture description. The first four pictures were used for practice; the remaining three pictures were used for the final analyses. The three pictures highlighted (a) "a table left out in the driveway under heavy rain", with the keywords rain, table, and driveway; (b) "three guys playing rock music with one guy singing a song and the other two guys playing guitars", with the keywords three guys, guitar, and rock music; and (c) "a long stretch of road under a blue sky with a lot of clouds", with the keywords blue sky, road, and cloud.

All of the recordings were conducted individually in a sound proof booth at the university, and digitally stored as WAV files. First, for each talker, we sampled 10 seconds of each picture description and then combined them into a single speech sample. The length of each sample (30 sec) was determined in line with the previous L2 speech research (e.g., Derwing & Munro, 2009; Trofimovich & Isaacs, 2012). The mean number of words for each file was 44.5, ranging from 35-61 words.

Global analyses. Five native speaking raters were recruited (2 males, 3 females) from an English speaking university in Montreal, Canada. All of them were undergraduate students with a mean age of 21.6 years. Along with the definition by Isaacs and Thomson (2013), these raters could be considered as *novice* in that they had not taken any linguistic courses prior to the project. Based on the language background questionnaire, they reported low familiarity with Japanese-accented English (*M* = 1.8 from 1 = *Not at all* to 6 = *Very much*). None of the raters reported any hearing problems.

To reflect their intuition of how easy it was to understand these speech samples and how nativelike the samples sounded, the raters first received a brief (but not detailed) training on the two global rubrics of 'comprehensibility' and 'accentedness' (Derwing & Munro, 2009). Then, the raters listened to all speech samples in a randomized order via the MATLAB computer software. After hearing each sample only once, they used a moving slider on a computer screen to rate for comprehensibility and accentedness on a 1000-point scale (*0 = hard to understand,*

heavily accented; 1000 = easy to understand, little accent). For training materials and onscreen labels, see Supporting Information (Appendix) adapted from Saito et al. (2015, 2016).

Fluency analyses. Unlike the global analyses conducted by novice raters, five experienced native speaking raters (3 males, 2 females) were recruited with their mean age of 28.3 years. In conjunction with Isaacs and Thomson's (2013) definition, these raters could be considered as experts in that they had much experience in pronunciation and vocabulary analyses (all of them were graduate students in the department of applied linguistics at the time of the project) and in ESL/EFL teaching (M = 4.0 years from 2 to 6 years). Our rationale for using experienced raters is that perceived fluency is a complex phenomenon which may require raters' relevant knowledge in applied linguistics. For example, while judging "optimal speech rate," raters need to understand and process various kinds of temporal information, such as the number of filled/unfilled pauses and repetition and timing of these pauses (mean length of run) (Derwing, Rossiter, Munro, & Thomson, 2004). None of the raters reported any hearing problems.

In the rating sessions, the raters listened to the speech samples randomly played through the MATLAB interface, and used the moving slider to judge the quality of the tokens based on a 1000-point scale (*0 = non-targetlike, 1000 = targetlike*). Due to the demanding nature of the rating task, they were allowed to replay each speech sample until they felt satisfied with their judgements. To increase the reliability and consistency of their assessment, all of the raters went through training prior to the rating sessions. First, they received a thorough explanation on the construct of fluency (i.e., speech rate) from a trained research assistant. Next, they practiced the procedure by using five speech samples (not included in the main dataset), for each of which they explained their decisions and received feedback from the assistant.

For training materials and onscreen labels for the fluency analysis, see Supporting Information (Appendix) adapted from Saito et al. (2015, 2016).

Results

Productive Vocabulary (Lex30) Scores and Oral Ability Task Scores

Table 1 presents the means, standard deviations, and ranges of Lex30 scores and L2 oral ability task scores. For L2 oral ability, a Chronbach's alpha reliability analysis was performed; the five novice raters demonstrated relatively high inter-rater agreement for comprehensibility (.91) and accentedness (.93). Their scores were thus averaged to generate a single comprehensibility/accentedness score for each talker. As summarized in Table 1, the 39 talkers produce substantially higher scores in terms of comprehensibility (M = 466) than accentedness (M = 283), a finding in line with previous L2 speech research (e.g., Derwing & Munro, 2009). Given the relatively high inter-rater agreement for speech rate among the five expert raters (Chronbach alpha = .88), a single mean score was calculated for each talker. Similar to their global scores, their fluency performance widely varied (see Table 1).

Table 1. *Lex30 and L2 oral ability measure scores*

	M	*SD*	Range	Reliability
Lex30				
Raw scores	42.54	10.33	22-60	
Percentage scores	42.57	7.55	26.97-56.19	
Global constructs				
Comprehensibility	467	141	189-781	.91

Accentedness	283	106	82-556	.93
Fluency				
Speech rate	567	178	213-834	.88

Correlation Analysis between Productive Vocabulary and L2 Oral Ability

To examine the vocabulary-proficiency link, a set of simple correlation analyses were conducted on the global rating scores (comprehensibility/accentedness) and fluency score (speech rate) as dependent variables, and the Lex30 scores (raw/percentage) as an independent variable. As summarized in Table 2, their raw scores were moderately correlated with speech rate (p = .033), but did not significantly predict comprehensibility (p = .095) or accentedness (p = .539). In contrast, none of the oral ability scores (comprehensibility, accentedness, and speech rate) were significantly correlated with percentage scores (p > .05).

Table 2. *Pearson Correlations between Productive Vocabulary Scores and L2 Oral Ability Measures*

	Raw scores		Percentage scores	
	r value	*p* value	*r* value	*p* value
Global constructs				
Comprehensibility	*r* =.271	*p* = .095	*r* =.222	*p* = .174
Accentedness	*r* =.034	*p* = .835	*r* =.040	*p* = .808
Fluency				
Speech rate	*r* = .342	*p* = .033*	*r* = .175	*p* = .286

* indicates $p < .05$

Temporal Correlates of Comprehensibility and Accentedness

Finally, a set of simple correlation analyses were conducted to investigate the temporal (speech rate) correlates of L2 comprehensibility and accentedness. As summarized in Table 3, whereas comprehensibility was strongly linked with fluency, no correlation between accentedness and fluency was found.

Table 3. *Pearson Correlations between Global Scores (Comprehensibility/Accentedness) and Fluency Score (Speech Rate)*

	Comprehensibility	Accentedness
Speech rate	*r* =.518**	*r* = .117

** indicates $p < .01$

Discussion

By interfacing L2 vocabulary and speech research perspectives, the current paper explored the extent to which L2 productive vocabulary knowledge could predict the global (comprehensibility, accentedness) and specific (optimal speech rate) components of L2 speech production. The results showed that our participants' Lex30 raw scores significantly, but moderately, correlated with fluency measure (speech rate), whereas neither comprehensibility nor accentedness showed significant correlations with learners' Lex30 scores. Conversely, the Lex30 percentage scores did not correlate with the speaking measures. The results here were

partially in line with the previous studies investigating the role of productive vocabulary in L2 speaking performance (De Jong et al., 2012; Koizumi & In'nami, 2013). More importantly, our study further suggested that productive vocabulary knowledge might predict, in particular, how fluently L2 learners can speak, but not how comprehensible nor how native-like their speech sounds.

The question has then become: Why does vocabulary knowledge play an integral role in L2 oral fluency? It has been previously suggested that L2 learners with larger vocabularies might be equipped with well-organized lexical networks (Meara, 1990, 2006), and therefore might be able to retrieve L2 words more economically and immediately in real-time communication (Levelt, 1989; de Bot, 1992; Kormos, 2006; Skehan, 2009). According to Meara (1990, 2006), fully-developed lexical networks can be defined as well-established word associations where lexical links between and within the network might be stronger than poor lexical networks. Since the L2 speech production system is lexically driven (de Bot, 1992; Kormos, 2006), such robust L2 lexicons might be of key importance for successful L2 speaking by making L2 words readily accessible for immediate lexical retrieval (Skehan, 2009). In this regard, the current study tentatively suggested that learners who speedily produced associated words within strict time constraints (30 seconds) during the Lex30 task might be equipped with well-organized, immediately accessible L2 lexical networks. Consequently, such developed L2 lexicons might allow them to retrieve L2 words without much difficulty during the speaking task and produce fluently articulated speech in a spontaneous manner.

Although we failed to find any significant correlations between productive vocabulary task scores and global rating scores (comprehensibility/accentedness), it is noteworthy that the relationship between comprehensibility and productive vocabulary knowledge showed more of a trend toward significance compared to accentedness. One reason for this weak (but not significant) relationship could be that comprehensibility is equally linked with not only fluency (as shown in the current study, see Table 3), but also the pronunciation, vocabulary and grammar aspects of language (Saito et al., 2015,2016; Trofimovich & Isaacs, 2012). Whereas the Lex30 scores could predict L2 learners' fluent delivery of speech without much pauses and repetition, their scores did not appear to reflect how accurately they could pronounce words (pronunciation accuracy), nor the extent to which they use words in a contextually appropriate manner (lexicogrammar appropriateness).

Pedagogical Implications

The findings provide several crucial implications for L2 speech assessment in classroom contexts. Although much research attention has been directed towards optimizing the process of L2 speech assessment via automated computer software, the construct validity of such automatic scoring systems has remained highly controversial (see Xi, 2007). While many L2 speech researchers (e.g., Derwing & Munro, 2009) have emphasized the importance of setting realistic learning goals for L2 speech learning, such as comprehensibility, it is noteworthy that native speaker judgement of learner speech requires a great amount of time (e.g., 1.5 hours for assessing 39 students in the study), and it may be difficult to apply such assessment tools to classroom contexts. In this regard, our study demonstrates the pedagogic value of the productive vocabulary test (i.e., Lex30) as a tool to predict certain aspects of L2 oral ability. When teachers implement Lex30 even in a large-scale classroom, the results will provide a broad estimate of students' fluency (speech rate) skill, which plays a crucial role in the overall comprehensibility of L2 speech (Saito et al., 2015, 2016; Trofimovich & Isaacs, 2012).

With respect to L2 oral teaching, the findings here suggest that the development of productive vocabulary knowledge may play a key role in improving L2 oral proficiency. Though scarce, there is some evidence that L2 learners' word association knowledge can progress as a function of their increased amount of input and interaction with other native and non-native speakers (e.g., Schmitt, 1998). As suggested by the usage-based account of SLA (e.g., Ellis, 2006), when L2 learners frequently encounter certain vocabulary items in meaningful input, the connections between related words become strong and comprise formulaic units, allowing for more automatic, refined and robust word associations. To include more communicatively authentic activities in L2 teaching syllabi, L2 researchers have suggested several options by elaborating teaching methods per se (e.g., task-based language teaching: Bygate, 2015) and making the most of technological advancements (e.g., language exchanges via video-based interaction: Akiyama & Saito, in press; Saito & Akiyama, in press). Future studies are warranted to examine whether such meaning-oriented activities can actually enhance L2 learners' productive vocabulary knowledge over time, which will in turn positively influence their L2 speech learning in the long run.

Future Direction

The area of L2 speaking has to date received growing attention in research (Pickering, 2012) and pedagogy (Paran, 2012). Although the role of vocabulary in L2 speaking was lightly researched, some researchers have begun to examine this area particularly in the field of language assessment (e.g., Li & Lorenzo-Dus, 2014). Yet, these studies tend to focus on the relationship between speech production and the use of vocabulary derived from the same speech (Koizumi, 2013). Given that vocabulary knowledge and the ability to use it can be different in that all the knowledge is not accessible for use (Chapelle, 1994), it is vital to measure and compare the two constructs (vocabulary and speaking) separately in order to give a deeper insight into the whole picture of the relationship between vocabulary knowledge and oral ability. A limited number of studies follow this fashion (De Jong et al., 2012; Koizumi & In'nami, 2013), but the relationship between vocabulary and speaking in terms of comprehensibility and accentedness remains to be seen. In this sense and considering the considerable popularity of receptive vocabulary tests, the current study took a first step to further probe the role of vocabulary in L2 oral ability by using an ecologically-valid productive vocabulary test (Lex30; Meara & Fitzpatrick, 2000) and L2 speech measures (comprehensibility/accentedness; Derwing & Munro, 2009) echoing the growing concept of English as an International Language (Jenkins, 2000). Our intent is not to be conclusive but indicative for further research which will be informed by and expanded on the current study. Thus, we address several future directions worthy of further investigation on the grounds of the results in two respects—(a) productive vocabulary measure and (b) L2 oral ability measures.

Productive Vocabulary Measure

Productive vocabulary knowledge in the current study was measured by Lex30 using two different types of scoring procedures (raw and percentage scoring). We followed the same procedure used in the previous Lex30 studies (Meara & Fitzpatrick, 2000; Fitzpatrick & Clenton, 2010), but in order to relate vocabulary task scores to oral ability, at least two issues need to be highlighted.

First, frequency-based measurement only might not provide sufficient information about learner's vocabulary knowledge for the potential relationship with oral ability to be made explicit. Such issue is illuminated by the fact that none of the L2 oral ability task scores

(comprehensibility, accentedness, and speech rate) significantly correlated with percentage scores of Lex30 in this study. The Lex30 raw scores and percentage scores are different in that the former taps into other aspects of language competence (e.g., fluency) as well as relative size of productive vocabulary; by contrast, the latter represents scoring which is not affected by varying corpora size across individuals (Clenton, 2010). More specifically, the percentage scoring procedure relies exclusively on a frequency-based principle (i.e., learning words occurs in order of frequency): it is only the proportion of low-frequency items to the total responses that feeds into the final output of Lex30 scores regardless of the total number of responses produced within the time constraint. Clearly, the absence of significant correlation between percentage scores and oral ability in the study indicates potential problems with the frequency-based principle for assessing vocabulary knowledge (Nation & Webb, 2011). Therefore, in order to further investigate the relationship between vocabulary knowledge and L2 speaking ability, it is important to be aware of other aspects of vocabulary knowledge besides frequency such as L1 cognate (Bardel, Gudmundson, & Lindqvist, 2012), speed of lexical retrieval (Fitzpatrick & Izura, 2011; Miralpeix & Meara, 2014), abstractness (Crossley, Salsbury, & McNamara, 2009), sense relations (Crossley, Salsbury, & McNamara, 2010), richness and sophistication (Lu, 2012), and multi-word units (Kyle & Crossley, 2014).

Second, given some neuropsychological evidence that the process of written and oral word production appears fairly distinctive (Shelton & Weinrich, 1997; Shelton & Caramazza, 1999), it is essential to assess vocabulary knowledge in the same modality with targeted language ability (Uchihara & Clenton, 2016). For instance, when relating vocabulary task scores to speaking, we should note that the task needs to be delivered in a spoken form, not a written form (Milton et al., 2010; Milton, 2013). Because of the lack of a spoken version of productive vocabulary test at the time of experiment, learner's vocabulary knowledge was elicited productively in a written form in this study, which might be partially attributed to weak or null correlations between productive vocabulary knowledge and L2 oral ability. To respond to such issues, a spoken version of Lex30 might meet our expectation in order to best relate such task scores to L2 speaking ability. As Milton and his colleagues (2014) suggest, the spoken Lex30 is ready to be developed given great amounts of existing data from a written version of Lex30 (Fitzpatrick & Clenton, 2010; Meara & Fitzpatrick, 2000; Fitzpatrick & Meara, 2004; Fitzpatrick 2007; Jimenez Catalán & Moreno Espinosa, 2005; and Walters, 2012) against which it can be compared. Accordingly, it is suggested that vocabulary knowledge needs to be assessed in an oral form for the potential relationship with L2 oral ability to be made fully clear.

L2 Oral Ability Measures

One obvious limitation in terms of the oral ability measure used in the study concerns the length of the speech samples. Although we aimed to satisfy the standard in L2 pronunciation research (30 sec), considerably longer speech samples have been used in L2 vocabulary and grammar research to better reflect the current state of L2 learners' vocabulary and grammar competence (e.g., 3 min for Crossley et al., 2009, 2010). Employing longer speech samples (> 3 min), therefore, future studies of this kind need to examine whether and to what degree the productive vocabulary test scores could predict the way L2 learners actually use vocabulary while speaking at a spontaneous mode (Uchihara, Saito, & Clenton, 2016). Drawing on the computational modelling of L2 vocabulary proficiency developed by Crossley, Salsbury, and McNamara (2014) and Lu (2012), Saito, Webb, Trofimovich and Isaacs (2015) recently examined lexical correlates of native speakers' comprehensibility judgement based on 40 Francophone learners of English in Canada. Their full-length picture descriptions (> 3 min) were

analysed via a set of comprehensive measures, tapping into various domains of L2 lexical usage. The results identified five specific dimensions—appropriateness, fluency, variation, sophistication, and sense relations—directly relevant with native speakers' successful understanding of L2 speech. Following this line of thought, it would be intriguing if future studies take a look at lexical profiles of actual speech samples from multiple angles in order to highlight the complex relationship between the Lex30 scores, comprehensibility, and a range of lexical dimensions of L2 speech.

Next, it is crucial to remember that any discussion in the study regarding L2 oral ability was based on one task modality—timed picture description. Though few in number, certain SLA studies have shown that adult L2 learners' speech performance substantially differs according to various task conditions. Derwing et al. (2004) demonstrated that L2 learners' comprehensibility and fluency scores were rated more positively in monologue- and dialogue-based tasks than in a picture-narrative task. Other studies pointed up the role of planning time. For example, whereas a sufficient amount of time for task completion (online planning) enables L2 learners to use more appropriate lexical items, ample preparation time (offline planning) leads to more fluent L2 speech at the sacrifice of accuracy and complexity (Ellis, 2009 for review). Similarly, Crowther, Trofimovich, Isaacs and Saito (2015) provided some evidence that same L2 learners likely make more lexicogrammar errors in a cognitively more demanding task format (i.e., TOEFL iBT integrated task) compared to a less complex task format (IELTS long-turn speaking task). Therefore, one future direction for the L2 speech/vocabulary interface research could be to test the generalizability of the preliminary finding in the small-scale study with a specific task context (timed picture descriptions by 39 Japanese EFL learners), for a large sample of L2 learners with different task modalities with various task demands (cf. Hulstijn, Schoonen, De Jong, Steinel, & Florijn, 2012).

References

Akiyama, Y., & Saito, K. (in press). Development of comprehensibility and its linguistic correlates: A longitudinal study of video-mediated telecollaboration. *Modern Language Journal*.

Baba, K. (2009). Aspects of lexical proficiency in writing summaries in a foreign language. *Journal of Second Language Writing*, *18*, 191–208.

Baker, W. (2010). Effects of age and experience on the production of English word-final stops by Korean speakers. *Bilingualism: Language and Cognition*, *13*, 263–278.

Bardel, C., Gudmundson, A., & Lindqvist, C. (2012). Aspects of lexical sophistication in advanced learners' oral production. *Studies in Second Language Acquisition*, *34*, 269–290.

Bauer, L., & Nation, I.S.P. (1993). Word families. *International Journal of Lexicography*, *6*, 253–279.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception. In O. Bohn, & M. Munro (Eds.), *Language experience in second language speech learning: In honour of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins.

Bosker, H. R., Pinget, A. F., Quené, H., Sanders, T., & De Jong, N. H. (2013). What makes speech sound fluent? The contributions of pauses, speed and repairs. *Language Testing*, *30*, 159–175.

Bundgaard-Nielsen, R., Best, C., & Tyler, M. (2011). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics, 32,* 51–67.

Bygate, M. (2015). Sources, developments and directions of task-based language teaching. *The Language Learning Journal*, 1-20.

Catalán, R. M. J., & Espinosa, S. M. (2005). Using Lex30 to measure the L2 productive vocabulary of Spanish primary learners of EFL. *VIAL, Vigo international journal of applied linguistics, 2,* 27–44.

Chapelle, C. A. (1994). Are C-tests valid measures for L2 vocabulary research? *Second Language Research, 10,* 157–187.

Clenton, J. (2010). *Investigating the construct of productive vocabulary knowledge with Lex30.* (Unpublished doctoral dissertation). University of Swansea, Swansea, U.K.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2009) Measuring L2 lexical growth using hypernymic relationships. *Language Learning, 59,* 307–34.

Crossley, S. A., Salsbury, T., & McNamara, D. S. (2010). The development of polysemy and frequency use in English second language speakers. *Language Learning, 60,* 573–605.

Crossley, S. A., Salsbury, T., & Mcnamara, D. S. (2014). Assessing lexical proficiency using analytic ratings: A case for collocation accuracy. *Applied Linguistics, 35*.

Crowther, D., Trofimovich, P., Isaacs, T., & Saito, K. (2015). Does a speaking task affect second language comprehensibility? *Modern Language Journal, 99,* 80-95.

de Bot, K. (1992). A bilingual production model: Levelt's "speaking" model adapted. *Applied Linguistics, 13,* 1–24.

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition, 34,* 5–34.

Derwing, T.M., Rossiter, M.J., Munro, M.J., & Thomson, R.I. (2004). L2 fluency: Judgments on different tasks. *Language Learning, 54,* 655–679.

Derwing, T. M., & Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching, 42,* 476–490.

Derwing, T. M., & Munro, M. J. (2013). The development of L2 oral language skills in two L1 groups: A seven-year study. *Language Learning, 63,* 163–185.

Doughty, C. (2003). Instructed SLA: Constraints, compensation, and enhancement. In M. Long & C. Doughty (Eds.), *Handbook of second language acquisition* (pp. 257–310). Malden, MA: Blackwell.

Ellis, N. C. (2006). Language acquisition as rational contingency learning. *Applied Linguistics. 27,* 1-24.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a Second Language: A psychometric study. *Studies in Second Language Acquisition, 27,* 141–172.

Ellis, R. (2009). The differential effects of three types of task planning on the fluency, complexity, and accuracy in L2 oral production. *Applied Linguistics, 30,* 474–509.

Farvardin, M. T., & Koosha, M. (2011). The role of vocabulary knowledge in Iranian EFL students' reading comprehension performance: Breadth or depth? *Theory and Practice in Language Studies, 1,* 1575–1580.

Field, J. (2005). Intelligibility and the listener: The role of lexical stress. *TESOL Quarterly, 39,* 399–423.

Fitzpatrick, T. (2007). Productive vocabulary tests and the search for concurrent validity. In H. Daller, J. Milton & J. Treffers-Daller (Eds.), *Modelling and assessing vocabulary knowledge* (pp. 116–132). Cambridge: Cambridge University Press.

Fitzpatrick, T., & Clenton, J. (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing*, *27*, 537-554.

Fitzpatrick, T., & Izura, C. (2011). Word association in L1 and L2. *Studies in Second Language Acquisition*, *33*, 373–398.

Fitzpatrick, T., & Meara, P. (2004). Exploring the validity of a test of productive vocabulary. *Vigo international journal of applied linguistics*, *1*, 55–74.

Harrington, M., & Carey, M. (2009). The on-line Yes/No test as a placement tool. *System*, *37*, 614–626.

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *Language Learning Journal*, *36*, 153–166.

Hulstijn, J. H., Schoonen, R., De Jong, N. H., Steinel, M. P, & Florijn, A. (2012). Linguistic competences of learners of Dutch as a second language at the B1 and B2 levels of speaking proficiency of the Common European Framework of Reference for Languages (CEFR). *Language Testing*, *29*, 203–221.

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, *10*, 135–159.

JACET (2003). *JACET list of 8000 basic words.* Tokyo: Japan Association of College English Teachers.

Jenkins, J. (2000). *The phonology of English as an international language*. Oxford: Oxford University Press.

Koizumi, R. (2013). Vocabulary and speaking. In C. A. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 1–7). Oxford, UK: Wiley-Blackwell.

Koizumi, R., & In'nami, Y. (2013). Vocabulary knowledge and speaking proficiency among second language learners from novice to intermediate levels. *Journal of Language Teaching and Research*, *4*, 900–913.

Kormos, J. (2006). *Speech production and second language acquisition*. Mahwah, NJ: Erlbaum.

Kyle, K., & Crossley, S. A. (2014). Automatically assessing lexical sophistication: indices, tools, findings, and application. *TESOL Quarterly*. Advance online publication. doi: 10.1002/tesq.194

Laufer, B., & Levitzky-Aviad, T. (2015, August). *What type of vocabulary knowledge predicts reading comprehension: Word meaning recall or word meaning recognition?* Paper presented at the 25th Annual Conference of the European Second Language Association EUROSLA 25, Aix-en-Provence, France.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*, 33–51.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* MA: MIT Press.

Levis, J. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, *39*, 367–377.

Li, H., & Lorenzo-Dus, N. (2014). Investigating how vocabulary is assessed in a narrative task through raters' verbal protocols. *System*, *46*, 1–13.

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners' oral narratives. *The Modern Language Review*, *96*, 190–208.

Meara, P. (1990). A note on passive vocabulary. *Second Language Research*, *6*, 150–154.
Meara, P. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and Competence in Second Language Acquisition* (pp. 35–53). Cambridge: Cambridge University Press.
Meara, P. (2006). Emergent properties of multilingual lexicons. *Applied Linguistics*, *27*, 620–644.
Meara, P., & Fitzpatrick, T. (2000). Lex30: an improved method of assessing productive vocabulary in an L2. *System*, *28*, 19–30.
Meara, P., & Jones, G. (1990). *The Eurocentres Vocabulary Size Tests*. *10KA*. Zurich: Eurocentres Learning Service.
Meara, P., & Milton, J. (2003). *X_Lex: The Swansea vocabulary levels test*. Newbury: Express Publishing.
Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.
Milton, J. (2010). The development of vocabulary breadth across the CEFR levels. In I. Vedder, I. Bartning & M. Martin (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 211–232). Second Language Acquisition and Testing in Europe Monograph Series 1.
Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use. New perspectives on assessment and corpus analysis* (pp. 57–78). EUROSLA Monographs Series 2.
Milton, J., Alexiou, T., & Mattheoudakis, M. (2014). Knowledge of spoken form. In J. Milton & T. Fitzpatrick (Eds.), *Dimensions of Vocabulary Knowledge* (pp.13–29). Basingstoke: Palgrave Macmillan.
Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, & M. Torreblanca-López (Eds.), *Further insights into non-native vocabulary teaching and learning* (pp. 83–98). Bristol: Multilingual Matters.
Miralpeix, I., & Meara, P. (2014). Knowledge of the written word. In J. Milton & T. Fitzpatrick (Eds.), *Dimensions of Vocabulary Knowledge* (pp. 30–44). Basingstoke: Palgrave Macmillan.
Mora, J. C., & Valls‐Ferrer, M. (2012). Oral fluency, accuracy, and complexity in formal instruction and study abroad learning contexts. *TESOL Quarterly*, *46*, 610–641.
Munro, M. J., & Derwing, T. M. (2001). Modeling perceptions of the accentedness and comprehensibility of L2 speech: The role of speaking rate. *Studies in Second Language Acquisition*, *23*, 451–468.
Nation, I. S. P. (1983). Testing and teaching vocabulary. *Guidelines***,** *5***,** 12–25.
Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Boston, MA: Heinle and Heinle.
Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.
Nation, I. S. P., & Webb, S. (2011). *Researching and Analyzing Vocabulary*. Boston: Heinle Cengage Learning.
Paran, A. (2012). Language skills: Questions for teaching and learning. *ELT journal*, *66*, 450–458.

Pickering, L. (2012). Second language speech production. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 335–348). New York: Routledge.

Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accents in an L2: A review. *Journal of Phonetics*, *29*, 191–215.

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

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, *52*, 513–536.

Saito, K. (2015a). The role of age of acquisition in late second language oral proficiency attainment. *Studies in Second Language Acquisition, 37*, 713-743.

Saito, K. (2015b). Experience effects on the development of late second language learners' oral proficiency. *Language Learning, 65*, 563-595.

Saito, K., & Akiyama, Y. (in press). Video-based interaction, negotiation for comprehensibility, and second language speech learning: A longitudinal study. *Language Learning*.

Saito, K., Trofimovich, P., & Isaacs, T. (2015). Using listener judgements to investigate linguistic influences on L2 comprehensibility and accentedness: A validation and generalization study. *Applied Linguistics*. Advance online publication. doi: 10.1093/applin/amv047

Saito, K., Trofimovich, P., & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics, 37*, 217-240.

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2015). Lexical profiles of comprehensible second language speech: The role of appropriateness, fluency, variation, sophistication, abstractness and sense relations. *Studies in Second Language Acquisition*. Advance online publication. doi: 10.1017/S0272263115000297

Schmitt, N. (2010). *Researching vocabulary*. Basingstoke: Palgrave Macmillan.

Schoonen, R., Van Gelderen, A., De Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing: The role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning*, *53*, 165–202.

Shelton, J. R., & Caramazza, A. (1999). Deficits in lexical and semantic processing: implications for models of normal language. *Psychonomic Bulletin & Review*, *6*, 5–27.

Shelton, J. R., & Weinrich, M. (1997). Further evidence of a dissociation between output phonological and orthographic lexicons: a case study. *Cognitive Neuropsychology, 14*, 105–130.

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*, 510–532.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*, 263–308.

Stæhr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning Journal*, *36*, 139–152.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition, 31,* 577–607.

Trofimovich, P., & Baker, W. (2006). Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition*, *28*, 1–30.

Trofimovich, P., & Isaacs, T. (2012). Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*, *15*, 905–916.

Uchihara, T., & Clenton, J. (2016, September). *Exploring the relationship between productive vocabulary knowledge and second language oral ability assessment*. Paper presented at Pacific Second Language Research Forum (PacSLRF), Tokyo, Japan.

Uchihara, T., Saito, K., & Clenton, J. (2016, September). *Investigating the potential relationships between productive vocabulary and speaking performance in second language speakers*. Paper presented at Vocab@Tokyo international conference on vocabulary acquisition, Tokyo, Japan.

Walters, J. (2012). Aspects of validity of a test of productive vocabulary: Lex30. *Language Assessment Quarterly*, *9*, 172–185.

Xi, X. (2007). Evaluating analytic scoring the TOEFL Academic Speaking Test (TAST) for operational use. *Language Testing*, *24*, 251–286.