# **Exploring L2 relationships between vocabulary size and academic speaking**

**Abstract**

The current paper explores second language (L2) learners’ academic vocabulary size and its potential relationship to speaking in academic contexts. Our participants were 62 first language (L1) Chinese undergraduates of intermediate English level. We elicited speech samples from monologue tasks in formal class settings. We elicited vocabulary knowledge using two yes / no tests, X\_Lex (Meara & Milton, 2003) and the Academic Vocabulary Size Test (AVST; Masrai & Milton, 2018), measuring general and academic vocabulary size, respectively. Our results indicate that AVST scores correlate moderately significantly with several speaking fluency measures and band-based frequency measures of vocabulary use, which outperform the explained variance of X\_Lex. The mid-frequency words (K3) forming 56% of the academic words list is perhaps the primary force for the strong predictive strength of the AVST. These findings provide critical pedagogical implications for L2 academic vocabulary learning and teaching in English for Academic Purposes (EAP) classrooms. The AVST emerges as a potential assessment to predict L2 speaking in academic contexts. We discuss future possible research directions with particular reference to vocabulary research in practice.

 *Keywords:* General vocabulary size, Academic vocabulary size, Lexical diversity, Lexical sophistication, Speaking fluency

**1.1 Introduction**

The current paper expands on the work of earlier investigations (Author, 2021a; de Jong et al., 2013) that have detailed the vocabulary-fluency link, with one important difference: we explore the relationship in academic settings. The current study is critical because it represents a first attempt to explore second language (L2) learners’ academic vocabulary size and its potential relationship to speaking in academic contexts. Earlier vocabulary-fluency studies have alluded to a relationship between vocabulary size and second language (L2) speaking, with a focus on vocabulary size covering a range of different frequency levels: The first frequent band (K1) - the third frequent band (K3) level (Koizumi & In’nami, 2013), K1-K4 level (Noreillie et al., 2020), K1-K5 level (Author, 2021a; Koizumi & In’nami, 2013), and K1-K10 level (de Jong & Mora, 2019; de Jong et al., 2013; Miralpeix & Munoz, 2018; Author, 2020). However, none of these studies have examined the specific contribution of academic vocabulary size to L2 speaking in academic environments. We suggest that studies investigating vocabulary size with a ceiling of 10,000 (10K) might cover academic word knowledge. Yet, the extent to which academic vocabulary size can predict speaking in academic contexts remains unclear and under-researched. While studies (e.g., Author, 2020; Noreillie et al., 2020) have examined spoken output profiles according to Lexical Frequency Profile (LFP, Laufer & Nation, 1995), only one such study (Author, 2021a), to our knowledge, reports spoken profiles with spoken academic word lists (i.e. Academic Spoken Word List (ASWL, Dang et al., 2017). Dang, Coxhead and Webb (2017) provide a powerful means to explore academic spoken words from a broad academic subject coverage. With the demand for verbalising opinions in academic language settings, the ASWL (Dang et al., 2017) offers a necessary means to explore academic vocabulary knowledge in academic contexts. We suggest that exploring the potential relationships between academic vocabulary size and speaking in academic contexts is now essential.

The current study responds to these critical research threads by comparing general vocabulary size (K1-K5 level) and academic vocabulary size (primarily K3 level; Masrai & Milton, 2018), on the one hand, and L2 speaking in academic context, on the other hand. The motivation for the scope of this investigation stems from two earlier studies. First, in an investigation of vocabulary size and its coverage, Milton (2009) proposed that 3000 words or more were required for full comprehension in an academic specialist context. Second, Koizumi and In’nami (2013) reported a significant finding of K1 to K5 frequency bands to L2 speaking proficiency.

## **2. Literature review**

## **2.1 L2 speaking fluency research and speaking size**

Fluency is considered a central scoring rubric in high-stakes assessments (e.g., International English Language Testing System, IELTS) and language ability descriptors (e.g., Common European Framework of Reference, CEFR). According to the IELTS (2021) speaking band descriptor, “fluency and coherence” is one of the critical assessment criteria, together with lexical resource, grammatical range and accuracy, and pronunciation. Fluency is also considered a central component in CEFR (2001) descriptors, in which they describe users, such as the proficient language users at mastery level (C2) who “can express him/herself spontaneously at length with a natural colloquial flow, avoiding or backtracking around any difficulty so smoothly that the interlocutor is hardly aware of it.” (p. 28-19) Despite its common usage, definitions of fluency are not straightforward and conventionally viewed as being either broad “a cover term for oral proficiency” (Lennon, 1990, p. 389) or narrow “one, presumably isolatable, component of oral proficiency” (Lennon, 1990, p. 389). Tavakoli and Skehan (2005) identify the measurable aspects of fluency as: breakdown (concerning the amount and the location of silent and filled pauses), repair (related to measures of repetitions, corrections, false starts and other strategies speakers use in monitoring and modifying their utterances), and speed (referring to the rate at which the speech is produced). We adopt Tavakoli and Skehan’s terms in the experiment reported below.

Studies seeking to explore potential relationships between vocabulary size and L2 speaking fluency (Milton et al., 2010; Miralpeix & Muñoz, 2018; Noreillie et al., 2020; Author, 2020) have tended to adopt one of two research approaches: a subjective approach, or an objective approach. We present a brief survey of studies adopting a subjective approach and an objective approach in our literature survey below.

Studies adopting a subjective approach tend to employ human raters to provide holistic scores based on overall performance (e.g., Milton et al., 2010; Noreillie et al., 2020). Comparing the explained variance of X\_Lex and A\_Lex, the written and aural Yes/No vocabulary size tests of the first five frequency bands (K1-K5), respectively, to IELTS speaking scores, Milton et al. (2010) reported strong and significant correlations between IELTS overall speaking proficiency (*r =* .71) and A\_Lex (*Mean* = 2384) but no such significant correlation between X\_Lex (*Mean* = 2844) and IELTS overall speaking proficiency. Working with French speakers, Noreillie et al. (2020) reported significant and moderate correlations between receptive vocabulary test scores and holistic ratings for one of their dialogic speaking tasks (a doctor’s visit) but not for the other speaking task (a job interview). Other human-rater studies detail the rating of several subcomponents, such as pronunciation, grammar, accuracy, fluency, and vocabulary (e.g., Miralpeix & Muñoz, 2018; Author, 2020). In Miralpeix and Muñoz (2018), upper-intermediate / advanced L2 Spanish learners’ speaking was were? assessed by ratings on fluency, pronunciation and grammar/vocabulary. Their L2 learners’ vocabulary size (*Mean* = 5100), as measured by X/Y\_Lex (Yes/No vocabulary size tests assessing words from the first five frequency bands (K1-K5) and the next five (K6-K10)) (Meara & Miralpeix, 2006) correlated with moderate significance with oral fluency (*r =* .485). Author’s (2020) advanced level L2 Japanese participants’ speaking was were? assessed by IELTS speaking descriptors covering fluency, pronunciation, vocabulary and grammar. Their L2 learners’ vocabulary size (*Mean* = 6058) correlated significantly with their human rating vocabulary aspect (*r =* .552).

Studies adopting an objective approach have tended to examine the vocabulary size-fluency link by calculating various aspects of breakdown, repair, and speed fluency (e.g., Author, 2021a; de Jong & Mora, 2019; Hilton, 2008; Koizumi & In’nami, 2013). de Jong and Mora’s (2019) and Author’s (2021) approaches examined the relationship between vocabulary size and three dimensions of utterance fluency. In the study of de Jong and Mora (2019), X/Y\_Lex was used to measure the participant vocabulary size of 10,000 words while the speech was elicited via three speaking tasks (a formal descriptive task of B1 level, a formal persuasive task, and an informal persuasive task of B2 level). The study of de Jong and Mora (2019) reported a significant correlation between vocabulary size (*Mean* = 6144) and mean syllable duration (255 syllables per ms; *r =* -0.311) while employing a similar speech elicitation method, Author’s (2021) found no evidence of a size-speaking link in their pre-intermediate level learners (*Mean* = 4048), in which X\_Lex was used to measure the vocabulary size of 5000 words. Other studies (Hilton, 2008; Koizumi & In’nami, 2013) have achieved consistent and significant results related to the relationship between vocabulary size and speaking. Elicited from a short video sequence description, Hilton (2008) reported that novice to advanced learners’ vocabulary size scores, as measured via DIALANG, correlated significantly with all of their fluency measures, especially with the temporal fluency measures – words per minute (*r = .*581), mean length of run (*r = .*668), percentage of hesitation (*r =* -0.551) and rate of hesitation per 1000 words (*r =* -0.661). Koizumi and In’nami (2013) also report that vocabulary size alone explained up to 63% of the variance in speaking proficiency with novice to lower-intermediate level L2 learners, and that vocabulary size could explain 29% (*r = .*54) to 48% (*r = .*69) of the speed fluency and 11% (*r = .*33) to 12% (*r = .*34) of the repair fluency.

The above studies highlight the mixed results regarding the explanatory power of vocabulary size to speaking fluency that span 0% - to 63%. Drawing on these studies, we observe that vocabulary size can predict many aspects of fluency, including words per minute, mean syllable duration, hesitation rate, self-correction, and repetition rate. In the current study, to explore objective evidence of the relationship between speaking and academic vocabulary size, we adopt an objective approach to measure quantifiable features of fluency and the extent they relate to the size of academic vocabulary that is yet studied.

## **2.2 Vocabulary use in L2 speaking**

Investigations exploring vocabulary use have tended to employ lexical diversity or lexical sophistication measures (e.g., Appel et al., 2019; Kyle & Crossley, 2015; Li & Lorenzo-Dus, 2014; Lu, 2012; Noreillie et al., 2020; Saito et al., 2016; Suzuki & Kormos, 2020). Lexical diversity focuses on the number of unique words in a text or the range of words. L2 speaking studies found indices of lexical diversity can be significantly associated with holistic human-rater judgments of speaking proficiency (Appel et al., 2019; Lu, 2012; Noreillie et al., 2020; Suzuki & Kormos, 2020). Based on three dimensions of lexical richness (density, diversity, and sophistication), Lu (2012) analysed L2 Chinese learners’ oral narrative proficiency with 26 lexical indices. She found test-takers oral narratives were significantly related to nine of 20 lexical diversity indices. Following Lu (2012) and using a series of lexical measures, Appel et al. (2019) identified that the Measure of Textual Lexical Diversity (MTLD) was the sole significant predictor of L2 comprehensibility and nativeness from two speaking tasks, a picture description and a Test of English as a Foreign Language (TOEFL) task. Similarly, Noreillie et al. (2020) reported that the diversity of their participant output could best predict raters’ holistic scores. However, when using an argumentative task to elicit 40 Japanese learners’ linguistic dimensions in their speaking performances, Suzuki and Kormos (2020) reported lexical diversity was not a prominent predictor. Instead, speed and breakdown fluency were perceived as their two best predictors of comprehensibility and perceived fluency. Such lexical diversity studies appear to support findings from lexical sophistication studies when predicting speaking proficiency (Li & Lorenzo-Dus, 2014; Kyle & Crossley, 2015; Saito et al., 2016). Lexical sophistication is defined as the proportion of high-frequency and low-frequency words used in a text (Read,2005, p.14). Li and Lorenzo-Dus (2014) found the use of “advanced words” from 25 raters’ verbal protocols related to vocabulary scores in speech. Kyle and Crossley (2015) defined lexical sophistication more broadly by including range, *n*-gram frequency, academic lists, and word information. Drawing on a TOEFL speaking corpus, Kyle and Crossley (2015) examined the predictive validity of 135 indices of lexical sophistication to speaking proficiency holistic scores. They report that almost half of their 101 indices of lexical sophistication measures statistically correlated with speaking proficiency holistic scores and that five indices could explain up to 52% of the variance. Using a series of variables including diversity, sophistication, and fluency, Saito et al. (2016)’s study found these three aspects, represented by MTLD (*r = .*72), word familiarity (*r = .*53), and filler ratio (*r = .*72), were all significantly correlated with L2 comprehensibility ratings, to varying degrees.

In the current study, we adopt a variety of vocabulary measures. We do this because we want to explore the extent to which our various measures apply to our participants’ contexts. This is in line with McCarthy and Javis (2010). They found lexical diversity measures (MTLD, vocd-D, and Maas) did not necessarily measure the same latent trait underneath. They, therefore, suggested several of its measures should be considered in one study. Following their suggestion, we included several diversity measures in the present study to better understand the relationship between vocabulary size and the diversity aspect of vocabulary used in L2 learners’ speaking. In terms of measures of lexical sophistication, we adopt Read’s (2000) definition of the narrow sense of sophistication that focuses on the proportion of frequency words in a learner’s text. Following Author (2020), we measured both count-based and band-based aspects (Crossley et al., 2013) of lexical sophistication in the current study.

## **Research questions**

Our literature survey highlights the view that vocabulary size and speaking study results, when taken together, are far from conclusive and that little is known about the extent to which academic vocabulary size and speaking in genuine academic contexts relate. Moreover, evidence from recent studies (e.g., Noreillie et al., 2020; Author, 2020) profiling spoken output information remains scarce in terms of academic use in an academic context. A further issue relates to participant levels; since most vocabulary-speaking studies report participants with higher proficiency levels (e.g., de Jong & Mora, 2019; Miralpeix & Muñoz, 2018; Author, 2020), it is unclear whether findings from previous studies applied to participants of intermediate level. Therefore, the current study aims to respond to these concerns by contributing to discussions related to vocabulary-speaking, investigating an academic context, and focusing on intermediate learners. Accordingly, the present study asks the following research questions:

1. Can vocabulary size predict L2 speaking fluency in an academic context?

2. Can vocabulary size predict vocabulary use, range-based diversity, and frequency-based sophistication, of L2 speaking in an academic context?

## **3. Method**

## **3.1 Participants**

Participants were 62 L1 Chinese undergraduates (35 female, 27 male) with an average of 490 at College English Test Level 6 (CET 6, Chinese national English proficiency test), equivalent to a B1-B2 level according to the CEFR (2001, p.24). According to a Kolmogorov-Smirnov test, these scores were normally distributed (Field, 2009, p. 147). Before the study, participants had received L2 English instruction for 90 minutes per week from L1 Chinese teachers for three semesters. Each semester had 16 teaching weeks. They did not use English regularly outside of the learning context.

## **3.2 Measuring vocabulary size**

To measure receptive vocabulary size, we used a Yes / No test format for two reasons: it has been widely used and accepted in earlier research (e.g., Milton & Hopkins, 2006; Mochida & Harrington, 2006; Roche & Harrington, 2018; Zhang et al., 2020); and, it receives validation from use in earlier vocabulary-fluency studies (e.g., Author, 2021a; de Jong & Mora, 2017; Author, 2020). A more recent developed Academic Vocabulary Size Test (AVST, Masrai & Milton, 2018), developed from the Yes / No test format, also reported good reliability, content, and construct validity.

### ***3.2.1 General vocabulary size (5000 words)***

Participants’ general vocabulary size was measured with X\_Lex (Meara & Milton, 2003), a Yes / No test, covering the first 5000 frequent words of Kilgarriff’s (2006) BNC lemmatised frequency list (accessible at <http://www.kilgarriff.co.uk>). 20 words were selected and combined from each frequency band with four pseudowords consisting of 100 actual words and 20 pseudo words. Each response (“Yes” to an accurateword) was given a credit of 50 points to achieve a total of 5000, and a false hit (“Yes” to a pseudo word) was given a penalty of 100 points to have an equivalent score of 5000 to adjust participants vocabulary size from guessing. No credit or penalty was given if participants missed (“No” to an accurate word) or rejected a pseudo word (“No” to a pseudoword) (See Appendix A for a sample image of the X\_Lex task).

### ***3.2.2 Academic vocabulary size (570 words)***

Participants’ academic vocabulary size was measured with the Academic Vocabulary Size Test (AVST, Masrai & Milton, 2018). Following a Yes / No test format and the same test development principle of X\_Lex, at a sample rate of 1:5, Masrai and Milton (2018) selected 19 words from each frequency band of the Academic Word List (AWL, Coxhead, 2000), leading to a total of 114 words representing 570 academic word families. To adjust the final score from guessing, 19 rare words were included. Unlike Masrai and Milton’s study (2018), we included pseudo words in this test as we believe the response to pseudowords could provide more convincing feedback to our participants. We adopted the same marking criteria as Masrai and Milton (2018) that a hit should be given a credit of 5 points to achieve a total of 570, and a false hit was given a penalty of 30 points for an equivalent score of 570.
Similarly, no credit or penalty was given if participants missed or rejected a pseudo word (See Appendix B for a sample image of the AVST test). The rationale for employing the AVST based on the AWL is that participants in the present study and many L2 learning contexts are more exposed to written academic English than spoken academic English. One reason is the dominant L1 instruction in the academic classes. Another potential reason is explained by the fact that L2 learners generally spend more time reading than listening.

## **3.3 Measuring academic speaking**

The current study employed an argumentative speech task (see Appendix C) to elicit participants’ academic speaking performance in the classroom. Participants were explicitly instructed to provide supporting examples/information for their arguments and were not given time to practice in advance or limited speaking time. All the elicited speech samples, 1-2 minutes long on average, were recorded on personal computers, were transcribed, and analysed using Praat (Boersma & Weenink, 2005). In the current study, L2 learners’ academic speaking was assessed using various aspects of speaking fluency; and with diversity and sophistication focused vocabulary to use in the speech production.

### ***3.3.1 Speaking fluency***

To measure speaking fluency, we targeted measures of total duration and three aspects of utterance fluency (breakdown, repair, and speed fluency) as used in recent fluency studies (e.g., de Jong & Mora, 2019; Author, 2021a). For breakdown fluency, we calculated silent pause duration between Analysis of Speech Units (ASU1, Foster et al., 2000), silent pause duration within ASU, mean silent pause duration, number of silent pauses per second, and number of filled pauses per second. We counted the number of repetitions and the corrections per second for repair fluency. Finally, we calculated the mean syllable duration (sounding time divided by the total number of syllables) to measure speed fluency.

Consistent with previous studies (i.e. Author, 2021a), we adopted a threshold of 350 ms for the silent pause. With the script written in Praat (Lennes, 2021), the duration of speaking and silent pausing time was marked and calculated automatically. Silences shorter than 350 ms were discarded. The speaking transcripts were then broken down into ASU (Foster et al., 2000) to calculate silent pause duration between and within ASU. The number of syllables, silent pauses, filled pauses, repetitions, and corrections were counted manually.

### ***3.3.2 Lexical diversity measures***

Traditional diversity measures are based on calculating the ratio of the number of Tokens and Types (TTR). One primary concern of this calculation is its sensitivity to sample size, as the ratio tends to decrease when the sample size increases (Zenker & Kyle, 2021). To respond to this issue, several revised indices have been proposed by researchers, including the Diversity (D) measure (Malvern et al., 2004), the Hypergeometric Distribution Diversity index (HD-D; McCarthy & Jarvis, 2007), and MTLD (McCarthy & Jarvis, 2010). Drawing on different text lengths, studies tended to agree that newer diversity measures (e.g., D, HD-D, MTLD) are least affected by the text length and are better indicators of the quality of leaners’ output (Koizumi & In’nami, 2012; McCarthy & Jarvis, 2007; 2010; Zenker & Kyle, 2021). We, therefore, employed these three indices (D, HD-D, and MTLD) of lexical diversity measures in the present study.

The D measure was calculated using the VocD utility of CLAN (MacWhinnery, 2000) after covering text files in the Codes for the Human Analyses of Transcripts (CHAT) format. The average of the three analyses served as the D value of each sample text. HD-D and MTLD were calculated using the Tool for the Automatic Analysis of Lexical Diversity (TAALED, Kyle et al., 2021). In general, a higher value of lexical variation indicates low word repetition and, therefore, a high proportion of more unique words.

### ***3.3.3 Lexical sophistication measures***

Following Author (2020), we used count-based and band-based frequency measures of lexical sophistication (Crossley et al., 2013). Regarding count-based frequency measure, each participant speaking text was given two corpus-based frequency scores, spoken-frequency-log Corpus of Contemporary American English (COCA) and spoken-frequency-log British National Corpus (BNC), obtained from Tool for the Automatic Analysis of Lexical Sophistication (TAALES, Kyle et al., 2018).

Each participant’s text was given a lexical frequency profile for band-based frequency measure based on a corpus-based word list. In contrast to Author (2020), where the L2 texts were profiled according to general vocabulary frequency levels (Laufer & Nation, 1995), the L2 spoken texts in the current study were profiled in line with the ASWL (Dang et al., 2017) covering four frequency levels (Level 1, 2, 3 and 4). We used the AntWord profiler (Anthony, 2021) for the band-based frequency profiling for all participant speaking texts.

## **4. Results**

Our independent variables were two receptive vocabulary size measures: X\_Lex and the AVST scores. To determine the extent to which receptive vocabulary size predicts variables in speaking fluency, the dependent variables include total duration (ms), silent pause duration between ASU (ms), silent pause duration within ASU (ms), number of silent pauses per second, number of filled pauses per second, number of repetitions per second, number of corrections per second and mean syllable duration (ms). To determine the extent to which receptive vocabulary size predicts variables in vocabulary use of speech, the dependent variables are: VocD, HD-D, and MTLD as measures of lexical diversity; and spoken-frequency-log (COCA) and spoken-frequency-log (BNC) as measures of count-based lexical sophistication; Level 1 to Level 4 as measures of band-based lexical sophistication.

In the following section, descriptive statistics of independent and dependent variables are presented, followed by correlation analyses to examine the relationship between vocabulary size and measures of L2 speaking. Finally, regression analyses were conducted to explore how much variance in the L2 academic speaking can be predicted by general and academic vocabulary size.

## **4.1 Descriptive analyses**

 Table 1 shows the descriptive statistics of all independent variables (X\_Lex and the AVST). As table 1 shows, participants’ vocabulary knowledge is an average of 4119 general words and an average of 421 academic words.

*[Insert Table 1 here]*

Descriptive statistics in terms of measures of speaking fluency are shown in Table 2. The mean total speaking duration was 77.4 seconds, with an average production of 138 tokens. We found that the silent pause duration within ASU was longer than the silent pause duration between ASU, indicating participants tended to have a fluency breakdown within a complete speech unit. The total breakdown fluency count (silent pauses and filled pauses) was 0.46 times per second, of which the number of silent pauses per second was about twice higher than the number of filled pauses. Of two measures of repair fluency, the repetitions strategy was used more frequently than the self-correction strategy. In terms of speed fluency, participants generally produced their speech at the rate of 287 syllables per millisecond.

 *[Insert Table 2 here]*

 The descriptive statistics of dependent variables related to lexical diversity and lexical sophistication are shown in table 3 and table 4. Higher values of three diversity indices (D, HD-D, and MTLD) and two sophistication indices (Frequency-log (COCA) and Frequency-log (BNC)) mean higher diverse output and a higher proportion of low-frequent words, respectively. The band-based frequency profile shows that the most spoken words produced by participants belonged to the Level 1 band (*Mean =* 128 words), weighing 95% of the total words. This is followed by the Level 2 band (*Mean =* 4 words), weighing approximately 3%.

 *[Insert Table 3 and 4 here]*

## **4.2 Correlation analyses**

Before conducting correlation analysis, a Kolmogorov-Smirnov (K-S) test was conducted to examine the normality of each sampling data (Field, 2009). The results in table 5 showed that while some variables were normally distributed (sig. > .05), some were not (sig. < .05). Therefore, we employed Pearson’s correlation analysis for normally distributed variables and Spearman’s for non-normally distributed variables in the correlation analysis.

*[Insert Table 5 here]*

***4.2.1 Correlation analyses between vocabulary size and speaking fluency measures***

 Table 6 reported the results from the correlation analyses between two vocabulary size tests and speaking fluency measures. We found that both X\_Lex (*r* (60) = -.29, *p* < .05) and the AVST (*r* (60) = -.30, *p* < .05) achieved negative and moderate significant correlations with number of repetitions per second. Compared with X\_Lex, more fluency measures showed significant correlations with the AVST. A moderate positive significant correlation was found between the AVST and the total duration (*r* (60) = .27, *p* < .05) and a moderate negative significant correlation between the AVST and the number of silent pauses per second (*r* (60) = -.30, *p* < .05), whereas X\_Lex did not achieve any significant relationship with these two fluency variables. The results yielded no evidence of a significant correlation between vocabulary size and speed fluency, as measured by mean syllable duration, which is in line with the results reported in Author (2021a).

 *[Insert Table 6 here]*

***4.2.2 Correlation analysis between vocabulary size and vocabulary use measures in the speech***

##  To respond to the second research question, we conducted correlation analyses between X\_Lex / the AVST and lexical diversity indices and between X\_Lex / the AVST and indices of lexical sophistication. Table 7 shows no significant correlation between vocabulary size and diversity measures or between vocabulary size and two count-based frequency measures. However, moderate and positive significant correlations were found between academic vocabulary size and two band-based frequency measures, Level 1 (*r* (60) = .33, *p* < .01) and Level 2 (*r* (60) = .36, *p* < .01), of which words used from Level 2 of the ASWL (Dang et al., 2017) achieved the highest correlations with the AVST.

 *[Insert Table 7 here]*

Overall, the correlation analyses suggest that participants who knew more words tended to have less repetition, but knowing more academic words would help participants produce longer spontaneous speech with fewer pauses. However, the articulation rate in L2 spontaneous speech did not relate to the vocabulary size, whether the words belonged to a more general frequency level or the academic word list. Moreover, knowing more academic words appeared to be helpful in assisting learners in producing more academic words in the academic speech, especially words from the mid-frequency (Level 2).

## **4.3 Regression analyses**

We conducted multiple regression analyses to investigate the predictive strength of vocabulary size over L2 speaking fluency. Based on the results of correlation analyses (Table 6), the dependent variables were total duration (ms), number of silent pauses per second, and number of repetitions per second. The independent variables were X\_Lex and the AVST, of which collinearity (VIF=1) was examined, so the multicollinearity was not violated (VIF≦1, Field, 2009, p. 224). Log10 transformations were applied, so assumptions of normality, linearity, and homoscedasticity of residuals were met before regression analyses.

Results from the regression analyses showed that AVST was the only significant predictor. X\_Lex did not appear to be a significant predictor in predicting L2 speaking fluency in an academic context. Table 8 summarises multiple regression analyses (stepwise method) with the AVST as the only significant predictor. The *R²* within each model showed that the AVST can predict 7% of the variance of the total speaking duration (*R²* = .07, *F*(1, 61) = 4.45, *p* = .039); 8% of the variance of the number of silent pauses per second (*R²* = .08, *F*(1, 61) = 5.03, *p* = .029); and 12% of the variance of the number of repetitions per second (*R²* = .12, *F*(1, 50) = 6.70, *p* = .013).

*[Insert Table 8 here]*

Similar steps were followed by examining the predictive strength of vocabulary size over L2 vocabulary use. We found that about 7% of the words at ASWL Level 1 and 12% at Level 2 could be predicted by the academic vocabulary size (Table 8).”

## **5. Discussion**

The primary aim of the current study was to investigate the vocabulary-speaking link in academic settings. This investigation included two aspects: the relationship between vocabulary size and L2 learner’s speaking fluency; and the relationship between vocabulary size and L2 learner’s vocabulary use in the speech. In contrast with previous research (Author, 2021a; de Jong & Mora, 2019; de Jong et al., 2013; Koizumi & In’nami, 2013; Miralpeix & Munoz, 2018; Noreillie et al., 2020; Author, 2020) which targeted the contribution of general vocabulary size to L2 learners’ speaking, the current study examined the speaking in an academic context, using argumentative speaking task to elicit the specific contribution of academic words to speaking fluency and vocabulary use.

## **5.1 Research question 1 - Vocabulary size and speaking fluency**

The first research question was designed to explore the extent to which vocabulary size could predict academic speaking fluency. The correlation analysis findings show that both vocabulary sizes achieve moderate and significant correlations with repair fluency (number of repetitions per second), with X\_Lex explaining 9.6% (*R2*) and the AVST 8.4% (*R2*) of the variance of repair fluency. However, when the two vocabulary size measures were combined, the AVST outperformed X\_Lex and became the single significant predictor of L2 speaking fluency in an academic context. These results are slightly lower than those from Koizumi and In’nami (2013), who found that 11% - 16% of the repair fluency was explained by general vocabulary size up to 5000 words. This difference might be explained by the different English levels of these two studies: participants were at an intermediate level in the current study, while in Koizumi and In’nami (2013)’s study, the participants were at a novice to lower-intermediate level. To compare the shared variance of two vocabulary size tests to academic speaking fluency, we found that academic vocabulary size related more closely to fluency variances. While X\_Lex only correlated with repair fluency, the AVST significantly correlated with speaking duration and breakdown fluency (number of silent pauses per second). In other words, learners with more academic vocabulary were likely to produce longer and smoother speech. This is in line with Lu (2012), who found learners with higher oral fluency ratings tended to produce longer speech.

A comparison of the results from the current study with the Author’s (2021a) study is also worthwhile considering both studies used very similar measures of speaking fluency and participants of similar vocabulary size (*Mean =* 4048 in Author’s study and *Mean =* 4119 in the current study). While Author (2021a) did not report a significant correlation between X\_Lex and their fluency variables, the results of the current study are encouraging since we report that other than X\_Lex, the AVST appeared to be more effective in supporting the vocabulary size-speaking fluency link, especially in explaining the variance of the speech length (total duration) and the breakdown fluency (number of silent pauses).

The current study found no supporting evidence regarding the relationship between vocabulary size and speed fluency. This finding is counter to de Jong and Mora’s (2019) finding of a significant relationship between vocabulary size and mean syllable duration (MSD = 255); however, the current study findings support Author (2021a), who also yielded no supportive evidence. A further comparison of these three studies (Table 9) suggests that the vocabulary size and speed fluency link among intermediate-level learners does not appear as close and prominent among advanced-level learners. We recommend that future studies explore this finding with learners, perhaps from a range of proficiency levels, to explore this consideration further.

*[Insert Table 9 here]*

## **5.2 Research question 2 - Vocabulary size and vocabulary use in the speech**

 The second research question was designed to explore the extent to which vocabulary size can predict range-based lexical diversity and frequency-based lexical sophistication in speech output. In response to this second research question, output words from Level 1 and Level 2 of ASWL achieved moderate significance with participant academic vocabulary size, but not with words from the same general frequency level. Results from the regression analyses further confirmed the predictive strength of academic vocabulary size over L2 vocabulary use in the speech, to the extent that the AVST could predict up to 7% of the Level 1 words and 12% of the Level 2 words of the spoken words often appeared in an academic context. This finding indicates that knowing a more significant number of academic words tends to help learners produce suitable words in academic contexts, highlighting the importance of academic word learning and teaching in EAP (English for Academic Purposes) classrooms. This finding, however, might best be further investigated with larger sample size, especially for words falling into both level 3 and level 4.

 We also found diversity and count-based frequency measures were not significantly related to participants’ vocabulary size. This finding is contrary to earlier studies (e.g., Appel et al., 2019; Kyle & Crossley, 2015; Lu, 2012; Noreillie et al., 2020; Suzuki & Kormos, 2020; Author, 2020) that identify a certain degree of oral proficiency can be explained by measures of diversity and sophistication. One potential reason for this absence of a finding might relate to the speaking task feature number, to the extent that the usage of one speaking task might not fully mirror participant vocabulary use (when compared to the use of multiple speaking tasks (e.g., two tasks in Appel et al., 2019). Therefore, it is possible that participants with an extensive vocabulary size might not necessarily produce text with rich lexical features due to the task feature, a further avenue for subsequent studies.

 Whilst examples of the use of one elicitation task are elsewhere in the vocabulary-speaking literature (e.g., Suzuki & Kormos, 2020), and the results remain encouraging, we still need to account for the non-significant relationship between vocabulary size and the lexical features in the current study. Compared to subjectively assessed studies (Appel et al., 2019; Kyle & Crossley, 2015; Lu, 2012; Noreillie et al., 2020; Suzuki & Kormos, 2020; Author, 2020), the current explorative study’s use of objective measures to calculate lexical characteristics of speaking output might be a challenge especially when the sample size appeared relatively small (*Mean =* 1 token from Level 3 and *Mean =* 1 token from Level 4 of ASWL) and the average text length was relatively short (*Mean =* 138 tokens). Unlike human rater studies, which can be more sensitive to unique/advanced words despite sample size and sample length, machine calculation is likely influenced by such factors and therefore become less sensitive to rare words usage in text. The current study finding is in line with Author (2020), who also report participant vocabulary size is not significantly correlated to words produced above the K3 level, even though their sample size was relatively large (*Mean =* 5127). Still, they found vocabulary size could explain up to 30% of vocabulary ratings from human raters. This highlights the complementary role of machine calculation to human judgments of speaking proficiency and further lends weight to encouraging further studies that explore both approaches concurrently.

 For the AVST, we reported a close relationship with various aspects of speaking, so it might be a potentially effective means to predict L2 learners’ academic speaking. To illustrate the difference between X\_Lex and the AVST, we also compared two-word lists, Kilgarriff’s first 5000 frequent word list and Coxhead’s AWL list, on which the two vocabulary size tests were developed. To make these two lists comparable, we processed them using the BNC-COCA via VocabProfile (Cobb, 2002) and employed the word family as the counting unit. Figure 1 shows the frequency distribution of these two lists on BNC-COCA as percentages. As the figure shows, about 56% of the words from the AWL are concentrated on the third *1000* frequency level (K3) and 23% on the second *1000* frequency level (K2), which is in line with Masrai and Milton’s study (2018) (that the AWL is heavily loaded with words at K3 levels). Therefore, we suggest that vocabulary at the K3 level is an essential predictor of L2 learners’ academic speaking, especially related to the total speaking length and the frequency of breakdown fluency. It may be the case that mid-frequency words cause the AVST to outperform X\_Lex in predicting L2 speech in an academic context. We would also add that this finding is somewhat tentative and worthy of further investigation to evaluate the claims we make here. For L2 learners spending much longer time in the target country, other studies might be necessary to explore the predictive strength of spoken academic word size to their academic speaking when learners have much longer time exposed to the academic spoken language. While we would add that objective measures have the advantage of reduced costs compared to human raters, we suggest that a significant benefit to using objective measures relates to the need to support the large numbers of English (L2) learners in L1 Chinese contexts; employing objective tools might be a more desirable option to cater for mass assessment inside the classroom or beyond.

## **Conclusions and limitations**

Our study primarily investigated the extent to which general vocabulary size / academic vocabulary size predicts speaking in an academic context as measured via fluency and vocabulary use. The current study results are encouraging in supporting the vocabulary size and speaking link and are supportive of the specific tasks we employed in our study. The strong predictive power of academic vocabulary size we report to various aspects of fluency measures and vocabulary used in the speech suggested the benefit of learning academic words, especially the mid-frequency words (K3), to improve aspects of L2 learner speaking fluency, breakdown, and repair fluency in particular; and speaking length in an academic context. We propose further studies designed to investigate academic usage in EAP classrooms.

Additionally, objective measures from the current study yielded similar results to previous studies that used subjective measures (e.g., Miralpeix & Munõz, 2018 on the link between vocabulary size and fluency; Noreillie et al., 2020, and Author, 2020 on the connection between vocabulary size and vocabulary use), the results from the current study are encouraging as they provide evidence for the need for large scale objective measurement of learner speaking proficiency. However, we recommend that future studies consider using subjective approaches concurrently with objective approaches, especially when sample sizes are small, to compensate for the potential insensitivity to rare words with machine calculation.

Finally, as measures in the current study were conducted in a Chinese higher education context, the finding provides insights into both domestic Chinese students and Chinese students abroad studying in an English academic context. The objective approach might offer an efficient means to assess students’ academic speaking to guide the academic study.

Given the scope of this research and the limitation of the dataset, our research has a few limitations. First, one speaking task type used in the assessment might have limited examining the lexical characteristics of learners’ output. Future studies could employ more tasks in examining the relationship between vocabulary size and academic speaking, covering more task types in an academic context. However, the results from the current study are still representative to a certain extent, as was found in many previous studies with a similar situation (e.g., Suzuki & Kormos, 2020). Second, the spoken text selected for this study was a monologue-type task. Our findings cannot be generalised to the relationship between academic vocabulary size and academic speaking in the broad context of academic speaking contexts. Instead, we believe that they contribute valuable insights into the importance of learning academic vocabulary in improving academic speaking in general monologic situations. Last, the present study’s focus was on single lexical items’ knowledge and their relationship to speaking. While multiword units are largely available in academic language, it is worthwhile to look into this vocabulary size-speaking link on n-grams (combination of n number of words), especially when collocation knowledge has proven useful in improving oral proficiency (e.g., Author, 2021b).

**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).

**Table 1**

*Descriptive statistics of independent variables (vocabulary size)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Predictor variables | Min | Max | Mean | SD |
| X\_Lex | 2737 | 4789 | 4119 | 439 |
| AVST | 204 | 542 | 421 | 83.50 |

**Table 2**

*Descriptive statistics of dependent variables - speaking fluency measures*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Speaking fluency measures | Min | Max | Mean | SD |
| Total duration (ms) | 27220 | 182540 | 77404 | 34859.27 |
| *Breakdown fluency* |  |  |  |  |
| Silent pause duration between ASU (ms) | 1130 | 19050 | 6457.26 | 3428.19 |
| Silent pause duration within ASU (ms) | 910 | 70710 | 14936.14 | 14483.97 |
| Mean silent pause duration (ms) | 433.44 | 1876.67 | 800.90 | 332.80 |
| Number of silent pauses per second | 0.11 | 0.52 | 0.33 | 0.10 |
| Number of filled pauses per second | 0.00 | 0.36 | 0.13 | 0.09 |
| *Repair fluency* |  |  |  |  |
| Number of repetitions per second | 0.00 | 0.18 | 0.05 | 0.05 |
| Number of corrections per second | 0.00 | 0.10 | 0.02 | 0.03 |
| *Speed fluency* |  |  |  |  |
| Mean syllable duration (ms) | 144 | 409 | 287 | 48.53 |

Note: ms refers to milliseconds.

**Table 3**

*Descriptive statistics of dependent variables - lexical diversity measures*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lexical diversity measures | Min | Max | Mean | SD |
| D | 18.45 | 135.57 | 57.25 | 20.76 |
| HD-D | 0.62 | 0.86 | 0.73 | 0.05 |
| MTLD | 18.83 | 98.58 | 42.26 | 14.65 |

**Table 4**

*Descriptive statistics of dependent variables - lexical sophistication measures*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lexical sophistication measures | Min | Max | Mean | SD |
| *Count-based* |  |  |  |  |
| Frequency-log (COCA) | 2.94 | 3.40 | 3.18 | 0.10 |
| Frequency-log (BNC) | -0.03 | 0.50 | 0.20 | 0.10 |
| *Band-based (Tokens)* |  |  |  |  |
| Level 1 | 54 | 331 | 128.24 | 59.18 |
| Level 2 | 0 | 18 | 3.95 | 3.53 |
| Level 3 | 0 | 4 | 0.77 | 1.03 |
| Level 4 | 0 | 7 | 1.37 | 1.55 |

**Table 5**

*Tests of normality*

|  |  |
| --- | --- |
|  | Kolmogorov-Smirnova |
| Statistics | df | Sig. |
| *Independent variables* |  |  |  |
| X\_Lex | .10 | 61 | .200\* |
| AVST | .08 | 61 | .200\* |
| *Dependent variables - fluency measures* |  |  |  |
| Total duration (ms) | .14 | 61 | .008 |
| Silent pause duration between ASU (ms) | .12 | 61 | .042 |
| Silent pause duration within ASU (ms) | .17 | 61 | .000 |
| Mean silent pause duration (ms) | .18 | 61 | .000 |
| Number of silent pauses per second | .09 | 61 | .200\* |
| Number of filled pauses per second | .10 | 61 | .200\* |
| Number of repetitions per second | .17 | 61 | .000 |
| Number of corrections per second | .19 | 61 | .000 |
| Mean syllable duration (ms) | .08 | 61 | .200\* |
| *Dependent variables - vocabulary use measures* |  |  |  |
| D | .11 | 61 | .082 |
| HD-D | .09 | 61 | .200\* |
| MTLD | .12 | 61 | .036 |
| Frequency-log (COCA) | .04 | 61 | .200\* |
| Frequency-log (BNC) | .06 | 61 | .200\* |
| Level 1 | .17 | 61 | .000 |
| Level 2 | .18 | 61 | .000 |
| Level 3 | .32 | 61 | .000 |
| Level 4 | .41 | 61 | .000 |

1. Lilliefors Significance Correction

**Table 6**

*Correlations between vocabulary size and speaking fluency measures*

|  |  |  |
| --- | --- | --- |
|  | X\_Lex | AVST |
| Speaking fluency measures | *r* | *r* |
| Total duration (ms) | .17 | .27\* |
| *Breakdown fluency measures* |  |  |
| Silent pause duration between ASU (ms) | -.02 | -.01 |
| Silent pause duration within ASU (ms) | .04 | .06 |
| Mean Silent pause duration (ms) | .00 | .04 |
| Number of silent pauses per second | -.21 | -.30\* |
| Number of filled pauses per second | -.04 | .40 |
| *Repair fluency measures* |  |  |
| Number of repetitions per second | -.29\* | -.30\* |
| Number of corrections per second | .04 | .12 |
| *Speed fluency measures* |  |  |
| Mean syllable duration (ms) | .06 | .02 |

Note: \* refers to significance at the 0.05 level (2-tailed).

 \*\* refers to significance at the 0.01 level (2-tailed).

**Table 7**

*Correlations between vocabulary size and vocabulary use measures*

|  |  |  |
| --- | --- | --- |
|  | X\_Lex | AVST |
| Vocabulary use measures | *r* | *r* |
| *Lexical diversity measures* |  |  |
| D | -.12 | -.08 |
| HD-D | -.12 | -.06 |
| MTLD | -.00 | -.05 |
| *Lexical sophistication measures* |  |  |
| *Count-based frequency measures* |  |  |
| Spoken-frequency-log (COCA) | .20 | .23 |
| Spoken-frequency-log (BNC) | .11 | .15 |
| *Band-based frequency measures* |  |  |
| Level 1 | .20 | .33\*\* |
| Level 2 | .20 | .36\*\* |
| Level 3 | .06 | .21 |
| Level 4 | .06 | .20 |

Note: \* refers to significance at the 0.05 level (2-tailed).

 \*\* refers to significance at the 0.01 level (2-tailed).

**Table 8**

*Summary of L2 speaking predicted by academic vocabulary size*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | R | R2 | Adjusted R2 | Std.Error of the Estimate |
| Total duration (ms) | .26 | .07 | .05 | .18 |
| Number of silent pauses per second | .28 | .08 | .06 | .13 |
|  Number of repetitions per second | .35 | .12 | .10 | .34 |
| Level 1 | .26 | .07 | .05 | .13 |
| Level 2 | .35 | .12 | .11 | .90 |

**Table 9**

*Correlation between vocabulary size and speaking fluency measures*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Vocabulary size | MSD | *r* between size and MSD |
| Author (2021a) | 4048 | 388 | 0.02 |
| Current study | 4119 | 287 | 0.06 |
| de Jong and Mora (2019) | 6144 | 255 | -0.31\* |

Note: \* refers to significance at the 0.05 level (2-tailed).

**Figure 1**

*Frequency distribution of Kilgarriff’s word list and AWL on BNC-COCA (percentages).*

#

# **References**

Author (2020)

Author (2021a)

Author (2021b)

Appel, R., Trofimovich, P., Saito, K., Isaacs, T., & Webb, S. (2019). Lexical aspects of comprehensibility and nativeness from the perspective of native-speaking English raters. *International Journal of Applied Linguistics,* *170*(1), 24-52. https://doi.org/10.1075/itl.17026.app

Anthony, L. (2021). AntWordProfiler (Version 1.5.1) [Computer Software]. https://www.laurenceanthony.net/software

Beeckmans, R., Eyckmans, J., Janssens, V., Dufranne, M., & Van de Velde, H. (2001). Examining the Yes/No test: Some methodological issues in theory and practice. *Language Testing*, *18*(3), 235-274. https://doi.org/10.1177/026553220101800301

Boersma, P., & Weenink, D. (2021). *Praat: Doing phonetics by computer* [Computer program]. Version 6.1.42. Retrieved from [www.praat.org/](http://www.praat.org/)

Cobb, T. (2002). *The Web Vocabulary Profiler*. https://www.lextutor.ca/vp/

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354-375. https://doi.org/10.1093/applin/21.3.354

Lennes, M. (2021) *Speech corpus toolkit for Praat* [Toolkit]. https://lennes.github.io/spect/

CEFR (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. https://rm.coe.int/16802fc1bf

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213-238. https://doi.org/10.2307/3587951

Crossley, S. A., Cobb, T., & McNamara, D. S. (2013). Comparing count-based and band-based indices of word frequency: Implications for active vocabulary research and pedagogical applications. *System, 41*(4), 965-981. https://doi.org/10.1016/j.system.2013.08.002

Dang, T. N. Y., Coxhead, A., & Webb, S. (2017). The academic spoken word list. *Language Learning*, *67*(4), 959-997. http://doi.org/10.1111/lang.12253

Daller, H., & Wang, Y. (2017). Predicting study success of international students. *Applied Linguistic Review*, *8*(4), 355-374, <https://doi.org/10.1515/applirev-2016-2013>

de Jong, N. H., & Mora, J. (2019). Does having good articulatory skills lead to more fluent speech in first and second languages? *Studies in Second Language Acquisition*, *41*(1), 227-239. https://doi.org/10.1017/S0272263117000389

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*(1), 5–34. https://doi.org/10.1017/S0272263111000489

de Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2013). Linguistic skills and speaking fluency in a second language. *Applied Psycholinguistics*, *34*(5), 893–916. https://doi.org/10.1017/S0142716412000069

Eyckmans, J. (2004). *Measuring receptive vocabulary size. Reliability and validity of the Yes/No vocabulary test for French-speaking learners of Dutch*. Utrecht: LOT.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). London, UK: Sage Publications Ltd.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied Linguistics, 21*(3), 354-375. https://doi.org/10.1093/applin/21.3.354

Hilton, H. (2008). The link between vocabulary knowledge and spoken L2 fluency. *The Language Learning Journal*, *36*(2), 153–166. https://doi.org/10.1080/09571730802389983

IELTS (2021, May) Speaking: Band descriptors (public version) https://www.ielts.org/-/media/pdfs/speaking-band-descriptors.ashx?la=en

Kilgarriff, A. (2006). *BNC database and word frequency lists* [Data set]. [http://www.kilgarriff.co.uk/bnc-readme.html#lemmatised](http://www.kilgarriff.co.uk/bnc-readme.html%22%20%5Cl%20%22lemmatised)

Koizumi, R., & In’nami, Y. (2012). Effects of text length on lexical diversity measures: Using short texts with less than 200 tokens. *System, 40*(4), 554–564. https://doi. org/10.1016/j.system.2012.10.012

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*(5), 900–913. http://doi:10.4304/jltr.4.5.900-913

Kyle, K., & Crossley, S. A. (2015). Automatically assessing lexical sophistication: Indices, tools, findings, and application. *TESOL Quarterly, 49*(4), 757-786. https://doi.org/10.1002/tesq.194

Kyle, K., Crossley, S. A., & Berger, C. (2018). The tool for the analysis of lexical sophistication (TAALES): Version 2.0. *Behavior Research Methods, 50*, 1030-1046. https://doi.org/10.3758/s13428-017-0924-4

Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity using direct judgements. *Language Assessment Quarterly, 18*(2), 154-170. https://doi.org/10.1080/15434303.2020.1844205

Laufer, B., & Nation, I. S. P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics*, *16*(3), 307–322. https://doi.org/10.1093/applin/16.3.307

Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, *40*(3), 387–417. <http://doi:10.1111/j.1467-1770.1990.tb00669.x>

Lennon, P. (2000). The lexical element in spoken second language fluency. In H. Riggenbach (Eds.), *Perspectives on fluency* (pp. 25-42). University of Michigan Press. <https://dialnet.unirioja.es/servlet/libro?codigo=662496>

Li, H., & Lorenzo-Dus, N. (2014). Investigating how vocabulary is assessed in a narrative task through raters’ verbal protocols. *System, 46*, 1-13. https://doi.org/10.1016/j.system.2014.06.006

Lu, X. (2012). The relationship of lexical richness to the quality of ESL learners’ oral narratives. *The Modern Language Journal*, *96*(2), 190-208. <https://doi.org/10.1111/j.1540-4781.2011.01232_1.x>

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk.* Mahwah, USA: Erlbaum.

Malvern, D. D., Richards, B. J., Chipere, N., & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment.* Houndmills, UK: Palgrave Macmillan. DOI: 10.1057/9780230511804

Masrai, A., & Milton, J. (2018). Measuring the contribution of academic and general vocabulary knowledge to learners' academic achievement. *Journal of English for Academic Purposes*, *31*, 44-57. <https://doi.org/10.1016/j.jeap.2017.12.006>

McCarthy, P. M., & Jarvis, S. (2007). A theoretical and empirical evaluation of vocd. *Language Testing, 24*(4), 459-488. https://doi.org/10.1177/0265532207080767

McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behaviour Research Methods, 42*, 381-392. https://doi.org/10.3758/BRM.42.2.381

Meara, P., & Miralpeix, I. (2006). *Y\_Lex: The Swansea Advanced Vocabulary Levels Test,* *v.2.05*. Swansea, UK: Lognostics.

Meara, P., & Milton, J. (2003). *The Swansea levels test*. Swansea, UK: Express publishing.

Milton, J. (2009). *Measuring Second Language Vocabulary Acquisition.* Bristol, UK: *M*ultilingual Matters. https://doi.org/10.21832/9781847692092

Milton, J., & Hopkins, N. (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *Canadian Modern Language Review*, *63*(1), 127-147. https://doi.org/10.3138/cmlr.63.1.127

Milton, J., Wade, J., & Hopkins, N. (2010). Aural word recognition and oral competence in a foreign language. In M. R. Chacón-Beltrán, C. Abello-Contesse, & M. M. Torreblanca-López (Eds.), *Further insights into non-native vocabulary teaching and learning* (pp. 83–98). Bristol, UK: 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), 1–24. https://doi.org/10.1515/iral-2017-0016

Mochida, A., & Harrington, M. (2006). The Yes/No test as a measure of receptive vocabulary knowledge. *Language Testing*, *23*(1), 73–98. https://doi.org/10.1191/0265532206lt321oa

Noreillie A. N., Desmet, P. & Peters, E. (2020). Factors predicting low-intermediate French learners’ vocabulary use in speaking tasks. *The Canadian Modern Language Review, 76*(3), 194-217. <https://doi.org/10.3138/cmlr-2019-0018>

Read, J. (2000). *Assessing vocabulary*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1017/CBO9780511732942

Roche, T., & Harrington, M. (2018). Offshore and onsite placement testing for English pathway programmes. *Journal of Further and Higher Education, 42*(3), 415-428. https://doi.org/10.1080/0309877X.2017.1301403

Saito, K., Webb, S., Trofimovich, P., & Isaacs, T. (2016). Lexical profiles of comprehensible second language speech: The role of appropriateness, fluency, variation, sophistication, abstractness, and sense relations. *Studies in Second Language Acquisition,* *38*(4), 677-701. https://doi.org/10.1017/S0272263115000297

Segalowitz, N. (2010). *Cognitive bases of second language fluency* (1st ed.). Routledge. https://doi.org/10.4324/9780203851357

Suzuki, S., & Kormos, J. (2020). Linguistic dimensions of comprehensibility and perceived fluency: An investigation of complexity, accuracy, and fluency in second language argumentative speech. *Studies in Second Language Acquisition, 42*(1), 143-167. https://doi.org/10.1017/S0272263119000421

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and Task Performance in a Second Language* (pp. 239-273). John Benjamins. https://doi.org/10.1075/lllt.11

Zenker, F., Kyle, K. (2021). Investigating minimum text lengths for lexical diversity indices. *Assessing Writing 47*,1-15. https://doi.org/10.1016/j.asw.2020.100505

Zhang, X., Liu, J., & Ai, H. (2020). Pseudowords and guessing in the Yes/No format vocabulary test. *Language Testing*, *37*(1), 6-30. <https://doi.org/10.1177/0265532219862265>

**Appendix A – X\_Lex test**

****

**Appendix B – Academic Vocabulary Size Test (AVST)**

**Appendix C – Speaking task**

It is a big advantage in life to be an only child in a family. To what extent do you agree with this statement? Use examples / evidence to support your argument. There is no time limitation.