International Journal of TESOL Studies (2023)
Vol. 5 (4) 37-54 https://doi.org/10.58304/ijts.20230403

Article

Exploring the Impact of Lexical Threshold on Listening Comprehension

Sonia W. L. Cheung
The University of Hong Kong, Hong Kong, China

Received: 23 September 2022; Accepted: 8 May 2023; Published: 5 November 2023

Abstract
Despite the prevalence of exam-focused textbooks and practice tests, few studies have examined how students react to their input. A listener may need to reach a certain proficiency level to be able to benefit from the input and listening practice. Given the importance of L2 vocabulary knowledge for listening comprehension, this study used multiple methods to investigate the relationship between lexical threshold and listening comprehension. A vocabulary level test was designed to measure whether the participants knew 2000 high frequency words in English. Four classes of 73 secondary students in Grade 11 in Hong Kong responded to four listening tasks during their end of term examination. Cluster analysis was used to group students on the basis of their listening proficiency. Three profiles of students were identified to infer what they could comprehend based on the amount of lexical knowledge they had. Clear trends emerged in terms of listening fluency. Students in the cluster with high listening scores reached or approached the lexical threshold. They displayed an ability to identify main ideas and specific details, taking notes fluently. The borderline group showed readiness for comprehension when listening support was provided. The third group showed low achievement in English learning largely unable to grasp key words. The paper discusses the challenges faced by students in listening comprehension with reference to lexical threshold. Implications for curriculum design and materials development are discussed.

Keywords
Listening fluency, vocabulary knowledge, needs analysis, practice tests, Chinese learners

1 Introduction

Language proficiency threshold and the provision of language support to students in English-medium university programmes is being increasingly investigated in light of second language (L2) learners' difficulties (e.g. Aizawa et al., 2020; Harrington & Roche, 2014; McKinley et al., 2021). One potential reason for language-related problems in tertiary education is insufficient vocabulary among students who enter English-medium instruction (EMI) universities directly from a mother tongue (L1) medium school.

Email: soniacheung2007@gmail.com

Research on language threshold has suggested that L2 learners are at an academic disadvantage until they reach adequate proficiency (Ardasheva et al., 2012).

In recent years, threshold concepts, which "represents a transformed way of understanding, or interpreting, or viewing something without which the learner cannot progress", has gained prominence in discussions about teaching and learning (Meyer & Land, 2006, p. 3). It is believed that by understanding certain important fundamentals of a discipline, learners will successfully integrate knowledge of that discipline and their learning will be transformed. Identifying and characterising threshold concepts are seen as critical steps in curriculum design (Meyer & Land, 2006). Carson (2017) provided a case of communicative language use as a threshold concept to conceptualize a curriculum for academic English programmes in a Japanese university. Students were engaged in discussions that aimed to transform language classes from learning language elements to using the language for communication. It was suggested that the threshold concept can provide a framework to support students for academic success. Although the concept of a threshold is different from threshold levels, both are central to second language learning. One reason is that the learner may not progress in communicative language use until they possess a critical mass of vocabulary. Research has tended to focus on a lexical threshold, i.e., requiring knowledge of a certain number of words, and reading comprehension (e.g. Laufer & Ravenhorst-Kalovski, 2010; Schmitt et al., 2011), rather than listening comprehension.

To make advances in curriculum, materials and pedagogy, the present study investigates the relationship between lexical threshold and listening comprehension, assuming that the listener needs to reach a certain level of vocabulary knowledge to be able to efficiently process a text in coursebooks and practice tests. "Threshold" in this study refers to the minimum level of vocabulary knowledge that students need to identify the main ideas of a listening passage, or to pass a test. The practical value of the study lies in its implications for curriculum design and materials development.

2 Literature Review

Studies have been conducted to identify the relative contribution of different factors to listening comprehension. Speed rate, unknown vocabulary, and insufficient contextual knowledge are among the difficulties facing second/foreign language (L2) learners when listening. Central to these investigations is that a better understanding of the factors contributing to L2 comprehension can guide curriculum design and classroom teaching. The present study focuses on three dimensions of interest, namely, the listener, the text, and the task to support student learning.

The "listener" refers to internal factors that affect the listener's L2 ability to process a text (Rubin, 1994). In Vandergrift and Baker's (2015) study on learner variables, L2 vocabulary knowledge had the strongest and most consistent influence on listening comprehension. L2 vocabulary knowledge is a stronger predictor than working memory, sentence processing speed, grammar, and metacognitive awareness (Andringa et al., 2012; Oh, 2016; Wang & Treffers-Daller, 2017; Wolfgramm et al., 2016). Studies have also suggested there is a strong association between L2 vocabulary knowledge and academic listening in results on standardised tests like CPE and IELTS (Stæhr, 2009; Teng, 2016). It is therefore important to expand learners' vocabulary size in the L2 classroom (Stæhr, 2009).

A listener needs to reach a certain level of proficiency to be able to efficiently process spoken language (Lynch, 1998). A body of research has suggested that language users need to know between 2,000- and 3,000-word families (rather than individual words) to be able to cope with the lexical demands of everyday conversation (van Zeeland & Schmitt, 2013); Nation (2013) suggests a minimum of 95% coverage is sufficient for listening comprehension. According to Dang and Webb's (2014) analysis of 160 lectures and 39 seminars of the British Academic Spoken corpus, reaching 95% coverage of academic

spoken English requires the knowledge of a minimum of 3000 word-families plus the *Academic Word List*. To compare, Averil Coxhead's *Academic Word List* contains 570 words frequently found in college reading (Hirsh, 2015), and Michael West's *General Service List* contains 2,000 basic words for learning English as a foreign language (Richards, 2001, p. 8). This latter list represents the minimum number of words that operate together in the greatest variety of contexts. Vocabulary knowledge is also categorised into six threshold levels according to the Common European Framework of Reference for Languages: beginner (A1 and A2), intermediate (B1 and B2) and mastery (C1 and C2) (Hawkins & Filipović, 2012). CEFR thresholds require a lexical syllabus to guide curriculum planning and materials development. The underlying principle of a lexical syllabus is to provide L2 learners with meaningful exposure to texts and tasks using the most common words. Importantly, a lexical approach advocates the use of corpora to inform pedagogical materials, as well as a recycling strategy to help students develop vocabulary knowledge (Harwood, 2002).

Providing students with vocabulary lists is not sufficient for improving their performance in listening assessments (Chang, 2007). One reason is that vocabulary lists present words to students in isolated written form without developing their ability to automatically recognise the phonological form (Chang & Millett, 2013). A component of vocabulary knowledge is fluency (Laufer & Goldstein, 2004), which involves the speed of comprehension, and the retrieval and production of words. Fluent word recognition occurs at sub-lexical levels, which is not subject to top-down influence from syntactic and semantic processing (Hulstijn, 2001, p. 265). When listeners do not recognise a word, they shift from automatic processing to strategic comprehension (Aryadoust, 2019).

Texts represent input for students to repeat encounters with the target language. Research on listening input includes text modification (Rubin, 1994), L2 captions (Hayati & Mohmedi, 2011), and audio glosses (Antes, 2014), and guidelines on simplification and modification have been devised (Buck, 2001, p. 169-171; Rost, 2011, p. 172). It is believed that texts should be scripted to "go in the direction of rendering a text accessible to learners" (McGrath, 2002, p. 105). Another criterion for preparing input is "authentic processing", especially for low proficiency listeners (Buck, 2001, p. 169). For authenticity, learners should be brought to the point where they can understand and respond to spoken English at normal speed (Lynch, 2009, p. 100). Yet, scripted texts lack textual and phonological characteristics of real world spoken English. It has been argued that incorporating unscripted texts into the teaching of listening comprehension promotes L2 learners' communicative competence (Wagner, 2014).

Apart from suitable texts, students need listening comprehension practice, particularly to identify key words. Listening for words requires bottom-up processes to extract information (Richards, 2005). Students need to develop rapid word recognition to match the aural form of a word with the word in their mental lexicon (Vandergrift, 2006). Tasks include, but are not limited to, narrow listening (Dupuy, 1999), dictogloss (Prince, 2013; Vasiljevic, 2010), and vocabulary instruction (Pan et al., 2018). It is, however, not clear whether metacognitive listening training has a positive effect on lower proficiency learners of English (Milliner & Dimoski, 2021). Much of the literature advocates the use of authentic materials in teaching. For example, Hubackova (2011) and Kuo (2010) explored gap-filling to scaffold authentic listening. Teachers specified component skills – phonological knowledge, word recognition, paraphrasing or fluency – and designed gap-filling exercises to engage university students in listening to radio broadcasts. Generally, listening activities may be designed to engage students to listening to a recording a second time to complete a cloze exercise, which raises their language awareness before moving on to using the texts for productive activities (Richards, 2005). However, previous research gives little information about the characteristics of the input and how students react to it.

Accordingly, the present study investigates the relationship between the lexical threshold and L2 listening comprehension. It examines the extent to which the input, or texts scripted for coursebooks and practice tests by educational publishers, is understood by the listener with reference to lexical threshold. Student responses to a listening test were collected to provide evidence of learning as well as serving as

data for a needs assessment (see Richards, 2013). Cluster analysis was used to identify learner profiles that would have potential to impact classroom practice. The research questions are:

1. What vocabulary knowledge do learners need to cope with comprehending scripted texts in their L2?
2. How do L2 learners react to content words in scripted texts, specifically nouns and verbs?

3 Study Context

In Hong Kong, the context of the present study, English is an official language in business, management, law and tertiary education. However, the medium of instruction in most secondary schools is Chinese, resulting in a deficiency in the knowledge of academic words in tertiary students (Lin & Morrison, 2010). Listening is an important skill when students advance to post-secondary education in Hong Kong and elsewhere where English is the main language of instruction. According to the current secondary school curriculum, students should acquire the ability to "listen for information, ideas, intended meanings, views, attitudes and feelings in a variety of spoken texts" (Curriculum Development Council & Hong Kong Examinations and Assessment Authority, 2007, p. 23). Listening exercises in secondary school English classes that simulate real-world scenarios are encouraged; therefore, most spoken English in textbooks simulate radio broadcasts, podcasts, interviews, and informal discussions. Textbook passages and test papers are mostly scripted. However, listening tasks contribute to only about 5% of the total number of tasks in textbooks (Chan, 2013). Teachers often play recordings and have students finish textbook exercises with a focus on accuracy (Tsang, 2021). Additionally, the teaching of English at secondary schools tends to be examination oriented. Students sit for the Hong Kong Diploma of Secondary Education Examination (HKDSE, note 1) at the end of their secondary education. Many teachers rely on exam-focused textbooks, practice tests, and grammar drills to prepare students (Leung & Andrews, 2012). Thus, one motivating factor behind the current study is to see what linguistic knowledge senior secondary school students possess for coping with their textbooks and listening tests.

4 Methods

4.1 Participants

The participants were 73 students from a secondary school where Chinese was the medium of instruction. They were recruited through an invitation email and a follow up telephone call. One school agreed to provide their S5 students' (Year 11) test papers for data collection. The school, the students, and their parents signed a consent form that included details about the purpose and procedures of the study. The participants' completed the listening tasks as part of their end of term English language assessment. Their English proficiency was measured by a vocabulary test in class one week prior to the assessment.

4.2 Listening comprehension test

The instrument was a set of commercially published HKDSE practice papers consisting of four listening tasks. The situations of Tasks 1, 2 and 4 were conversations that simulated radio programs and a meeting, and Task 3 included monologues about two cities. Students wrote their answers in the form of bullet points and tables to simulate notetaking and report writing. The recording was played once lasting 37 minutes, including instructions and reading time. The speech was a standard British accent

at approximately 120 – 150 wpm. Among the 54 items, 20 items (37%) were gap-filling and sentence completion exercises; 20 items were form-filling exercises; and 13 items (24%) were comprehension questions. Around four in 10 test items were short answers with one to three words. While the necessary information for 14 items (25.9%) was mentioned once, the stimulus was often reiterated, including exact repetition (21 items, 38.9%), and paraphrasing or elaboration (19 items, 35.2%). Examples of test items are reported in the Appendix 1. The answers were then marked by the class teachers using dichotomous scoring (1 = correct, 0 = not correct).

4.3 Vocabulary level test

To measure the extent to which the participants knew the first 2,000 most frequently used words in English, the author designed three measures, namely, a receptive vocabulary (RV) test, a productive vocabulary (PV) test, and a C-test (see Appendix 2). Table 1 summarises the vocabulary levels of the three measures. Notably, some K2 words (the second 1,000 most frequent words of English) and AWL words (words in the *Academic Word List*) fell in the range of CEFR Level B2 and C1. Most of the words assessed in the C-test were at K1.

The RV-test had 30 test items divided into groups of six words to assess students' receptive knowledge, that is, their ability to recognize the meaning and form of a word. The matching of six words to three meanings reduced the chances of guessing correctly (Nation, 1990, as cited in Fan, 2001). The words were defined with 1000-level words. The PV- and C-tests measured productive vocabulary knowledge, that is, an ability to retrieve the form and meaning of a word from memory (Hirsh, 2015).

Table 1

Number of Vocabulary Test Items by Lexical Frequency

	VocabProfile			**CEFR**				Total
	K1	K2	AWL	A1/A2	B1	B2	C1/Off-list	
RV-test	13	10	7	7	17	6	0	30
PV-test	4	5	6	3	9	3	0	15
C-test	20	3	5	15	8	4	1	28
No. of items	37	18	18	25	34	13	1	73

The RV-test and PV-test were adapted from Schmitt et al.'s (2001) Receptive and Productive Vocabulary Level Test. The target words in the two tests were the same and were not meant to be administered at the same time. Some words in the two tests appeared randomly (see Stoeckel et al., 2020) (e.g. "ancient" and "holy"). To reflect vocabulary use in the current context, some words were replaced by words shortlisted from past papers of the HKDSE listening assessment as well as a vocabulary list from the participating school. Three short paragraphs from junior secondary coursebooks were selected for the C-test. Although the C-test is often used as a measure of general language proficiency (Daller & Wang, 2017), it was used as a measure of productive vocabulary in this study. The C-test engaged the students in cognitive processes of reading comprehension. Every second, third or fourth word had missing letters which required the students to have an understanding of words in more depth in relation to words in the surrounding text. The reliability of the instrument was evaluated by Cronbach's alpha. The three measures presented high correlation coefficients ranging between .73 and .79, indicating good internal consistency (α=.882) of the instrument. Principal components analysis suggested that the three sub-tests were unidimensional, measuring a single, psychometric construct.

4.4 Data analysis

The data were conceptualised along two dimensions: student ability (listening and vocabulary scores) and text characteristics (e.g. lexical coverage, lexical frequency). Data processing included transcription of the recordings and coding of lexical features. The study took a student-oriented approach to investigate the relationship between lexical threshold and listening comprehension. Two techniques were used.

Content analysis was used to reduce student output and listening passages into categories based on coding. The texts were systematically analysed for making replicable and valid inferences (Krippendorff, 2013). Quantification led to a more transparent analysis (Guest et al., 2012, p. 15). First, online text analysis tools generated variables such as vocabulary levels to measure listening difficulties. Second, gap-filling provided a record to investigate how well the input was understood by individual students (Ellis, 2003, p. 51). Students' responses to test items were closely read to identify similarities and differences in cognitive and linguistic abilities required for listening, and then gaps between what students could do and what they needed to do were inferred. It was expected that students would process language selectively focusing on nouns to grasp essential information (Brown, 2008), and they would process lexical items for meaning before processing grammatical form (VanPatten, 2015). Consequently, a smaller sample of items was shortlisted to study the processing of content words. The criteria were items with verbs and noun phrases with inflectional morphemes. Data analysis involved tabulating, counting, quantifying, and drawing inferences.

Cluster analysis was used to group students on the basis of their listening proficiency. SPSS TwoStep cluster analysis was run to classify students based on the sub-total scores of the four listening tasks. This statistical technique makes use of two commonly used cluster methods, namely, hierarchical cluster analysis and K-means cluster analysis, to classify groups in a data set (Crowther et al., 2020). The students in the same cluster displayed pattern of strengths and weaknesses in listening comprehension that were very different from those in different clusters. The data analysis sought to uncover how the students in the different clusters processed key words in the scripted texts.

5 Results

5.1 Vocabulary knowledge required for L2 listening comprehension

Table 2 shows the students' performance together with a summary of information about the four tasks presenting a matrix revealing the relationship between students' ability and text difficulty. Text difficulty was measured by text analysis tools (see note 2). Lexical coverage refers to the percentage of 2,000 high frequency words measured by *VocabProfile*. The four listening passages differed in their degree of familiarity. Sorting the tasks in descending order of lexical coverage helped identify negative cases. Task 4 had the highest coverage (95.2%), and Task 1 was the easiest.

Cluster analysis provided a three-cluster solution. Students in Clusters 1, 2 and 3 represented low, mid and high ability in listening comprehension. When the mean scores were highlighted using light grey (20-30%), grey (40-55%) and dark grey (>65%), a pattern across rows emerges. For Tasks 1, 2 and 4, students in Cluster 3 likely knew 90% of words in the texts, which were graded CEFR Level B1. There was a dramatic decrease in their score in Task 3 (45.1%) when compared with the other three tasks (>65%). Similarly, students in Cluster 2 scored 40-55% on average in three of the tasks; however, they experienced a dramatic drop to 21.2% in Task 3. Clearly, students did not understand the input at low percentages of lexical coverage.

Table 2

Relationship between Lexical Coverage and Listening Ability

	No. of items	Text difficulty		Listening score (%)		
		CEFR	Lexical coverage	Cluster 1 (n=28)	Cluster 2 (n=28)	Cluster 3 (n=17)
Task 4	13	B1	95.2%	19.2(13.0)	47.5(9.83)	70.5(16.3)
Task 2	12	B1	94.8%	29.1(10.7)	41.6(7.17)	68.6(8.60)
Task 1	14	B1	90.3%	24.7(13.6)	55.1(13.3)	82.3(9.81)
Task 3	15	B2	88.4%	7.1(6.52)	21.2(10.4)	45.1(11.4)
Total	54			19.1(7.78)	40.1(5.82)	65.4(8.4)

Figures 1 and 2 present the students' vocabulary knowledge as a factor contributing to their listening performance. The boxplots show the 25th, 50th (median), 75th percentiles, the minimum and maximum values. Students in Cluster 1 attained a vocabulary score far below the 85% threshold level, i.e., receptive vocabulary (mean 64.4, SD 21.6) and productive vocabulary (mean 36.3, SD 13.1). Students in Cluster 2 attained the 85% threshold in receptive vocabulary (mean 85.9, SD 11.6) but not productive vocabulary (mean 56.1, SD 12.6). Students in Cluster 3 scored higher in receptive vocabulary (mean 95.9, SD 4.05) and productive vocabulary (mean 74.0, SD 7.14). Notably, scoring 85% in receptive vocabulary was not sufficient for complete understanding despite the listening support. A minimum score of 60% productive vocabulary appeared necessary to pass the test. Therefore, lexical threshold helps to explain the variation. It appears that when students reached a greater understanding of the running words, they went up one CEFR level.

Figure 1

Receptive Vocabulary Knowledge by Cluster

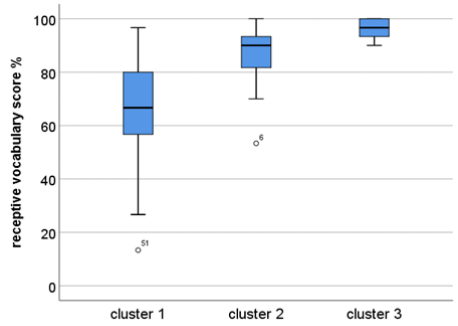
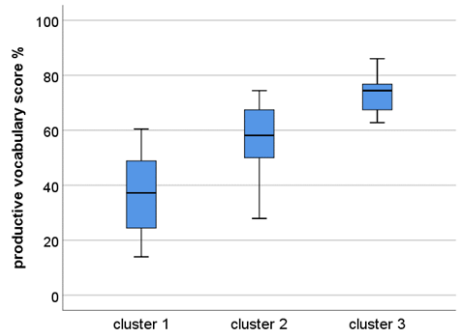


Figure 2

Productive Vocabulary Knowledge by Cluster



Tables 3 and 4 show the mean scores of students' vocabulary knowledge fell below 85% in the CEFR Levels ranging from A1 to B2. A Pearson correlation coefficient was computed to assess the linear relationship between vocabulary knowledge and listening comprehension. A positive and significant correlation between receptive and productive vocabulary knowledge and the listening test score was found (p<.001). The correlations between receptive vocabulary knowledge increased progressively at CEFR Level A1/A2 (r=.55), Level B1 (r=.61) and Level B2 (r=.63). The relationship with productive vocabulary knowledge at CEFR Level B1 (r=.79) and B2 (r=.75) was stronger than that at A1 (r=.59) and A2 (r=.54). In other words, a student who achieved sufficient vocabulary knowledge at CEFR Level B1 or B2 performed better in the listening test. Multiple regression analysis revealed that vocabulary knowledge using the PV-test (β=.375, p=.001) and C-test (β=.500, p<.001) contributed significantly to the model, explaining 66.6% of the variance in the listening score, adjusted R^2=.666, F(2,70)=72.9, p<.001. Thus, vocabulary knowledge had a strong positive relationship with listening comprehension.

Table 3

Descriptive Statistics for the Receptive Vocabulary Test, and Correlation with Listening Comprehension

Level	N	Mean(%)	Standard deviation	Correlation
A1&A2	7	83.6	17.4	.556*
B1	17	81.6	21.6	.614*
B2	6	73.7	26.3	.636*
Overall	30	80.5	19.8	.663*

Table 4

Descriptive Statistics for the Productive Vocabulary Test, and Correlation with Listening Comprehension

Level	N	Mean(%)	Standard deviation	Correlation
A1	11	61.6	22.2	.596*
A2	7	56.8	21.7	.545*
B1	17	52.9	22.0	.797*
B2	7	39.1	25.6	.750*
Overall	43	53.2	18.8	.821*

Table 5

Regression Analysis for Vocabulary Knowledge Predicting Listening Comprehension

Variable	R^2	adj R^2	B	Standardized beta	t	*p*
Model 1	.683	.669				
RV-test			.123	.125	1.247	.217
PV-test			.295	.326	2.924	.005
C-test			.456	.450	4.034	.000
constant			-11.161			
Model 2	.676	.666				
PV-test			.339	.375	3.581	.001
C-test			.506	.500	4.774	.000
constant			-6.277			

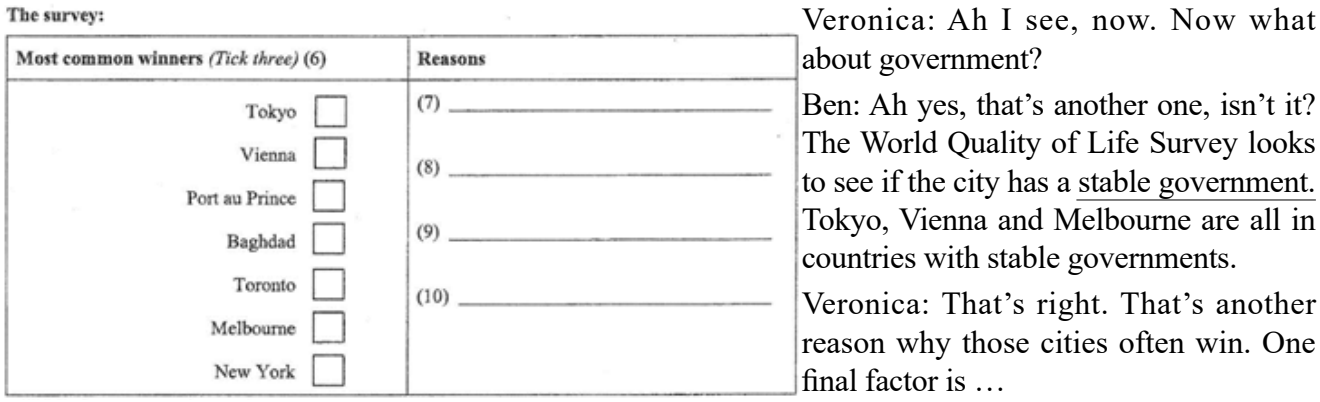
Overall, the results suggest that students needed to know 2,000 high frequency words in English to cope with the test. Students in Cluster 3 who reached or approached the lexical threshold passed the test.

5.2 How L2 learners react to content words in scripted texts

The study further investigated what the students could achieve with their existing vocabulary knowledge. To explore this issue, one representative example appears in Figure 3 showing an item whose correct answer was "stable government" (p=.58). The rates of attaining the correct answer in Clusters 1, 2 and 3 were: 28.6%, 67.9% and 94.1% respectively. Student errors indicated that the test item operated like a dictation requiring skills in spelling, word recognition, and semantic processing. Errors included: "good goverment" (Cluster 1), "statable government" (Cluster 2), and "steady governments" (Cluster 3). It is possible that the use of rhetorical questions together with repetition of the necessary information made the listening task less demanding, which gave students time to guess at the spelling of unfamiliar words.

Figure 3

Listening Support



To investigate the extent to which students recognised key words, the use of nouns and verbs were coded. Table 6 shows students' processing of nouns. A gradual change was observed in the accuracy rate of retrieving target words (TW) by students in Cluster 1, 2 and 3 respectively: 26.3%, 53.3% and 79.8%. When mistakes in plural 's' (WF) were ignored, the rate of accuracy increased to 45.1%, 74.1% and 87.5% respectively. Thus, differences between all clusters were evident.

The variability in student performance may be explained by target-word familiarity and automaticity of word form. Notably, easier items were associated with familiar words (e.g. "air pollution") and exact repetition (e.g. "offices"). The percentage of correct responses often fell below 30% (p < .30) in the case of unfamiliar words (e.g. "divided", "cousin", "variety", "rules") and prepositional phrases (e.g. "at 2 pm", "per week", "around the world").

Erroneous responses explain why lower proficiency students had difficulty with processing information. Students in Cluster 1 lacked sufficient vocabulary knowledge to retrieve and produce the words heard. Explicitness brought by exact repetition might have enhanced listening performance. Responses such as "compen", "international compary" and "internet companies" (item 4) revealed an ability to segment what was heard with a certain degree of phonological awareness. Although some students in Cluster 1 were able to identify relevant words in Task 3, their answers were incomplete, such as "two part", "devited 2 parts" (item 33). Erroneous responses suggest that insufficient vocabulary knowledge impeded the students' notetaking.

Table 6

Frequency of Nouns Decoded by Cluster

		Cluster 1 (n=28)			Cluster 2 (n=28)			Cluster 3 (n=17)		
Items	*p*	TW	WF	M	TW	WF	M	TW	WF	M
CEFR A1:										
18: 10 years [NR]	.82	22	2	4	26	1	1	17	0	0
3: 205 cities [NR]	.34	9	4	15	13	2	13	15	0	2
28*: three new areas [NR]	.24	1	5	22	11	3	14	8	2	7
12: variety of restaurants [ER]	.19	0	9	19	3	18	7	10	6	1
41*: five famous museums [R]	.16	3	9	16	7	14	7	6	7	4
33*: divided into two parts [NR]	.27	9	9	10	15	6	7	15	1	1
51: Her cousin called her at 2 am [R]	.24	14	0	14	18	0	10	12	0	5
CEFR A2:										
4: international companies [ER]	.61	8	9	11	22	4	2	17	0	0
22: open to new ideas [ER]	.49	5	13	10	16	9	3	17	0	0
39*: 105 paintings [R] around the world	.06	2	8	18	9	14	5	12	4	1
5: set up new offices [ER]	.72	13	3	12	27	0	1	15	0	2
42: increase/bring … tourists [R]	.38	8	6	14	18	5	5	17	0	0
20: attend two meetings [NR] per week	.26	1	4	23	9	10	9	17	0	0
CEFR B1:										
50: reduce air pollution [ER]	.65	16	0	12	25	1	2	17	0	0
49: improve recycling programme [ER]	.35	6	1	21	17	3	8	17	0	0
36*: changed its rules [R]	.09	1	2	25	3	3	22	5	1	11
Overall noun (%)		**26.3**	18.8	54.9	**53.3**	20.8	25.9	**79.8**	7.7	12.5

Note:

1. TW = target word (underlined), WF = target word with incorrect form; M = missing the target word, ER = exact repetition, R = lexical repetition, NR = no reiteration, * Task 3
2. The items are listed in descending order by (1) lexical frequency, (1) noun and verb phrases, and (3) item difficulty, which is the value of *p*, as a proportion of correct items (see note 3).

Table 7

Frequency of Verbs Decoded by Cluster

		Cluster 1 (n=28)			Cluster 2 (n=28)			Cluster 3 (n=17)		
	p	TW	WF	M	TW	WF	M	TW	WF	M
CEFR A1/A2:										
52: walked[R] to MTR station alone	.35	4	5	19	14	8	6	12	1	4
49: improve[NR] recycling programme	.35	2	2	24	13	1	14	13	0	4
51: Her cousin called[R] her at 2 am	.24	4	6	18	9	8	11	9	1	7
36*: changed[R] its rules	.09	2	13	13	2	16	10	5	8	4
CEFR B1:										
50: reduce[NR] air pollution	.65	9	2	17	24	1	3	17	0	0
42: increase/bring[ELA] … tourists	.38	20	0	8	27	0	1	17	0	0
33*: divided[R] into two parts	.27	1	0	27	9	2	17	14	1	2
20: attend[NR] two meetings per week	.26	2	2	24	13	2	13	14	0	3
Overall verb (%)		**19.6**	13.4	67.0	**49.6**	17.0	33.5	**74.3**	8.1	17.6

Note: * Task 3; TW = target word (underlined), WF = target word with incorrect form; M = missing the target word, NR = no reiteration, R = lexical repetition, ELA = elaboration

Table 7, which summarises how students reacted to verbs, shows the accuracy rate of decoding verbs was lower than that of nouns. Among the eight items, seven had a *p*-value below .40 indicating the items were difficult. Similarly, there were differences in the accuracy rate in decoding verbs by students in Cluster 1 (19.6%), Cluster 2 (49.6%) and Cluster 3 (74.3%). When the past 'ed' was not considered, the accuracy rate increased to 33.0%, 66.6% and 82.4% respectively.

Students in Cluster 1 (67%) and Cluster 2 (33%) often failed to identify the verb. Elaboration appeared useful when words were familiar. For example, they were able to substitute "bring" for a CEFR B2 word "increase". Students in both Clusters 1 and 2 had difficulty attending to inflectional morphemes that did not contribute substantially to the meaning of the passage. Students in Cluster 3 often showed an ability to encode grammatical relationships among words. Many of them were able to identify lower frequency vocabulary (e.g. "divided" and "attend" at CEFR Level B1), and attention to the past 'ed' (e.g. "called", "walked") was automatised.

As verbs are more morphologically complex, students' level of grammatical knowledge might have had an impact on how quickly and effectively they could parse syntactic patterns. For example, item 36 asked what happened in 1989: "The government in East Germany _____". Students heard:

> By 1989 there was a change of rules in East Germany. It allowed people to cross into West Germany. As soon as the government changed its rules, people began climbing over and destroying the wall.

This text is dense, demanding the listener to interpret the speech signal quickly to understand the input. Some students were able to identify specific details, for example, "allow to cross the wall" (Cluster 1), "allow people to cross to West Germany" (Cluster 2) and "allow people crossing the wall" (Cluster 3). However, the variation of the syntactic structure demanded a higher sensitivity to grammatical details.

Overall, students in Cluster 3 who scored 95% receptive vocabulary and 74% productive vocabulary were able to decode key words fluently. Word familiarity probably helped the students become more sensitive to grammatical details in the clues. Accuracy in producing 's' and 'ed' endings decreased when they were drawn to identify words in more complex texts (e.g. * items). One possible reason was that students first processed content words for information before they processed grammatical form (VanPatten, 2015). The results also appear to contradict the view of selective listening where listeners processed nouns more than verbs to cope with spoken input (Brown, 2008). Familiar words appeared more salient for the students.

6 Discussion

The study investigated the relationship between lexical threshold and listening comprehension, especially at the threshold 2,000-word level. The four listening tasks required the students (n = 73) to produce key words for notetaking and answering comprehension questions. The results reveal that many parts of the listening input are not successfully decoded until a listener reaches a certain threshold. The matrix of lexical coverage and content words captured the dynamic processing of text comprehension and word recognition at a given period. They functioned like implicational scales revealing what linguistic elements students take in from the spoken input and what elements are most difficult in processing (see Rost, 2011, p. 260). Such results then provide information to generate listener profiles for needs assessment.

Cluster analysis showed the three student groups differed in their ability to comprehend. Students in Cluster 3 displayed good listening fluency, that is, an ability to process input automatically while reaching a satisfactory degree of performance (Chang, Millett, & Renandya, 2019). They had greater success processing noun phrases and clauses as well as comprehending complex texts beyond their current vocabulary level. Recognising content words probably facilitated a chunking of information, which may have helped the students remember more words. Their decoding of verbs may be attributed to their higher cognitive skills because it involved an automaticity of grammatical knowledge. Their receptive knowledge of plural and tense markings likely turned into productive knowledge through a period of extensive input processing. The students' lexical awareness may be explained by a possible accuracy threshold at CEFR B1 (see Thewissen, 2013). The findings indicate that attaining the lexical threshold increases possibilities for processing words in a text.

The students in Cluster 2, who appeared to understand at the CEFR Level B1, were often able to identify relevant ideas and details when listening support was available. An ability to recognise two-thirds of the key words distinguished them from Cluster 1 students. However, they lacked sensitivity to plural and tense markings even after many years of exposure to English. One possible reason is that L2 learners in the early and intermediate stages of acquisition do not attend to form in the input (VanPatten, 2015).

The students in Cluster 1 showed low achievement in English learning. Results indicate that their vocabulary knowledge was insufficient for decoding even half of the key words. They likely could not connect different pieces of information, nor could they accurately interpret the meaning of the incoming stimulus. Although they occasionally segmented the sound stream into meaningful units and selected potential word candidates, they lacked spelling skills. They were at risk because the input would not result in intake. They processed the text at a superficial level where little new information would be encoded to the lemma, leading to a negative impact on their learning.

Theoretically, this study showed that lexical threshold may mark a critical point of automatic decoding incoming speech signals. The findings show that when students had reached the threshold, their ability to extract and organise information from a text improved. Grammatical sensitivity increased as proficiency approached the lexical threshold. Vocabulary growth and grammatical development are dynamic systems (Larsen-Freeman, 2009). In Schmitt, Jiang and Grabe (2011), results did not indicate a vocabulary knowledge threshold beyond which reading comprehension increased dramatically. Thus, the use of multiple methods – cluster analysis, content analysis, graded texts, and a vocabulary test – may have helped discover a type of threshold that is not easily detected. Regression analysis suggested that vocabulary knowledge explained over half of the variance observed within the L2 listening scores as in previous studies (Matthews, 2018; Zhang & Zhang, 2020). The analysis of student responses to the test items further provides compelling evidence in favour of L2 listening being more an issue of language than listening (Vandergrift, 2006). While there are many relevant factors contributing to listening comprehension, the study sheds light on the lexical threshold, which can explain significant differences in listening performance.

6.1 Pedagogical implications

The findings provide empirical support to the notion of a lexical threshold for L2 listening comprehension. Students need to know 2,000 – 3,000 high frequency words, or CEFR Level B1 vocabulary, to cope with listening to English in standardised tests like the HKDSE examination for university admission. Thus, lexical threshold deserves more pedagogical attention when designing syllabi, materials, and tasks. Education authorities should provide teachers with specific guidelines or professional development workshops concerning lexical thresholds to advance students' L2 listening ability. Since listening syllabus design is often skill-based defining the competence students should have, the study may imply a shift is needed in the practice of listening instruction from focusing on skills to linguistic competence. Increasing teachers' and learners' awareness of the importance of meaningful exposure to texts and tasks with the 2,000 – 3,000 most common words in English is required. The findings have important implications for instructional design and materials development focusing on: (1) words that should be taught; (2) students' vocabulary level; (3) the use of repetition to help listeners make sense of the message; and (4) information in authentic contexts that may be given only once (Richards, 2013). Considering the characteristics of the listener profiles, different kinds of listening input and practices catering to learners of diverse proficiency may be required.

The results shed light on the relationship between the listener and the text. The findings suggested that students' processing capacity constrained what they extracted from input. Attaining a certain lexical threshold may be a prerequisite for a learner to benefit from the spoken input of coursebooks and practice tests. Although Stoeckel et al. (2020) suggested specifying texts at four levels – 100%, 95-99%, 85-95% and below 85% coverage – the present study revealed 85% - 95% coverage may be further divided into two difficulty levels. A minimum coverage of 95% is necessary for students who have not reached the lexical threshold. A text with a larger coverage of familiar words would likely improve automatised lexical access, empowering the listener to recognise a sufficient percentage of the words in a text for processing at higher levels.

Texts scripted in principled ways can enhance comprehensibility and help students develop listening fluency. Based on the findings then, five guiding questions for preparing listening texts are proposed:

1. Content words: Does the text serve the purpose of building essential vocabulary?
2. Lexical frequency: Is the text written at a frequency level that matches the vocabulary knowledge of the learners?
3. Explicitness: Are low proficiency students provided sufficient listening support?

4. Function words: Are there sufficient instructions on grammatical structures in context?
5. Word forms: Are there opportunities to listen to how plurals, participles and other inflectional features are spoken?

The purpose of these questions is to provide students comprehensible input for vocabulary acquisition through listening. When the learner has control over processing for meaning, processing space is freed up. According to VanPatten's (2015) model of input processing, L2 learners process content words before anything else in the input. Research has suggested that the retention of target words increases after three to five repetitions (Vidal, 2011). Listening to improve grammar needs more rehearsal than for improving content word knowledge (Schütze, 2016). Repetition can improve automaticity of lexical access within contextualised and communicative activities (Segalowitz & Hulstijn, 2005). A text that is comprehensible may provide linguistic data to be automatically fed into the learners' internal system leading to language acquisition.

Apart from preparing suitable texts, teachers should focus their practice to improve students' processing capacity. An analogy of focused practice is scale practice for musicians, which aims to improve overall skill and fluency. Carefully designed focused practice can scaffold the processing of unfamiliar words in spoken English. Word recognition practice may help students activate their receptive vocabulary knowledge. Gap-filling of increasing difficulty may be designed, ranging from recognising keywords to short answers where students are asked to paraphrase key points (see Buck, 2001; Cai, 2012). Focused practice may be given a few days before or after performing listening tasks, applying a spacing technique to engage students in recalling the newly learned words. As word familiarity increases, the text is likely more accessible to the listener. This means mock exam practice on questioning formats and test-taking strategies may not be as effective in boosting students' listening scores as previously believed. Intensive instruction on vocabulary related to listening may be a more useful instructional tool.

6.2 Limitations and concluding remarks

This study used multiple analytic techniques to uncover patterns of listening performance while creating three profiles to describe challenges facing L2 listeners. The study has certain limitations. The limited data from the naturalistic inquiry using summative assessment at a local school may have produced unreliable results. In addition, few test items were available to measure the listening ability of low proficiency students. Further, the use of a written vocabulary test may have overestimated the students' vocabulary size in listening because it did not measure students' ability to recognise the phonological form of the target words. Accordingly, researchers may consider intentionally designing a vocabulary test with a higher percentage of related items in a listening test (Schmitt et al., 2011). Apart from lexical coverage, spoken texts may also be measured and graded by temporal variables such as articulation rate and pause time (Préfontaine & Kormos, 2015). To conclude, the study findings suggest that a vocabulary threshold exists for L2 listening comprehension. Attaining the lexical threshold may transform students' communicative language use, which in turn offers an opportunity to reflect on the design of L2 listening curriculums to ensure the achievement of learning outcomes.

Note

1. The Hong Kong Diploma of Secondary Education Examination (HKDSE) sets the English language standard for secondary education or the workplace in Hong Kong. In the HKDSE, the performance of candidates is reported with criterion-referenced standards in five levels (level 1 to 5), with 5 being the highest. HKDSE Level 3 is a minimum requirement for admission to university while Level 2

indicates eligibility to work for the civil service in Hong Kong. HKDSE Levels 2 and 3 are equivalent to IELTS band 4.79 – 5.07 and IELTS band 5.48 – 5.68 scores, respectively (HKEAA, 2013). Candidates who do not meet the standard required for HKDSE Level 1 are graded 'unclassified'.

2. Lexical frequency and lexical coverage were computed by three text analysis tools: *VocabProfile* https://www.lextutor.ca/vp/eng/, *Text Inspector* https://languageresearch.cambridge.org/wordlists/text-inspector, and *The Text Analyzer* https://cefrlevels.com/textanalysis/index.html.
3. Item difficulty is measured by a proportion of correct items. The value of *p* ranges from 0 to 1. An item is difficult if its value falls below .30, meaning less than 30% of students get a test item correct. Items with a value below .10 do not discriminate student ability.

References

Aizawa, I., Rose, H., Thompson, G., & Curle, S. (2020). Beyond the threshold: Exploring English language proficiency, linguistic challenges, and academic language skills of Japanese students in an English medium instruction programme. *Language Teaching Research*, https://doi.org/10.1177/1362168820965510

Andringa, S., Olsthoorn, N., van Beuningen, C., Schoonen, R., & Hulstijn, J. (2012). Determinants of success in native and non-native listening comprehension: An individual differences approach. *Language Learning*, *62*(s2), 49-78.

Antes, T.A. (2014). Audio glossing during information-gap activities: The effect on learner output. *System*, *45*, 163-174.

Ardasheva, Y., Tretter, T. R., & Kinny, M. (2012). English language learners and academic achievement: Revisiting the threshold hypothesis. *Language Learning*, *62*(3), 769-812.

Aryadoust, V. (2019). An integrated cognitive theory of comprehension. *International Journal of Listening*, *33*(2), 71-100.

Brown, G. (2008). Selective listening. *System*, *36*(1), 10-21.

Buck, G. (2001). *Assessing listening*. Cambridge University Press.

Cai, H. (2012). Partial dictation as a measure of EFL listening proficiency: Evidence from confirmatory factor analysis. *Language Testing*, *30*(2), 177-199.

Carson, L. (2017). Second language use as a threshold concept: Reconceptualising language learning journeys. *International Journal of Education*, *9*(2), 34-43.

Chan, J. Y. H. (2013). The role of situational authenticity in English language textbooks. *RELC Journal*, *44*(3), 303-317.

Chang, A. C. (2007). The impact of vocabulary preparation on L2 listening comprehension, confidence and strategy use. *System*, *35*(4), 534-550.

Chang, A. C., & Millett, S. (2013). The effect of extensive listening on developing L2 listening fluency: Some hard evidence. *ELT Journal*, *68*(1), 31-40.

Chang, A., Millett, S., & Renandya, W. A. (2019). Developing listening fluency through supported extensive listening practice. *RELC Journal*, *50*(3), 422-438.

Crowther, D., Kim, S., Lee, J., Lim, J., & Loewen, S. (2021). Methodological synthesis of cluster analysis in second language research. *Language Learning*, *71*(1), 99-130.

Curriculum Development Council & the Hong Kong Examinations and Assessment Authority (2007). *CDC-HKEAA English Language Curriculum and Assessment Guide (Secondary 4 – 6)*. https://www.hkeaa.edu.hk/doclibrary/sba/hkdse/eng_dvd/doc/hkeaa_eng_lang_final.pdf

Daller, M., & Wang, Y. (2017). Predicting study success of international students. *Applied Linguistics Review*, *8* (4), 355374.

Dang, T. N. Y., & Webb, S. (2014). The lexical profile of academic spoken English. *English for Specific Purposes*, *33*, 66-76.

Dupuy, B. C. (1999). Narrow listening: An alternative way to develop and enhance listening comprehension in students of French as a foreign language. *System*, *27*(3), 351-361.

Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.

Fan, M. Y. (2001). An investigation into the vocabulary needs of university students in Hong Kong. *Asian Journal of English Language Teaching*, *11*, 69-85.

Guest, G., McQueen, K. M., & Namey, E. E. (2012). *Applied thematic analysis*. SAGE.

Harrington, M., & Roche, T. (2014). Identifying academically at-risk students in an English-as-a-Lingua-Franca university setting. *Journal of English for Academic Purposes*, *15*, 37-47.

Harwood, N. (2002). Taking a lexical approach to teaching: Principles and problems. *International Journal of Applied Linguistics*, *12*(2), 139-155.

Hawkins, J. A., & Filipović, L. (2012). *Critical features in L2 English: Specifying the reference levels of the Common European Framework* (Vol. 1). Cambridge University Press.

Hayati, A., & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of EFL learners. *British Journal of Educational Technology*, *42*(1), 181-192.

Hirsh, D. (2015). Researching vocabulary. In B. Palfridge & A. Phakiti (Eds.), *Research methods in Applied Linguistics: A practical resource*. Bloombury.

Hong Kong Examinations and Assessment Authority (2013). Benchmarking studies on international examinations. https://www.hkeaa.edu.hk/en/recognition/benchmarking/hkdse/ielts/

Hubackova, S. (2011). Listening texts in the teaching of German language. *Procedia – Social and Behavioural Sciences*, *28*, 34-38.

Hulstijn, J. H. (2001). Intentional and incidental second-language vocabulary learning: A reappraisal of elaboration, rehearsal and automaticity. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 258-286). Cambridge University Press.

Krippendorff, K. (2013). *Content analysis: An introduction to its methodology* (3rd ed.). Sage.

Kuo, Y. (2010). Using partial dictation of an English teaching radio program to enhance EFL learners' listening comprehension. *Asian EFL Journal*, *47*, 4-29.

Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy, and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 579-589.

Laufer, B., & Goldstein, Z. (2004). Testing vocabulary knowledge: Size, strength, and computer adaptiveness. *Language Learning*, *54*(3), 399-436.

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

Leung, C.Y., & Andrews, S. J. (2012). The mediating role of textbooks in high-stakes assessment reform. *ELT Journal*, *66*(3), 356-365.

Lin, L. H., & Morrison, B. (2010). The impact of the medium of instruction in Hong Kong secondary schools on tertiary students' vocabulary. *Journal of English for Academic Purposes*, *9*(4), 255-266.

Long, M. (2015). *Second language acquisition and task-based language learning*. Wiley-Blackwell.

Lynch, T. (1998). Theoretical perspectives on listening. *Annual Review of Applied Linguistics*, *18*, 3-19.

Lynch, T. (2009). *Teaching second language listening*. Oxford University Press.

Matthews, J. (2018). Vocabulary for listening: Emerging evidence for high and mid-frequency

vocabulary knowledge. *System*, *72*, 23-36.

McGrath, I (2002). *Materials evaluation and design for language teaching*. Edinburgh University Press.

McKinley, J., Rose, H., & Zhou, S. (2021). Transnational universities and English Medium Instruction in China: How admissions, language support and language use differ in Chinese universities. *RELC Journal*, *52*(2), 236-252.

Meyer, J., & Land, R. (2006). *Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge*. Routledge.

Milliner, B., & Dimoski, B. (2021). The effects of a metacognitive intervention on lower-proficiency EFL learners' listening comprehension and listening self-efficacy. *Language Teaching Research*, https://doi.org/10.1177/136216882110046

Nation, I. S. P. (1990). *Teaching and learning vocabulary*. Heinle and Heinle.

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

Oh, E. (2016). Comparative studies on the roles of linguistic knowledge and sentence processing speed in L2 listening and reading comprehension in an EFL tertiary setting. *Reading Psychology*, *37*(2), 257-285.

Pan, Y. C., Tsai, T. H., Huang, Y. K., & Liu, D. (2018). Effects of expanded vocabulary support on L2 listening comprehension. *Language Teaching Research*, *22*(2), 189-207.

Préfontaine, Y., & Kormos, J. (2015). The relationship between task difficulty and second language fluency in French: A mixed methods approach. *The Modern Language Journal*, *99*(1), 96-112.

Prince, P. (2013). Listening, remembering, writing: Exploring the dictogloss task. *Language Teaching Research*, *17*(4), 486-500.

Richards, J. C. (2001). *Curriculum development in language teaching*. Cambridge: Cambridge University Press.

Richards, J. C. (2005). Second thoughts on teaching listening. *RELC Journal*, *36*(1), 85-92.

Richards, J. C. (2013). Curriculum approaches in language teaching: Forward, central, and backward design. *RELC Journal*, *44*(1), 5-33.

Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Pearson Education.

Rost, M. (2014). Developing listening fluency in Asian EFL settings. In T. Muller, J. Adamson, P. S. Brown & S. Herder (Eds.), *Exploring EFL fluency in Asia* (pp. 281-296). Palgrave Macmillan.

Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, *78*(2), 199-221.

Schmitt, N., Jiang, X., & Grabe, W. (2011). The percentage of words known in a text and reading comprehension. *The Modern Language Journal*, *95*(1), 26-43.

Schmitt, N., & Schmitt, D. (2014). A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, *47*(4), 484-503.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language Testing*, *18*(1), 55-88.

Schütze, U. (2016). *Language learning and the brain: Lexical processing in second language acquisition*. Cambridge University Press.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. De Groot (Ed.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371-388). Oxford University Press.

Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English for a foreign Language. *Studies in Second Language Acquisition*, *31*(4), 577-607.

Stoeckel, T., McLean, S., & Nation, P. (2021). Limitations of size and levels tests of written receptive vocabulary knowledge. *Studies in Second Language Acquisition*, *43*(1), 181-203.

Teng, F. (2016). An in-depth investigation into the relationship between vocabulary knowledge and academic listening comprehension. *TESL-EJ*, *20*(2), 1-17.

Thewissen, J. (2013). Capturing L2 accuracy developmental patterns: Insights from an error-tagged EFL learner corpus. *The Modern Language Journal*, *97*(1), 77-101.

Tsang, A. (2021). The roads not taken: Greater emphasis needed on "sounds", "actual listening", and "spoken input". *RELC Journal*, https://doi.org/10.1177/0033688221104027

van Zeeland, H., & Schmitt, N. (2013). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics*, *34*(4), 457-479.

Vandergrift, L. (2006). Second language listening: Listening ability or language proficiency? *The Modern Language Journal*, *90*(1), 6-18.

Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, *65*(2), 390-416.

VanPatten, B. (2015). Input processing in adult SLA. In B. VanPatten & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (2nd ed.) (pp. 113-134). Routledge.

Vasiljevic, Z. (2010). Dictogloss as an interactive method of teaching listening comprehension to L2 learners. *English Language Teaching*, *3*(1), 41.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning*, *61*(1), 219-258.

Wagner, E. (2014). Using unscripted spoken texts in the teaching of second language listening. *TESOL Journal*, *5*(2), 288-311.

Wallace, M. P. (2020). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning*, *72*(1), 5-44.

Wang, Y., & Treffers-Daller, J. (2017). Explaining listening comprehension among L2 learners of English: The contribution of general language proficiency, vocabulary knowledge and metacognitive awareness. *System*, *65*, 139-150.

Wolfgramm, C., Suter, N., & Göksel, E. (2016). Examining the role of concentration, vocabulary and self-concept in listening and reading comprehension. *International Journal of Listening*, *30*(1-2), 25-46.

Zhang, S., & Zhang, X. (2022). The relationship between vocabulary knowledge and L2 reading/listening comprehension: A meta-analysis. *Language Teaching Research*, *26*(4), 696-725.

Sonia Cheung is a EdD graduate at The University of Hong Kong. She has conducted research in universities e.g. Wise Assessment Practices, learning Chinese as a second language, and the enhancement of generic learning outcomes. Her research interests include assessment for learning, needs analysis, and materials design and development.