Investigating the extent to which vocabulary knowledge and skills can predict aspects of fluency for a small group of pre-intermediate Japanese L1 users of English (L2)

Clenton, J.; Jong, N.H. de; Clingwall, D.; Fraser, S.; Booth, P.

11 Investigating the extent to which vocabulary knowledge and skills can predict aspects of fluency for a small group of pre-intermediate Japanese L1 users of English (L2)

Jon Clenton, Nivja H. de Jong, Dion Clingwall and Simon Fraser

Introduction

The words second language speakers choose to use when speaking may have consequences for their speaking fluency (e.g. Seifart et al., 2018). A number of studies (e.g. De Jong et al., 2013; De Jong & Mora, 2017; Miralpeix & Muñoz, 2018; Milton et al., 2010; Segalowitz & Freed, 2004; Uchihara & Saito, 2016) explore the ways in which the relation between vocabulary knowledge and fluent speech can be evaluated objectively. Such evaluation is important because of the variety and volume of second language speakers, especially of English, whose fluency needs to be assessed, with the importance vocabulary plays in such assessment being absolutely central: 'while without grammar very little can be conveyed, without vocabulary nothing can be conveyed' (Wilkins, 1972: 111–12). Measures of vocabulary knowledge and fluency provide stakeholders, such as those involved in research, pedagogy, and assessment, with essential information to discriminate between users of second languages and their respective proficiency levels. Much research, therefore, is designed to explore the specific features necessary to distinguish between second language users with different levels of language ability. The study we report here adds to this body of research by examining the relationship between the vocabulary knowledge of pre-intermediate Japanese learners of English and their oral fluency.

In this chapter, then, we present a small-scale study in which we employ various vocabulary knowledge tasks as well as fluency elicitation tasks. We compare the results from a number of elicitation tasks not conventionally employed together in the hope not only that this combination of tasks is better suited to the users whose second language we measure, but also that the findings are informative in our investigation of the ways in which vocabulary knowledge relates to aspects of second language fluency.

Before we turn to describing how we understand the term fluency, we briefly outline how we approach the vocabulary knowledge investigated in

this study. As we go on to show, a number of fluency papers (e.g. De Jong et al., 2013) report significant relationships between vocabulary knowledge and aspects of fluency. Such papers elicit learner knowledge from vocabulary tasks alongside fluency tasks and then report that, for instance, learners with specific vocabulary knowledge of X items consistently demonstrate particular aspects of fluency. In this study, we add an additional component to our investigation, one, we believe, that is both novel and unique. In addition to adopting the same approach as referenced above in the current chapter by reporting relationships between vocabulary knowledge and aspects of fluency, we compare the vocabulary used in response both to the vocabulary tasks and the fluency elicitation tasks. We attempt to go to the heart of the vocabulary knowledge of our subject population, and we refer to a recent approach (Fitzpatrick & Clenton, 2017) that supports this investigation.

Before we begin by measuring vocabulary and fluency, we need to detail what is currently meant by the construct of 'vocabulary knowledge'. The construct of vocabulary knowledge is far from straightforward, as Fitzpatrick and Clenton (2017: 844–5) point out, because the 'simplicity' of vocabulary task scores is inconsistent with the multitude of interpretations possible to the extent that 'subtle and informed interpretation is required'. Fitzpatrick and Clenton raise several related concerns when highlighting such complexity. The first pertains to vocabulary measures (e.g. the Vocabulary Levels Test (VLT), Nation, 1983; the Productive Levels Test (PVLT), Laufer & Nation, 1999) that conventionally base elicitation on the assumption that the difficulty of vocabulary items generally relates to their frequency of occurrence in corpora. Second, they highlight the commonly shared view (e.g. Nation 2001, 2013; Read, 2000; Webb, 2009) that vocabulary knowledge is 'multidimensional'. Fitzpatrick (2007) makes this point clearly when comparing three vocabulary tasks according to a revised version of Nation's (1990) 'aspects of word knowledge', indicating that despite being designed to elicit the same, all three tasks test different aspects of the construct. Fitzpatrick and Clenton suggest that Nation's (2001, 2013) 'aspects of word knowledge' is important in this discussion, because it attempts to list the multiple aspects of word knowledge. Fitzpatrick and Clenton add that the complexity of the construct includes a range of factors including those related to the way words are 'organized in the mental lexicon (Meara, 1996), and related to this, speed and, ultimately automaticity of retrieval' (Qian, 2002: 846). The current chapter reports an attempted analysis of the final two of this 'list' of factors: speed and automaticity of retrieval. We return to this specific question in our research questions below. A further concern relates to a distinction commonly made when discussing vocabulary knowledge, of that between productive and receptive knowledge.

'Productive' and 'receptive' are widely-used terms that appear to have gathered currency within the field of vocabulary research. Such terms, however, might need reconsidering in light of suggestions (e.g. Fitzpatrick, 2010; Fitzpatrick & Clenton, 2017) that elicitation relates to the aspects of

vocabulary knowledge measured. Productive vocabulary tasks (e.g. the PVLT; Lex30, Meara & Fitzpatrick; G_Lex, Fitzpatrick & Clenton, 2017; The Lexical Frequency Profile, Laufer & Nation; 1995; The Vocabulary Knowledge Scale, Paribakht & Wesche, 1996) broadly relate to spoken or written output, and receptive vocabulary tasks (e.g. the EVST, Meara & Jones, 1987; the VLT; XY_Lex, Meara & Miralpeix, 2016) broadly relate to reading and listening.

Fluency and vocabulary knowledge

For this study we consider fluency in the narrow sense as opposed to broad sense (Lennon, 1990). While the broad sense of fluency appears to relate to overall or global proficiency, the narrow sense of fluency (for diagnostic purposes) 'refer(s) to one, presumably isolatable, component of oral proficiency' (p. 389). Within narrow fluency, fluency is often measured as a component of speech with multiple aspects, referring to quick and perhaps smooth delivery of speech with or without filled or unfilled pauses, repetitions, and repairs. Some researchers, such as Skehan (2003) and Tavakoli and Skehan (2005), suggest fluency can therefore be measured according to three main characteristics: (i) breakdown fluency (referring to how often speech 'breaks down', or the number of pauses); (ii) speed fluency (referring to the speed of speech between these pauses, therefore articulation rate); and, (iii) repair fluency (referring to the number of times a speaker recognizes and repairs speech). The Common European Framework of Reference (CEFR; Council of Europe, 2001), for instance, refers to proficient users (e.g. C2) as being able to express themselves 'spontaneously, very fluently, differentiating finer shades of meaning even in more complex situations' (p. 5); independent users (e.g. B2) as being able to 'interact with a degree of fluency and spontaneity that makes regular interaction with native speakers quite possible without strain for either party' (p. 5); and basic users (e.g. A2) as being able to 'make themselves understood in short contributions, even though pauses, false starts and reformulations are very evident' (p. 31).

Several papers on fluency (e.g. De Jong et al., 2013; Segalowitz & Freed, 2004; Uchihara & Saito, 2016) have shown strong and significant correlations between fluency measures and productive vocabulary knowledge. De Jong et al. (2013) explore fluency according to several fluency characteristics and report strong and significant correlations between fluency and a newly constructed Dutch version of a sentence completion task (The Productive Vocabulary Levels Test (PVLT), Laufer & Nation, 1999). Their 'intermediate to advanced level' proficiency participants (learners of Dutch as an L2) responded to a variety of tasks, with the study designed to explore linguistic skills and speaking fluency. Fluency was measured from speaking performances in which participants were required to respond to eight computer-administered, semi-spontaneous speaking tasks ranging in terms of complexity, formality, and discourse type. De Jong et al. report a number of limited strength but nevertheless significant (and negative) correlations between productive vocabulary

knowledge and the different fluency measures (silent pauses (r=-0.39), filled pauses (r=-0.33), corrections (r=-0.43), repetitions (r=-0.24), and mean syllable duration (r=-0.58)): participants with higher vocabulary scores tend to produce fewer hesitations, pauses, and a lower mean duration of syllables. More recently, in their study of a participant group whose proficiency was widely varied, Uchihara and Saito (2016) found that fluency, as measured by 'optimal speech rate' (Saito et al. 2015, 2016), moderately predicts productive vocabulary task (Lex30) scores (r=0.34). They compared their fluency analyses with Lex30 task scores on the basis that scores have been shown to be 'representative of [each subject's] productive mental lexicon' (Fitzpatrick & Clenton, 2010: 548). Saito et al.'s oral ability measures required participants to respond to a timed picture description task. Their raters then judged optimal speech rate according to temporal information such as proportion of un/filled pauses and mean length of pauses (Derwing et al., 2004). Taken together, these two studies (De Jong et al., 2013; Uchihara and Saito, 2016) suggest that fluency relates to productive vocabulary knowledge and more broadly to vocabulary skills. The reported fluency measures involved pausing and syllable duration (articulation rate inversed), as well as speech rate. The vocabulary measures consisted of productive vocabulary measures, lexical access speed, and lexical access efficiency. The fluency studies reported in this section appear to broadly reflect a relationship between fluency and productive vocabulary knowledge. Fluent speech appears to correlate with vocabulary task scores to the extent that a second language speaker with a higher vocabulary score hesitates and pauses less and produces a lower mean syllable duration (i.e., a higher articulation rate).

Comparisons between aspects of fluency and receptive vocabulary measures (e.g. De Jong & Mora, 2017; Miralpeix & Muñoz, 2018; Milton et al., 2010), however, are somewhat less consistent. De Jong and Mora (2017) used three of the same speaking tasks as those from earlier studies (De Jong et al., 2013, 2015) and compared data with XY_Lex vocabulary size measures (Meara & Miralpeix, 2016). Their upper-intermediate to advanced proficiency subject vocabulary size scores (M= 6144, *Range* =3350–8200) were shown to correlate moderately significantly (r=-.311) with one aspect of fluency (mean syllable duration) but not with other fluency measures. Miralpeix and Muñoz (2018) investigated the relationships between vocabulary size (using XY_Lex vocabulary measures) with reading, writing, listening, and speaking measures. Their upper-intermediate proficiency subject vocabulary size scores (M=5127, *Range*=2500–7200) were shown to correlate moderately significantly with oral fluency (r=.485). In a comparison of two versions of Yes/No (aural and written) tasks, Milton et al. (2010) compared the vocabulary size (M=2844) with the IELTS tasks designed to elicit knowledge of the four skills (reading, writing, listening, and speaking). While Milton et al. did not find a significant correlation between the written form of the Yes/No task and the speaking scores, their study found a significant correlation between the aural form of the Yes/No task (r=.71) and the

speaking scores. These findings indicate that tasks designed to elicit a specific skill are sensitive to the mode of elicitation, which is supported by the fact that the correlations between X_Lex and reading and writing scores were very similar (r=.70 and r=.76).

The construct of productive vocabulary knowledge

In their discussion of *productive* vocabulary knowledge tasks, Fitzpatrick and Clenton (2017) point to different elicitation tasks eliciting different mean proportions of infrequent items. They suggest that mean score differences relate to different tasks not tapping into the same qualities of word knowledge and therefore '[not sampling] the learner lexicon in the same way' (p. 858). Fitzpatrick and Clenton (2017) have devised a 'Vocabulary Test Capture Model' (see pp. 859–61 for details) in which they adapt a model (the Vocabulary Knowledge Scale, VKS; Paribakht & Wesche, 1993, 1997) originally designed to rank learner knowledge of individual items. The vertical dimension of this scale relates to the nature of the task, to the extent that words are produced in response to a specific task if learner knowledge relates to the four levels. Accordingly, items produced in response to the Lex30 task, which elicits single word responses, might relate to learner knowledge of all of the four levels. Word knowledge would likely be populated with highly frequent items for lower proficiency learners, with an emerging lexicon exhibited by progression through the vertical levels. The horizontal dimension relates to what learners have the capacity to produce in response to each elicitation task. Lex30 activates a different semantic field for each of its 30 cues. On the basis that this chapter discusses fluency studies that have used both Lex30 and the PVLT as their productive vocabulary measures, we compare the PVLT with Lex30 in this section. The findings we present here are from a recent study (Clenton, Elmetaher, and Uchihara, 2019) which reports the different proportions of infrequent items each task elicits (n=107; Lex30 score = 18.41 (SD –10.66); PVLT score = 12.41 (SD –5.83) and show that scores on the two tasks correlate moderately significantly (r = .575, p <.01), to the extent that the capture map might better explain differences between these tasks. Therefore, a task such as the PVLT with its 18 elicitation sentence gaps over its five levels might indicate a less broad capture zone, in contrast to Lex30. Fitzpatrick and Clenton (2017) also suggest that their long arrows (see Figure 11.1) indicate the multiple dip activation events by which responses to tasks such as Lex30 require learners to repeatedly return to the 'same subset of lexical resource, pulling out consecutive items that are closely related' (p. 862); this 'same subset of lexical resource' is not available to PVLT task takers. Correct responses to the PVLT might indicate that participants can demonstrate semantic as well as grammatical mastery of their vocabulary knowledge, indicated by levels 3 and 4 in the model. Figure 11.1 shows a revised vocabulary test capture model, serving to highlight task differences and reasons behind those differences.

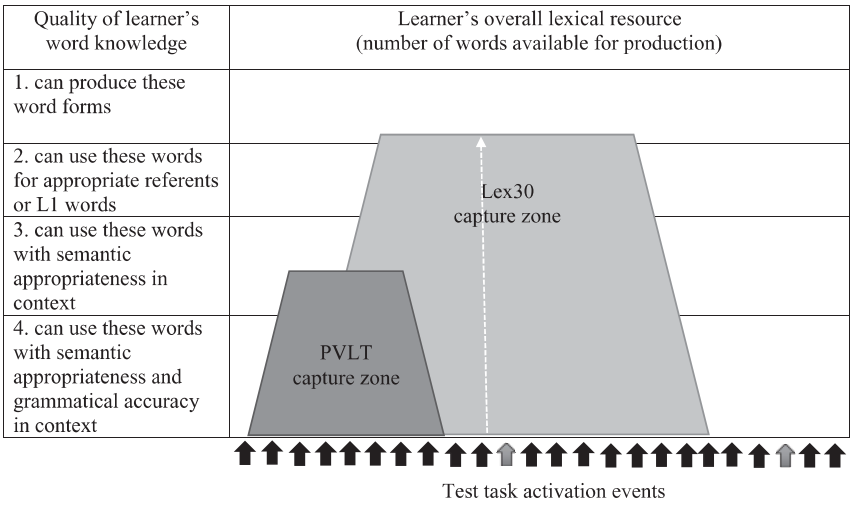

Figure 11.1 A revised vocabulary test capture model: Lex30 and the PVLT.

The study

Aims and research questions

Our main aim is to explore the potential relationships between the knowledge elicited from a productive vocabulary knowledge task and the aspects of fluency elicited from speaking (fluency) tasks. Our second aim is to compare the vocabulary produced in response to the productive vocabulary knowledge task with the vocabulary produced in response to the speaking (fluency) tasks. We also intend to explore findings from earlier papers on fluency in two additional respects, by: (i) comparing receptive knowledge with aspects of fluency; and (ii) exploring the speed and retrieval automaticity, and so including response latency and response duration measures in picture naming tasks in the investigation. The current study, therefore, focuses on the following four questions:

1. Can productive vocabulary knowledge task scores predict aspects of speaking fluency?
2. Can receptive vocabulary knowledge task scores predict aspects of speaking fluency?
3. To what extent do vocabulary skill measures (e.g. response latency and response duration in picture naming tasks) predict aspects of fluency?
4. Is there an overlap between vocabulary used in response to the productive vocabulary task and the vocabulary used in the speaking fluency task?

Methodology

Participants

The participants in the study were 30 pre-intermediate undergraduate adult L1 Japanese learners of English (*M* age = 19, *SD* = 1.3) with an average of 6.5 years' experience of learning English in a school environment; learners had received L2 English instruction for approximately three to four hours a week from L1 Japanese teachers in Japan. They did not use English regularly outside of the learning context. Their X_Lex scores (*M*= 4048, *Range*=2400–4800) also indicated they were of a pre-intermediate proficiency.

Speaking tasks

We chose three speaking tasks from those employed in De Jong et al. (2013), which varied in terms of their task demands: a formal descriptive task (describing a crime scene to a policeman); a formal persuasive task (responding in a town hall meeting to whether a new casino should be built next to an elementary school); and an informal persuasive task (responding to a view on climate change). All tasks were completed on a personal computer. All participants were required to prepare a response and then speak the response aloud. All outputs were recorded. The recordings were subsequently transcribed and analysed using PRAAT (Boersma & Weenink, 2005). The participants were instructed to complete the tasks themselves, and to follow the directions presented on the computer screen. Each task began by presenting participants with a detailed explanation of the situation. Participants were asked to imagine they were speaking for the situation presented. Participants then had a 30-second period within which to prepare their response, indicated by a colour time bar at the bottom of the screen. At the beginning of each task, this coloured bar indicated a time period of two minutes, with the approaching deadline indicated by changing colours, requiring participants to provide their response within the given time.

To measure fluency, using PRAAT (Boersma & Weenink, 2005), syllables were counted manually for all participants. The threshold for a silent pause was set to 350 ms (as De Jong, 2012), and silent pauses were measured manually. All instances of sounds uttered such as *ehh, uhh, mm*, and *umm* were indicated and counted as filled pauses. Similarly, repetitions and repairs were counted manually. All measures were collated over the three tasks. Subsequently, articulation rate was calculated per second of speaking time (total time minus total silent pausing time). Following De Jong and Mora (2017), for all fluency measures indicating hesitations, the total counts were normalized per second total speaking time. Finally, mean silent pause durations for each participant were calculated.

Vocabulary skills tasks

Picture naming: measuring lexical retrieval speed

The same task was used as in De Jong et al. (2013) and De Jong and Mora (2017). From the picture set produced by Snodgrass and Vanderwart (1980), we selected 35 pictures of items all participants were expected to know (i.e. these were highly frequent items). E-Prime was used to present the pictures, one by one. Before the experiment proper commenced, participants were familiarized with the pictures and their names. In this first round, a fixation cross was presented in the middle of the screen for 1000 ms, after which a picture appeared in the centre of the screen, and after yet another 2000 ms, its name was presented underneath the picture. Participants would press the space bar to proceed to the next picture. In the second round, after familiarization, participants were instructed to name (i.e. speak out and name) the pictures as fast and as accurately as possible. In this second round, first, a fixation cross was presented in the middle of the screen for 1500 ms. Then the picture appeared, which was presented for 2000 ms. After the picture, a blank screen followed for 500 ms. The pictures were presented in a random order identical for all participants (but in a different order from the first familiarization round). The time between the appearance of the picture and the beginning of the response was measured manually with the use of PRAAT. Per participant, the mean of all correct responses was used as the measure of lexical retrieval.

Delayed picture naming task: measuring speed of articulation

The materials and apparatus were the same as the ones used for the lexical retrieval measure (picture naming). Following the same picture naming procedure, participants carried out the picture naming task once more. This time, however, they were asked to prepare their response to naming a picture but wait with the actual naming of the picture until the cue was given. A fixation cross was presented in the middle of the screen for 500 ms. Then the picture appeared and remained on the screen for 2000 ms. After 2000 ms, the participant heard a short beep, and a green frame appeared on the screen around the picture. The beep together with the green frame formed the cue for participants to give their response. The picture (with the green frame) remained on the screen for another 1000 ms, during which time the participants responded. The pictures were presented in a random order identical for all participants, but in a different order from the procedures for familiarization and lexical retrieval speed. The experimenter noted incorrect responses and other deviations from the intended responses. Response latency was measured as the latency between the auditory cue and the beginning of the response. Response duration was measured as the duration of the response, i.e. the latency between the beginning and the end of the response.

Both duration measures were measured manually with the use of PRAAT. For each participant, the mean of all correct responses was calculated for both response latency and response duration.

Vocabulary knowledge tasks

Unlike the earlier De Jong et al. (2013) study, which used a Dutch version of Laufer and Nation's Productive Levels Test, we decided to use Lex30 as our productive vocabulary task. We chose Lex30 for four main reasons: (i) Lex30 task scores have been shown to relate to fluency measures (Uchihara & Saito, 2016); (ii) the scores are 'more aligned to the ability to "use" words compared to the PVLT' (Clenton et al., 2019); (iii) we felt the task would better relate to the pre-intermediate proficiency level of our participants (compared, for example, to use of the PVLT in other advanced participant populations (e.g. De Jong et al., 2013); and (iv) to explore the extent to which the vocabulary produced in response to the Lex30 task would match the vocabulary produced in response to the speaking task. Lex30 was created by Meara and Fitzpatrick (2000) in response to issues with other existing productive measures (i.e., PVLT, LFP) at the time of publication. Lex30 has since been used in a wide variety of different papers (Clenton, 2010; Fitzpatrick & Clenton, 2010, 2017; Fitzpatrick & Meara, 2004; Jiménez Catalán & Moreno Espinosa, 2005; Uchihara & Saito, 2016; Walters, 2012). The task requires participants to respond with up to four words to each of the 30 Lex30 cues. Each set of Lex30 responses, a potential 120 items, is processed by correcting misspellings, lemmatizing according to Bauer and Nation's (1993) criteria, and profiling online according to frequency using the Web VP Classic (www.lextutor.ca/vp/eng/). Following the original Lex30 procedure (Meara & Fitzpatrick, 2000), responses contributed to a Lex30 score if they fell outside the first 1000 frequency band and were not proper nouns.

We also used a receptive vocabulary task. A number of studies exploring the relationship between second language fluency and vocabulary knowledge have used receptive measures (e.g. De Jong & Mora, 2017; Milton et al., 2010). We measured the receptive vocabulary of our participants using X_Lex (Meara and Milton, 2003). X_Lex is a computer-based test in which participants are required to respond to whether (120) words presented one at a time are known or unknown. Word knowledge is tested for items from the 1,000-frequency band to the 5,000-frequency band. X_Lex includes pseudo words, and scores are adjusted when such items are identified as genuine.

As well as measuring vocabulary using the productive vocabulary knowledge tasks, we wanted to explore whether this data correlated with the vocabulary used in response to the three speaking tasks. We therefore transcribed the vocabulary produced by participants in response to the three scenario tasks. Corpora generated from our speech data were treated in the same way as in the standard Lex30 task. With the concern that any comparison between Lex30 (written) data and scenario description (spoken) data

is not without difficulties, we turn to an earlier paper (Fitzpatrick & Clenton, 2010). Fitzpatrick and Clenton (p. 546) compared two formats of Lex30, a written and a spoken format. A paired t-test analysis (t = .751, p = .457) indicated that the means between the two tasks did not significantly differ. However, they report that a correlation analysis between the two task scores (r = 0.391, p <.01) was significant but weak, and might have been explained by their participants reluctantly having to respond to their classroom teacher. We base our comparison between the speaking fluency task data and Lex30 data on Fitzpatrick and Clenton's findings, but of course note Fitzpatrick and Clenton's (2010) warning that '(w)e should not assume, then, that the sample of vocabulary produced by a test taker in written mode will exactly mirror that which they produce in spoken mode' (p. 547). Based on this assumption, that the two modes may not exactly mirror one another, we tentatively compared our Lex30 data on the basis that our participants' written responses might approximately reflect their spoken responses. An additional issue we needed to address was which corpus to use in order to make this comparison. Lex30 conventionally uses corpora based on writing. Rather than comparing the vocabulary produced by our participant population with written corpora, we wanted to process data using a spoken word list. We used Dang et al.'s (2017) Academic Spoken Word List (ASWL), on the basis that 'there is a clear-cut difference between the linguistic features of academic speech and academic writing' (p. 978) and, the ASWL 'represented (as closely as possible) the academic speech that EAP learners from a wide range of academic disciplines are likely to encounter in their academic study in English-medium events' (p. 968).

Results

To determine the extent to which measures of vocabulary knowledge and vocabulary skills predict fluency variables in speaking, our dependent variables were: Silent pause duration between ASU (transcriptions were broken down into analysis of speech units[1]), Silent pause duration within ASU, Number of silent pauses per second, Number of filled pauses per second, Number of repetitions per second, Number of corrections per second (all, per second speaking time), and Mean syllable duration. The vocabulary measures used as predictor variables were two measures for vocabulary knowledge: Lex30-score (raw score), and X_Lex-score (corrected score); and three measures for vocabulary skills: LRS (Response Latency – picture naming), RL (Response Latency – delayed picture naming), and RD (Response duration – delayed picture naming). Regarding the extent to which vocabulary *use* in speaking can be predicted by measures of vocabulary knowledge and skills, we use vocabulary as the dependent variable.

Table 11.1 shows the descriptive statistics of all dependent variables as measured from the speaking fluency performances (all fluency variables). Table 11.2 shows the descriptive statistics of all predictor variables (three vocabulary knowledge and three (timed) vocabulary skills measures).

Table 11.1 Descriptive statistics of all dependent variables as measured from the speaking performances (all fluency variables)

Fluency variables	*Mean*	*SD*
Silent pause duration between ASU (ms)	2565.9	1120.34
Silent pause duration within ASU (ms)	1759.2	439.55
Number of silent pauses per second	0.88	0.28
Number of filled pauses per second	0.12	0.14
Number of repetitions per second	0.04	0.01
Number of corrections per second	0.015	0.007
Mean syllable duration (ms)	388	78.08

Table 11.2 Descriptive statistics of all predictor variables (vocabulary knowledge and (timed) vocabulary skills measures)

Vocabulary knowledge	*Mean*	*SD*
Lex30 raw score	40.06	10.43
X_Lex score	4048	476
LRS: Response Latency – picture naming	513.4	141.63
RL: Response Latency – delayed picture naming	749.6	104.30
RD: Response Duration – delayed picture naming	516.1	22.86

Correlations between vocabulary knowledge and skills with fluency measures

Table 11.3 shows the bivariate correlations between the fluency measures, on the one hand, and the vocabulary knowledge and skills measures, on the other. As can be seen in the table, for two measures of fluency, significant correlations with vocabulary knowledge and skills were found. For number of silent pauses per second, the higher the participants scored on the Lex30 task, the fewer pauses were found in their speech samples. At the same time, shorter latencies in the delayed picture task were associated with few silent pauses. Finally, the latency measure in the delayed picture naming task was negatively related to mean syllable durations in the speech samples: participants with short latencies tended to speak slower (with longer syllable durations).

Correlations between productive vocabulary knowledge and vocabulary use

Table 11.4 shows the bivariate correlations between the vocabulary used in the speaking fluency performances on the one hand, and the productive vocabulary knowledge measure (Lex30), on the other. As can be seen from Table 11.4, for the vocabulary used in the speaking task, significant correlations with productive vocabulary knowledge were found. For words used from the level 2 and level 4 of the Academic Spoken Word List (ASWL), the higher participants

Table 11.3 Correlations between vocabulary knowledge and skills with fluency measures (N = 30)

	Silent pause duration between ASU	*Silent pause duration within ASU*	*Number of silent pauses per second*	*Number of filled pauses per second*	*Number of repetitions per second*	*Number of corrections per second*	*Mean syllable duration*
Vocabulary knowledge							
Lex30-score	-.06	-.15	-.39*	-.17	.22	-.02	.12
X_Lex-score	.24	.18	-.16	-.14	.13	.12	.02
Vocabulary skills							
LRS: Response Latency – picture naming	-.04	.08	.31	.18	.08	.28	-.34*
RL: Response Latency – delayed picture naming	.27	.16	.37*	-.12	-.03	-.22	-.44*
RD: Response duration – delayed picture naming	.04	.09	-.21	-.21	-.15	-.09	-.22

*: $p < 0.05$

Table 11.4 Correlations between productive vocabulary knowledge and vocabulary use ($N = 30$)

	Speaking fluency task vocabulary (ASWL levels)				
	Level 1	*Level 2*	*Level 3*	*Level 4*	*Level 0*
Lex30-score	0.341	0.389*	0.234	.0.154	0.400*

*: $p < 0.05$

had scored on the Lex30 task, the more words from this band were found in the speaking performances.

Discussion

The current study was designed to further investigate the extent to which vocabulary knowledge and skills can predict aspects of fluency using several tasks. We have reported on an experiment in which the participants carried out three speaking tasks, and responded to tasks designed to capture their vocabulary skills (picture naming to measure lexical retrieval speed and delayed picture naming task to measure articulation speed), as well as two vocabulary tasks (a productive vocabulary task (Lex30; Meara & Fitzpatrick, 2000), and a test of vocabulary size (X_Lex; Meara & Milton, 2003). We also included an analysis in which the vocabulary used in response to the speaking fluency tasks was correlated with the vocabulary knowledge and skills measures. We can now respond to each of our four research questions.

We first asked whether productive vocabulary knowledge task scores predict aspects of speaking fluency. In broad terms, the findings from the current study are to some extent consistent with earlier fluency studies (e.g. De Jong et al., 2013; De Jong & Mora, 2017). The current study, while using different productive vocabulary knowledge measures, supports De Jong et al.'s (2013) finding that a higher vocabulary score correlates negatively and significantly with the number of silent pauses (Lex30). Regarding this specific correlation, we suggest it relates to Lex30 tapping into aspects of fluent speech to the extent that our pre-intermediate participants potentially used a similar set of highly frequent items from the same or similar frequency bands for the written and fluency tasks. In using Lex30, the current study supports Clenton et al.'s (2019) suggestion that it appears more aligned to the ability to use the words than other productive vocabulary knowledge tasks. This implication we feel is borne out by the significant correlations between the vocabulary used in response to the speaking fluency task and the Lex30 score (Table 11.4), on the basis that participants' lexical resource appears to be shown both in response to Lex30 and to the speaking fluency task. We suggest, however, that at higher levels of proficiency such overlap might not exist to this same extent between

productive vocabulary knowledge task corpora and speaking fluency task corpora, because of the vocabulary size of highly proficient users. While we appreciate that our finding might be exclusive to the proficiency of the participants in the current study, we suggest that this interpretation is important because it appears that the Lex30 task might tap the vocabulary knowledge available to such proficiency groups. Clenton et al. (2019) suggest that some aspects of vocabulary acquisition might lag others to the extent that certain aspects of vocabulary knowledge (e.g. form, which we believe Lex30 accesses) come before others (e.g. semantic, and grammatical knowledge, which the PVLT accesses). We also sense that the current study confirms that Lex30 scores predict aspects of fluency at a pre-intermediate level of proficiency, at least for the specific participants examined here in the current study. However, we suggest that future studies explore suggestions (e.g. Webb & Chang, 2012; Zhang & Lu, 2013) that aspects of vocabulary knowledge develop inconsistently with increases in proficiency. We propose that for studies involving higher-level learners a test such as the PVLT (alongside other productive vocabulary tasks such as Lex30) might help to inform the extent to which the quality of vocabulary knowledge develops with increases in proficiency.

Our second research question was designed to explore the findings from earlier papers on fluency (e.g. De Jong & Mora 2017) that found a significant correlation between receptive vocabulary knowledge task scores and one aspect of speaking fluency. The current study, however, did not find any significant correlations between receptive vocabulary knowledge task scores and the various aspects of speaking fluency. We refer readers to the discussions of our first and second research questions in this case, because we believe that the lack of correlations with the receptive vocabulary measures might relate to the specific proficiency level of our participant group and that this might relate to differences in developing lexicons. Previous fluency related studies (e.g. De Jong et al., 2013; De Jong & Mora, 2017; Miralpeix & Muñoz, 2018) have tended to examine more proficient participants. Such higher-level participants might have developed a receptive vocabulary resource which, we suspect, while not only being larger than that of the pre-intermediate participants that were the focus of the current study might also be more closely related to their productive vocabulary knowledge. The lack of any significant correlation between Lex30 and X_Lex (r = 0.371) might support this finding and runs somewhat counter to earlier Lex30 studies (e.g. Fitzpatrick & Clenton, 2010; Fitzpatrick & Clenton, 2017) that tend to show significant correlations between the receptive and productive vocabulary measures. We suggest that follow-up studies explore this specific finding with perhaps leaners of different (lower and higher) proficiency participants.

Our third research question was designed to explore the extent to which vocabulary skill measures (e.g. response latency and response duration) predict aspects of fluency. This specific question investigates Qian's (2002) suggestion that vocabulary knowledge relates to speed and automaticity of retrieval. The findings here all relate to the timed picture naming tasks in

which participants were required to name pre-primed pictures presented on a screen. Our investigation showed three significant correlations. First, it yielded a significant moderate correlation between response latency-delayed picture naming and the number of silent pauses per second in the speaking tasks (r = 0.37, p < 0.05). In other words, participants who were slower in their response in naming pictures tended to use more silent pauses in their speaking performances. Second, there was also a significant correlation between response latencies in delayed picture naming and mean syllable duration (r = -0.44, p < 0.05). This negative correlation is counterintuitive, in that fast picture-naming speed is related to a slow articulation rate (long syllable duration). The findings we report here are different to those reported in De Jong et al (2013), who found ten significant relations (with n=179), the largest being .32. We speculate that such differences may relate to the different participant proficiency levels and the sample sizes. Accordingly, we suggest that the three findings we report in this chapter are worthy of further examination in additional studies to determine whether aspects of fluency, such automaticity of retrieval and speed of naming, relate differently at different proficiency levels.

Our fourth and final research question asked whether the vocabulary used in response to the productive vocabulary task predicted the vocabulary used in the speaking fluency task. Our findings here show that there is some degree of overlap between responses to the Lex30 task and the speaking fluency task at levels 2 and 0 of the Academic Spoken Word List (ASWL; Deng et al., 2017). This finding, however should be tempered by the comments we presented earlier in our discussion (e.g. Fitzpatrick & Clenton, 2010) that speaking output may not mirror written output. The current study, however, was originally designed to test our first research question, to evaluate the extent to which productive vocabulary knowledge predicts aspects of fluency with perhaps a measure appropriate to the specific proficiency of our participant group. We maintain that this specific finding is, however, worth exploring further and that future such studies could, of course, adopt a spoken Lex30 format in order to test this specific claim. We do suggest, however, that there are potential limitations to this finding that relies on comparing data from the productive vocabulary knowledge task with the speaking fluency task. For our pre-intermediate proficiency participants, we propose that this kind of approach might fit, to the extent that we can observe some degree of overlap. However, with a highly proficient group, we argue that there might only be limited overlap between the productive vocabulary knowledge task and the speaking fluency task. Arguably, because of the limitations of the lexical resource, this approach might only be relevant for lower proficiency levels. We wonder, therefore, up until which proficiency levels this specific approach is relevant. We might suppose, then, that up to a specific proficiency, Lex30 provides a useful indication of the available lexical resource. The extent to which this finding can relate to other proficiencies and to other productive vocabulary tasks, would, we feel, be worthy of further exploration.

Limitations

We acknowledge that, inevitably, there are limitations with the current study, which should not go unreported. The first of these limitations relates to the sample size in the current study. With N = 30, adopting a power of .8 and alpha level of 0.05, we can only expect to find quite large correlations (at least r = 0.49). A second limitation relates to the fact that the current study only explored the vocabulary knowledge of a participant group with the same L1 (Japanese). It is therefore difficult for us to extend the results to other first language groups, because the findings we report here might be limited to L1 Japanese learners. Accordingly, we encourage replications of the current study with different first language populations in order to explore the extent to which our findings represent a potentially bigger picture of the relationships between vocabulary knowledge and fluency. We also propose that other studies consider additional and different vocabulary measures to explore whether different proficiency levels demonstrate greater (or lesser) word knowledge. We suggest that by doing so, such studies might clarify and support the findings we present here. In short, despite its limitations, we believe the current study represents an important development in determining which aspects of vocabulary relate to second language fluency.

Conclusion

This chapter has explored relationships between vocabulary knowledge and fluent speech, but we cannot overextend our findings given the limitations of this small study. We can now report three, albeit tentative, findings. First, based on comparisons of our results with those of earlier studies (e.g. De Jong, 2013) we suggest that relations between vocabulary knowledge and fluent speech may to some extent be proficiency dependent. This can be followed up in future research designed to investigate the potential interaction between proficiency level and the relation between vocabulary knowledge and fluency. Second, there appears to be some degree of overlap between the productive vocabulary used in response to a productive vocabulary task as well as a speaking fluency task. We do not, however, suggest that this finding would be consistent across all proficiency levels, as we discuss above. We suggest, again, that a series of studies of participants at different proficiency levels with the same tools employed in the current study might help shed some light on this finding. Third, we propose that the responses in (delayed) picture naming might relate to vocabulary knowledge in terms of speed and automaticity of retrieval (i.e. in vocabulary skills). We suggest that such measures are interesting and worthy of more research to the extent that different 'vocabulary skills' (such as automaticity of retrieval) relate differently for participants at different proficiency levels.

In order to further vocabulary research within the field of speaking fluency, we urgently need a range of studies to address the issues raised in this chapter. Specifically, we suggest that follow-up studies employ the same fluency and

vocabulary skills tasks as those used here, but we would add that using concurrent productive and receptive vocabulary knowledge tasks at a range of different proficiency levels, and with different first language populations, might shed additional light on our findings. We also encourage research to explore the relationships between an individual's lexical resource, their vocabulary knowledge, and their vocabulary skills. More studies of vocabulary skills are needed to explore the relationships between lexical resource, speed, and automaticity of retrieval.[2]

Notes

1 An AS-unit is 'a single speaker's utterance consisting of an independent clause, or a subclausal unit, together with any subordinate clause(s) associated with either' (Foster, Tonkyn, and Wigglesworth, 2000).

2 The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the JSPS [project number 16K02922].

References

Bauer, L., & Nation, P. (1993). Word families. *International Journal of Lexicography*, *6*(4), 253–279. doi:10.1093/ijl/6.4.253

Boersma, P., & Weenink, D. (2016). PRAAT: Doing phonetics by computer [Computer program]. Version 6.0.23. Retrieved from www.praat.org/.

Clenton, J. (2010). *Investigating the construct of productive vocabulary knowledge with Lex30* (Order No. 10797989). Available from ProQuest Dissertations & Theses A&I. (2008472046). Retrieved from https://search.proquest.com/docview/2008472046?accountid=15247

Clenton, J., Elmetaher, H & Uchihara, T. (2019). Capturing and quantifying individual differences in bilingual learners: A task-based study. Paper presented at Capturing and Quantifying Individual Differences in Bilingualism workshop. Tromsø: Arctic University of Norway,.

Dang, T. N. Y, Coxhead, A., & Webb, S (2017). The academic spoken word list. *Language Learning*, *67*(4), 959–997. ISSN 0023-8333

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). The effect of task complexity on functional adequacy, fluency and lexical diversity in speaking performances of native and non-native speakers. In Housen, A., Kuiken F. &Vedder, I. (eds) *Dimensions of L2 performance and proficiency* (pp. 120–142). Amsterdam: John Benjamins.

De Jong, N. H. (2016). Fluency in second language assessment. In Tsagari, D., and Banerjee, J. (eds) *Handbook of second language assessment* (pp 203–218). Boston, MA/Berlin: Mouton de Gruyter.

De Jong, N. H., Steinel, M. P., Florijn, A., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*(5), 893–916.

De Jong, N.H., Groenhout, R., Schoonen, R., & Hulstijn, J. H. (2015), Second language fluency: Speaking style or proficiency? Correcting measures of second language fluency for first language behavior, *Applied Psycholinguistics, 36*(2): 223–243. doi: 10.1017/S0142716413000210

De Jong, N., & Mora, J. (2017). Does having good articulatory skills lead to more fluent speech in first and second languages? *Studies in Second Language Acquisition*, *1*–13.doi:10.1017/S0272263117000389

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004), Second language fluency: Judgments on different tasks. *Language Learning*, *54,* 655–679. doi:10.1111/j.1467-9922.2004.00282.x

Ellis, N. C. (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *SSLA, 24*, 143–188. doi: 10.1017/s0272263102002024.

Fitzpatrick, T. (2007). Productive vocabulary tests and the search for concurrent validity. In Daller, H.,. Milton, J., & Treffers-Daller, J. (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*(4), 537–554. doi:10.1177/2F0265532209354771

Fitzpatrick, T., & Clenton, J. (2017). Making sense of learner performance on tests of productive vocabulary knowledge. *TESOL Quarterly*, *51*, 844–867. doi:10.1002/tesq.356

Fitzpatrick, T., & Meara, P. (2004). Exploring the validity of a test of productive vocabulary. *Vigo International Journal of Applied Linguistics*, *1*, 55–74.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics*, *21*(3), 354–375.

Heatley, A., Nation, I. S. P., & Coxhead, A. (2002). Range and frequency programs. Available at www.victoria.ac.nz/lals/staff/paul-nation.aspx .

Jimenez Catalan, R., & Moreno Espinosa, S. (2005). Using Lex30 to measure the L2 productive vocabulary of Spanish primary learners of EFL. *Vigo International Journal of Applied Linguistics*, *2*, 27–44.

Kremmel, B., & Schmitt, N. (2017). Vocabulary levels test. In Liontas, J. I. (ed.) *The TESOL Encyclopedia of English Language Teaching*. Hoboken, NJ: Wiley. doi:10.1002/9781118784235.eelt0499

Laufer, B., & Nation, P. (1995). Vocabulary size and use lexical richness in L2 written production. *Applied Linguistics*, *16*, 307–322.

Laufer, B., & Nation, P. (1999). A vocabulary-size test of controlled productive ability. *Language Testing*, *16*(1), 33–51. doi:10.1177/026553229901600103

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning, 40*, 387–417.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. ACL-MIT Press series in natural-language processing. Cambridge, MA: MIT Press.

Meara, P. (1996). The dimensions of lexical competence. In Brown, G., Malmkjaer, K., & Williams, J. (eds) *Performance and competence in second language acquisition* (pp. 35–53). Cambridge: Cambridge University Press.

Meara, P., & Fitzpatrick, T. (2000). Lex30: An improved method of assessing productive vocabulary in an L2. *System*, *28*(1), 19–30. doi: 10.1016/S0346-251X(99)00058-5

Meara, P., & Jones G. (1987) Tests of vocabulary size in English as a foreign language. *Polyglot, 8*(1), [24.4].

Meara, P., & Milton, J. (2003). *X_Lex: The Swansea Levels Test*. Newbury: Express.

Meara, P., & Miralpeix, I. (2016). *Tools for researching vocabulary*. Multilingual Matters.

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In Chacón-Beltrán, R., Abello-Contesse, C., & Torreblanca López, M. (eds) *Further insights into non-native vocabulary teaching and learning* (pp. 83–98). Bristol: Multilingual Matters.

Miralpeix, I., & Muñoz, C. (2018). Receptive vocabulary size and its relationship to EFL language skills. *International Review of Applied Linguistics in Language Teaching, 56*, 1–24.

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 & Heinle Publishers.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge: Cambridge University Press.

Paribakht, T., & Wesche, M. (1993). Reading comprehension and second language development in a comprehension-based ESL program. *TESL Canada Journal*, *11*(1), 9–29. doi: 10.18806/tesl.v11i1.623.

Paribakht, T. S., & Wesche, M. (1996). Enhancing vocabulary acquisition through reading: A hierarchy of text-related exercise types. Canadian Modern Language Review *52*(2), 155–178.

Paribakht, T., & Wesche, M. (1997). vocabulary enhancement activities and reading for meaning in second language vocabulary acquisition. *Second Language Vocabulary Acquisition,* 174–200. doi: 10.1017/cbo9781139524643.013.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning, 52*, 513–536. doi:10.1111/1467–9922.00193

Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.

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*. doi: 10.1093/applin/amv047

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016). Lexical correlates of comprehensibility versus accentedness in second language speech. *Bilingualism: Language and Cognition, 19*(3), 597–609. doi:10.1017/S1366728915000255

Segalowitz, N., & Freed, B. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition, 26*(2), 173–199. doi:10.1017/S0272263104262027.

Seifart, F., Strunk, J., Danielsen, S., Hartmann, I., Pakendorf, B., Wichmann, S., & Bickel, B. (2018). Nouns slow down speech across structurally and culturally diverse languages. *Proceedings of the National Academy of Sciences*, *115*(22), 5720–5725.

Skehan, P. (2003). Task-based instruction. Language Teaching, *36*(1), 1–14. doi:10.1017/S026144480200188X

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 174–215. http://dx.doi.org/10.1037/0278-7393.6.2.174

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure and performance testing. In Ellis, R. (ed.) *Planning and task performance in a second language* (pp. 239–277). Amsterdam: Benjamins.

Uchihara, T., & Saito, K. (2016). Exploring the relationship between productive vocabulary knowledge and second language oral ability. *Language Learning Journal*. doi:10.1080/09571736.2016.1191527

Walters, J. (2012). Aspects of validity of a test of productive vocabulary: Lex30. *Language Assessment Quarterly*, *9*(2), 172–185. doi:10.1080/15434303.2011.625579

Webb, S. (2005). Receptive and productive vocabulary learning: The effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*(1), 33–52. doi:10.1017/S0272263105050023

Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, *28*(1), 46–65. doi:10.1093/applin/aml048

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, *30(*1), 79–95. doi:https://doi.org/10.1017/S0272263108080042.

Webb, S. (2009). The effects of receptive and productive learning of word pairs on vocabulary knowledge. *RELC Journal*, *40*(3), 360–376. https://doi.org/10.1177/0033688209343854

Webb, S., & Chang, A. C.-S. (2012). Second language vocabulary growth. *RELC Journal*, *43(*1). doi: https://doi.org/10.1177/0033688212439367.

Wilkins, D. A. (1972). *Linguistics in language teaching*. Cambridge: MFT Press

Zhang, X., & Lu, X. (2013). A longitudinal study of receptive vocabulary breadth knowledge growth and vocabulary fluency development. *Applied Linguistics*, *35*, 283–304. doi: https://doi.org/10.1093/applin/amt014.

12 Re-examining the relationship between productive vocabulary and second language oral ability

Takumi Uchihara, Kazuya Saito and Jon Clenton

Introduction

Vocabulary knowledge has been viewed as one of the most essential elements of second language (L2) proficiency and development (Meara, 1996), as research investigating the relationship between vocabulary and L2 proficiency has developed to support the long-standing view that vocabulary serves as a proxy for communicative language ability (Miralpeix & Muñoz, 2018). A growing body of research in this area relates vocabulary knowledge to overall proficiency benchmarks (e.g., Common European Framework of Reference for Languages (CEFR) levels; Milton, 2010), in-house placement tests (e.g., Harrington & Carey, 2009), and standardized language proficiency examinations (e.g., International English Language Testing System (IELTS); Milton et al., 2010, or Test of English as a Foreign Language (TOEFL); Qian, 2002), and each of the four language skills (e.g., Laufer & Levitzky-Aviad, 2017 for reading; Wang & Treffers-Daller, 2017 for listening; Baba, 2009 for writing; Uchihara & Clenton, 2018 for speaking). This line of research has been largely devoted to investigating the relationship between vocabulary and reading, yet surprisingly little is known about the role of vocabulary in speaking (cf. Uchihara & Clenton, 2018). Under recent frameworks of L2 speech, L2 oral ability is considered multifaceted in nature, since it comprises a range of different skills related to phonological, fluency, and lexicogrammar (Crossley et al., 2015; Saito et al., 2017). To move the research agenda in L2 vocabulary and speech ahead, the current study attempts to explore whether, to what degree, and how L2 learners' productive vocabulary knowledge is associated with global (comprehensibility), temporal (speed, breakdown fluency), and lexical (appropriateness, variation, sophistication) aspects of L2 oral ability, and to then discuss implications for vocabulary assessment and future research.

Productive vocabulary measures

Researchers agree that vocabulary knowledge encompasses a wide array of word knowledge, characterized by knowledge of form, meaning, and use