

Exploring the relationships between L2 vocabulary knowledge, lexical segmentation, and L2 listening comprehension

Kriss Lange

University of Shimane Matsue Campus, Japan

<https://orcid.org/0000-0001-7201-1751>

k-lange@u-shimane.ac.jp

Joshua Matthews

School of Education, University of New England, Australia

<https://orcid.org/0000-0002-2260-2331>

joshua.matthews@une.edu.au

Abstract

The capacity to perceive and meaningfully process foreign or second language (L2) words from the aural modality is a fundamentally important aspect of successful L2 listening. Despite this, the relationships between L2 listening and learners' capacity to process aural input at the lexical level has received relatively little research focus. This study explores the relationships between measures of aural vocabulary, lexical segmentation and two measures of L2 listening comprehension (i.e., TOEIC & Eiken Pre-2) among a cohort of 130 tertiary level English as a foreign language (EFL) Japanese learners. Multiple regression modelling indicated that in combination, aural knowledge of vocabulary at the first 1,000-word level and lexical segmentation ability could predict 34% and 38% of total variance observed in TOEIC listening and Eiken Pre-2 listening scores respectively. The findings are used to provide some preliminary recommendations for building the capacity of EFL learners to process aural input at the lexical level.

Keywords: second language listening; aural vocabulary; lexical segmentation; listening comprehension

1. Introduction

For some time there has been a general acknowledgement of a robust relationship between foreign or second language (L2) vocabulary breadth and L2 listening (Stæhr, 2009). More recent examinations of this relationship have improved our understanding of its strength and specificity. Indeed, recent research examining the relative strength of the link between L2 listening and multiple variables of assumed importance, such as auditory discrimination, working memory, metacognitive awareness, L1 vocabulary knowledge and L2 vocabulary knowledge have presented L2 vocabulary knowledge as arguably the most important (Vandergrift & Baker, 2015; Wallace, 2020). Furthermore, there is a growing appreciation of the specific relationship between L2 listening and aural vocabulary knowledge. Recent research has demonstrated that aural vocabulary knowledge is more predictive of L2 listening than is word knowledge measured in the written form alone (Cheng & Matthews, 2018) and should therefore be utilized more in listening research.

Recognizing and knowing the meaning of individual words from speech is an essential foundation for listening, but so too is lexical segmentation. Here we define lexical segmentation as the ability to identify multiple consecutive words in connected speech (Andringa, Olsthoorn, van Beuningen, Schoonen, & Hulstijn, 2012; Field, 2003). Although it is dependent upon adequate levels of single word knowledge, it is arguably just as important. This is because lexical segmentation entails accurately recognizing the boundaries between single words and the resultant capacity to map recognized words onto existing representations in the listener's mental lexicon, known as decoding (Field, 2008a). Lexical segmentation is especially challenging for L2 learners as authentic spoken language is typically not produced as discrete phonological word forms, but mostly as streams of connected, phonologically modified lexis. Words within fluent speech become co-articulated, with adjacent phonemes influencing each word's phonological form (Field, 2008a). Additionally, the speech signal is transient making it necessary for the listener to segment words rapidly, with an average rate of native speech reaching over six syllables per second (Pellegrino, Coupé, & Marsico, 2011). Lexical segmentation is a complex skill, "which requires a context-sensitive representation of phonemes and phoneme clusters both within and across word boundaries" (Hulstijn, 2003, p. 420). Considering these challenges, it is unsurprising that lexical segmentation of connected speech causes considerable difficulty for L2 listeners (Field, 2008b; Lange, 2018).

It is assumed here that aural vocabulary knowledge and lexical segmentation ability are both important in supporting successful L2 listening. However, the relationship between these constructs and multiple measures of L2 listening performance has not thus far been adequately explored. The study reported in

this paper seeks to begin filling this gap in the literature by examining these relationships among a group of tertiary level Japanese EFL learners.

2. Literature review

2.1. L2 vocabulary knowledge and L2 listening

Measures from written, receptive vocabulary tests have been shown to possess a relatively strong and consistent relationship with measures of L2 listening comprehension across a range of learning contexts. For example, Stæhr (2009) investigated the strength of association between advanced Danish EFL students' L2 listening and vocabulary size (Vocabulary Levels Test; Schmitt, Schmitt, & Clapham, 2001) and vocabulary depth (Word Associates Test; Read 1993, 1998), and determined that these correlated strongly and significantly ($r = .70$ and $r = .65$, respectively). The generalizability of the strength of association between receptive L2 vocabulary size and L2 listening comprehension was further demonstrated by Andringa et al. (2012). While investigating the determinants of L2 listening comprehension among 113 non-native Dutch speakers with 35 different first language groups, they found that scores from a 60-item receptive L2 vocabulary test correlated strongly with L2 listening comprehension ($r = .69$). The depth and size of receptive L2 vocabulary knowledge, measured orthographically in various contexts, appears to have a moderate to strong relationship with L2 listening comprehension.

2.2. L2 aural vocabulary knowledge and L2 listening

The robust relationship between L2 listening comprehension and receptive L2 vocabulary knowledge, as measured with written receptive vocabulary tests, is relatively well established. However, researchers engaged in previous related studies have tended not to use measures of aural vocabulary knowledge (Stæhr, 2009). This is likely because most vocabulary tests have been solely delivered through the medium of writing (Milton, 2013). This tendency is a significant limitation (Stæhr, 2009; Vandergrift & Baker, 2015) as scores from L2 aural vocabulary tests are more strongly associated with L2 listening comprehension than equivalent written measures of receptive L2 vocabulary knowledge. In a study undertaken within the Chinese tertiary EFL context among 250 participants, Cheng and Matthews (2018) demonstrated that scores from vocabulary tests that required test-takers to process aural stimulus were more strongly correlated with listening ($r = .71$) than scores from comparable written vocabulary tests ($r = .55$).

Other research that has explored links between L2 listening comprehension and L2 aural vocabulary knowledge has also demonstrated a strong link between

these constructs. For example, Vandergrift and Baker (2015) investigated the learner variables that predicted L2 listening comprehension among 157 learners of French. They tapped into a number of factors including receptive aural L2 (French) and L1 (English) vocabulary knowledge, L1 and L2 listening ability, auditory discrimination ability, working memory and metacognition. L2 vocabulary knowledge proved to be the strongest correlate of L2 listening. The mean magnitude of correlation between L2 listening comprehension and L2 vocabulary knowledge ($r = .51$) across three cohorts of learners was more than double that of all other variables that reached a statistically significant level (L1 vocabulary, $r = .23$; metacognition, $r = .23$; and auditory discrimination, $r = .22$). Matthews and Cheng (2015) demonstrated that partial dictation test scores measuring knowledge of high-frequency words correlated strongly with IELTS listening test scores among a cohort of 167 tertiary level Chinese EFL learners ($r = .73$). McLean, Kramer, and Beglar (2015) demonstrated that their Listening Vocabulary Levels Test, which requires test-takers to process aural stimulus material, correlated strongly ($r = .54$) with parts one and two of the listening component of the Test of English for International Communication (TOEIC). Finally, in the Japanese EFL context, Wallace (2020) examined the relationship between various factors, such as aural L2 vocabulary knowledge, metacognitive awareness, memory, attentional control, self-reported topical knowledge, and L2 listening. Results of structural equation modelling analysis indicated that vocabulary knowledge accounted for the most variability in L2 listening performance. These studies have helped to demonstrate the significant relationship that aural receptive L2 vocabulary knowledge has with L2 listening comprehension ability.

2.3. Lexical segmentation and L2 listening

The research reviewed above demonstrates that there is a relatively strong relationship between aural vocabulary knowledge and L2 listening comprehension across a range of contexts. However, a limitation of previous studies is that they have only measured individual words and not the capacity to segment multiple words in connected speech. This gap is important to address as spoken language is almost always delivered in concatenated intonation units (Rost, 2002). Connected words are often acoustically very different from their discrete citation form due to phonological modification (e.g., reduction, assimilation, elision, etc.). For this reason, being able to accurately segment strings of connected lexis is an important objective for L2 listeners, and is indicative of high levels of listening proficiency (Field, 2008b).

Investigations of L2 learners' capacities to segment and extract meaning from samples of connected speech suggest that phonological modification is

strongly associated with listening ability (Field, 2008a; Lange, 2018). For example, Sheppard and Butler (2017) used paused transcription tasks to investigate the capacity of 77 L2 learners to segment strings of four or five words in connected speech. Results indicated that only 67% of the words were correctly transcribed. Other research by Wong et al. (2017) showed that reduced forms dictation (i.e., lexical segmentation with attributes of phonological modification) was the strongest correlate with listening ($r = .63$) from among several others measured such as receptive knowledge of written vocabulary ($r = .50$) and minimal pairs discrimination ($r = .32$). These studies suggest that the ability to segment words in connected speech and specifically to mitigate the effects of phonological modification plays an important role in L2 listening. As previously mentioned, Andringa et al. (2012) explicitly addressed the relationship between lexical segmentation and L2 listening and demonstrated that segmentation accuracy and L2 listening comprehension were strongly correlated ($r = .64$). However, segmentation was assessed by the test-takers' ability to accurately count the number of words in a string of target speech. Therefore, the method did not directly measure the recognition of specific word forms in connected speech, which is an important factor in L2 listening. In contrast, test formats such as paused transcription can be used to measure a learner's capacity to segment sequences of multiple words presented in connected speech (Field, 2008c). Importantly, such tests can cast light on practical questions, such as which test-takers perceive "attracts investment" as "a tax investment" (Matthews & O'Toole, 2015, p. 371) and which recognize "don't always notice" as "don't always know this" (Sheppard & Butler, 2017, p. 92). This information is not provided by tests that measure knowledge of single target vocabulary items. For this reason, data gathered from tests that measure knowledge of both single and multiple vocabulary items are likely to offer useful insight into the lexical capabilities of L2 listeners, and how these relate to listening comprehension success.

3. The study

3.1. Purpose and research questions

This study seeks to address some of the many questions that still remain around the relationship between L2 learners' capacity to handle lexical input and L2 listening comprehension. Firstly, it seeks to measure aural receptive L2 vocabulary knowledge and lexical segmentation ability among a single cohort of L2 language learners. This will allow us to determine the relative strength of association, as well as the predictive capacities, of these two measures with respect to L2 listening comprehension. Further, unlike previous investigations of the relationship

between vocabulary knowledge and a single criterion measure of listening comprehension (e.g., Andringa et al., 2012; Stæhr, 2009; Vandergrift & Baker, 2015), the current study uses two different measures of L2 listening comprehension. The listening tests that have been chosen for this study, the TOEIC and Eiken, are both relevant to the study's context, namely tertiary level EFL in Japan. The Eiken test is not well-known outside of the Japanese EFL context and therefore further information about the test will be provided in section 3.3.4. Gathering participant scores on multiple criterion measures of L2 listening comprehension and examining the relationship of these with the lexical capacities mentioned above might provide a more generalizable picture of these relationships. This may then inform testing and teaching practice in the context of the study. In an effort to do so, the following research questions will be addressed:

1. What is the relative strength of association between aural receptive vocabulary knowledge, lexical segmentation ability and the two criterion measures of L2 listening among the study cohort?
2. To what degree does aural receptive vocabulary knowledge and lexical segmentation ability predict the two criterion measures of L2 listening?

3.2. Participants

All of the 130 participants (70% females, 30% males) in this study were first-year Japanese university students enrolled in a general English course at a university in western Japan. The participants generally had six years of English education before entering university. An analysis of the participants' average TOEIC listening (229.71, $SD = 46.14$) and reading (151.27, $SD = 44.11$) scores indicated their level of English ability was A2 (basic user, waystage) in terms of the *Common European Framework of Reference for Languages* (CEFR) (Educational Testing Service, 2015).

3.3. Instruments

3.3.1. Measure of listening vocabulary level

Aural receptive vocabulary knowledge was measured with the Listening Vocabulary Levels Test (McLean et al., 2015). This test contains 150 items and was designed to measure Japanese learners' lexical knowledge of the first five 1,000-word frequency levels of the British National Corpus/Corpus of Contemporary American English (BNC/COCA) (Nation, n.d.) and the Academic Word List (Coxhead, 2000). Each of the sections from the first 1,000-word frequency level to the fifth 1,000-word frequency level contains 24 items and the final section

measuring academic word knowledge contains 30 items. The test uses a multiple-choice format which was based on the Vocabulary Size Test (Nation & Beglar, 2007). Each item consists of the target vocabulary, a non-defining sentence containing the target word and four answer choices (written in Japanese). The target word and non-defining sentence are presented once aurally but are not written on the test paper. Test-takers choose the word, which best represents the meaning of the English target word, from among four options, as shown in the example below (English translations added here for clarity):

1. (Test-taker hears: “waited: I waited for a bus.”)
 - a. 食べた (ate)
 - b. 待った (waited)
 - c. 見た (saw)
 - d. 寝た (slept)

There is a five-second pause between each item and a 15-second pause between test sections (for turning the page). The last section, testing the Academic Word List, contains 30 items and all sections can be completed in about 30 minutes. The audio files were recorded by a native speaker of American English, which was appropriate for the cohort of the current study as this is the dialect of English most commonly taught in Japanese EFL. As a demonstration of the validity of the test, a correlation of .54 was reported between the Listening Vocabulary Levels Test and Parts 1 and 2 of the TOEIC listening section (McLean et al., 2015).¹

3.3.2. Paused transcription tests

Lexical segmentation ability was assessed using a paused transcription test with five sections produced in-house by the authors. The paused transcription test format utilizes a listening text in which pauses have been inserted at irregular intervals. The pause is placed directly following a target item and the test-taker attempts to recall the last three to five words before the pause and transcribe them on the answer sheet. After the paused interval, the recording resumes playback and the test-taker continues listening until another pause is heard during which the preceding phrase is transcribed and this continues for all of the test items. A unique aspect of the paused transcription testing format is that it allows the test-taker to utilize comprehension of the aural co-text and background knowledge to assist in transcribing the target items (Field, 2008c). Other types of listening tests relying on transcription, such as standard dictation tests or partial dictation tests,

¹ See the IRIS database for the test (<https://www.iris-database.org/iris/app/home/detail?id=york%3a937862&ref=search>).

generally require the listener to transcribe target items using limited co-text or contextual information that could facilitate the application of top-down knowledge.

The audio for each of the five sections of the paused transcription test was recorded in a question-answer format between a Japanese native speaker asking the questions and a North American English native speaker answering them. The audio for each section of the paused transcription test was between 10 to 12 minutes. Each section of the test contained 12 target phrases of three words each for a total of 180 items. Following the intonation unit containing each target phrase, a 15-second pause was inserted in the audio text. In order to standardize the acoustic features of the target phrases, all pauses were inserted in the speech of the English native speaker.

The content of the dialogues included personalized anecdotes as well as many topics related to Japan that would be familiar to the study cohort. A partial sample of the dialogue used in the first section of the Paused Transcription Test is provided in Appendix A. Note that test-takers were not reading the transcript and filling in blanks while listening to the dialogues; the dialogues were only heard and the test-takers wrote their transcriptions onto a numbered answer sheet. For example, the listeners heard the following question and answer followed by a beep and a 15-second pause during which they attempted to transcribe the target phrase immediately preceding the beep, *we could play*:

Speaker 1: *What was it like?*

Speaker 2: *So growing up in St Louis was fun I lived in a neighborhood with a few kids so we could play* (beep)

When designing the test dialogues, the use of high-frequency vocabulary was prioritized in order to reduce the number of potential errors in lexical segmentation caused by inadequate vocabulary knowledge. Frequency data for all vocabulary used in the test was analyzed using the online computer program Compleat Web VP (Cobb, 2018) based on the combined COCA/BNC 1-25K corpus. Results determined that 94.8% of the 5,278 tokens used in the test were within the first 1,000-word frequency band, 3.30% were in the second, 0.60% in the third, 0.30% in the fourth, 0.50% in the fifth, and 0.10% in the sixth 1,000-word frequency band with the remaining 0.44% of words not within the corpus (i.e., offlist). In a separate frequency analysis of only the words contained in the 60 target phrases, 97.2% of the 180 target words were within the first 1,000-word frequency band, 1.70% were in the second and 0.60% in the third. Only five target words were beyond the first 1,000-word frequency band. All 60 target phrases are listed in Appendix B.

In order to ensure that the target phrases were representative of authentic language in connected speech, each phrase was designed to contain one of

three types of phonological modification: reduced function words, transitions between words (i.e., assimilation and elision) or linking (i.e., liaison). These categories of co-articulation are known to be problematic for L2 learners (Sheppard & Butler, 2017; Wong et al., 2017). The target item length was set at three words to reduce the difficulty of the transcription task while adequately representing phonological modification occurring between words.²

3.3.3. TOEIC listening

The Test of English for International Communication (TOEIC) Listening and Reading Test is used widely in Japan with approximately 3,400 organizations and educational institutions administering the test in 2017 (Institute for International Business Communication, 2018). The TOEIC listening section takes about 45 minutes to complete and contains four parts with 100 multiple-choice items. Part 1 contains 10 items in which the test-taker selects the most accurate description of a photograph. Part 2 contains 30 items which assess the listener's ability to select the best response to a question. Part 3 contains 10 dialogues with three questions each and Part 4 consists of 10 monologues with 3 questions each to assess listening comprehension. There are 495 points possible for the TOEIC Listening section.

3.3.4. Eiken Pre-2 listening

The Eiken test is an English proficiency test developed in Japan and widely used in Japanese secondary schools. There are 7 grades of difficulty from Grade 5 (easiest) to Grade 1 (most difficult). This makes it possible, in contrast to TOEIC, for a test level to be selected that aligns with the known proficiency level of a given cohort. The listening section of the Eiken Pre-2 grade, used in the current study, is ranked between Grade 3 and Grade 2 and adequate achievement on the test positions a test-taker at roughly an A2 level on the CEFR (Eiken Foundation of Japan, 2016), which was the estimated proficiency level of most of the participants in this study. The listening section consists of three parts, each containing 10 multiple-choice questions. In Part 1, the test-taker listens to short conversations and chooses the best response from three options. In Part 2, the test-taker hears longer conversations and selects the correct answer to questions. Finally in Part 3, the test-taker hears a monologue and selects the best answers to questions about it. The listening section takes approximately 20 minutes to complete.

² The audio files and materials for the paused transcription test developed for this study are available online in the Mendeley Data repository (<https://data.mendeley.com/datasets/g278w62zpg/1>; Lange & Matthews, 2020).

3.4. Procedures

This study involved the administration of four test instruments: two listening comprehension tests and two lexical measures. For the purposes of analysis, measures of listening comprehension (TOEIC & Eiken Pre-2) were identified as outcome variables, and the two lexical measures were identified as predictor variables. The Listening Vocabulary Levels Test was used to measure aural vocabulary knowledge, and the Paused Transcription Test was used to measure lexical segmentation ability.

Tests were administered in the order of Eiken Pre-2, TOEIC and Listening Vocabulary Levels Test. The five sections of the Paused Transcription Test were administered approximately once every two weeks over the course of the 15-week semester. All tests, except for the TOEIC were administered during class and necessarily spaced to reduce the cognitive burden on students and allow time for other teaching activities. Table 1 lists the instruments, their purposes and time of administration. Formal approval from the university ethics committee was obtained for this study.

Table 1 Procedure summary

Test	Construct	Administration timing
Eiken Pre-2	L2 listening comprehension	Week 2
TOEIC Listening	L2 listening comprehension	Week 12 (outside of class)
Listening Vocab Levels Test	Aural vocabulary knowledge	Week 13
Paused Transcription Test	Lexical segmentation ability	Weeks 3, 5, 7, 9, 11

The directions for all tests, besides the TOEIC, were provided in Japanese with clear examples to illustrate the listening task as well as time to ask any questions about the test. The TOEIC was administered following the standardized rule booklet provided by the testing company and only English instructions for each part of the listening section were supplied in the test booklet and spoken aloud on the test CD. The audio for all tests was administered by audio file or CD to the whole class through high-quality speakers in a quiet classroom environment.

The criterion listening tests and two vocabulary tests used multiple-choice formats so scoring was unambiguous. However, the three-word target item transcriptions for the Paused Transcription Test required the development of a scoring protocol to ensure a standard scoring method. A scoring protocol, based on principles described in Matthews, O'Toole, and Chen (2017, pp. 42-43), was devised to facilitate consistent scoring (see Appendix C). This was not a test of spelling, and so correctly spelled target words and words with minor spelling errors which clearly reflected the phonological form of the target word (e.g., *uniek* for *unique*) received one point each. A score of 0.50 was given to recognizable

but more ambiguous representations of the target word (e.g., *unik* for *unique*). A deduction of 0.25 was applied if one of the three target words was transcribed out of order or if additional words were added within the target phrase. Other incorrect words or blanks received zero points. The first author scored the Paused Transcription Test and the second author scored a subset of 10%. The correlation between the two authors' scores was very high ($r = .997$), demonstrating strong levels of inter-rater agreement.

The final scores provided by the TOEIC testing institution, rather than raw scores, were used in this study with a possible score range of 5 to 495. The other three assessments utilized raw scores and their possible range of scores are listed as follows: Eiken Pre-2 listening section 0 to 30, Listening Vocabulary Levels Test 0 to 150 and Paused Transcription Test 0 to 180.

3.5. Analysis

Correlation and multiple regression were the two statistical techniques applied in the current study. The necessary assumptions associated with linearity, multivariate normality, multicollinearity, and homoscedasticity for regression analysis were confirmed to be unviolated for this data (Tabachnick & Fidell, 2007). The sample size of 130 exceeds the rule of thumb for regression analysis stated by Green (1991) in which N should be greater than $104 + m$ (where m is the number of predictors) and thus satisfies recommendations for the number of cases-to-independent variables.

4. Results

Table 2 shows the minimum, maximum, mean and standard deviation of scores obtained from each test used in the analyses. All instruments had an adequate Cronbach's alpha level of 0.70 or above (Cortina, 1993).

Table 2 Descriptive statistics for test variables

Test	Construct	N	Min	Max	Mean	SD	Mean %	α
TOEIC Listening	L2 listening comprehension	122	115	350	229.71	46.14	45.12	.72
Eiken Pre-2	L2 listening comprehension	130	7	29	18.18	4.88	60.60	.70
Listening Vocab Levels Test	Aural vocabulary knowledge	123	68	126	101.84	11.24	67.90	.75
Paused Transcription Test	Lexical segmentation ability	113	1	139	82.66	25.32	46.00	.86

Table 3 shows that z-skewness values for each test fall below 3.29, which indicates normal distribution for medium-sized samples ($50 < N < 300$) and therefore suitable for further statistical analysis (Kim, 2013).

Table 3 Skewness and Kurtosis statistics for test variables

Test	N	Skewness	SE skewness	z-skewness	Kurtosis	SE kurtosis	z-kurtosis
TOEIC Listening	122	-.20	.22	-.91	.04	.44	.10
Eiken Pre-2	130	-.07	.21	-.33	-.57	.42	-1.36
Listening Vocab Levels Test	123	-.62	.22	-2.84	.22	.43	.50
Paused Transcription Test	113	-.54	.23	-2.35	.39	.45	.87

4.1. Research question 1: What is the strength of association between the variables that were measured?

The correlations between all four measures are presented in Table 4. To standardize descriptions of the magnitude of these correlations, Cohen’s (1992, p. 157) interpretation of small ($r = .10$), medium ($r = .30$) and large ($r = .50$) effects was used. Firstly, the two measures of listening comprehension were strongly correlated ($r = .52$). Despite aural vocabulary knowledge and lexical segmentation ability each being measures dependent upon processing stimulus through the aural modality, a small ($r = .18$) but significant correlation was observed between them.

Correlations between aural vocabulary knowledge and measures of L2 listening were small and significant ($r = .15$ and $r = .12$). Correlations between lexical segmentation ability and L2 listening were medium to strong and significant ($r = .39$ and $r = .51$). The trend in the magnitude of the correlation coefficients between the two lexical measures and both measures of L2 listening was the same: lexical segmentation ability (stronger) and then aural vocabulary knowledge (weaker).

Table 4 Summary of intercorrelations between measures from each test instrument used in analyses

Test	1	2	3	4
1. TOEIC Listening Test				
2. Eiken Pre-2 Listening Test	.52**			
3. Listening Vocabulary Levels Test	.15*	.12*		
4. Paused Transcription Test	.39**	.51**	.18**	

Note. * $p < .05$, ** $p < .01$

Correlations between the Listening Vocabulary Levels Test and the tests of listening comprehension were too small to warrant further investigation with regression analysis. However, as previous research has shown that high-frequency

aural vocabulary test scores correlate strongly with scores from standardized L2 listening tests (Matthews, 2018), the strength of correlation between each level of the Listening Vocabulary Levels Test and listening test scores was investigated. Table 5 shows that scores from the first 1,000, second 1,000 and third 1,000-word frequency levels of the Listening Vocabulary Levels Test correlated significantly at a medium level with scores from the TOEIC listening test and the Eiken Pre-2 listening test. For both listening tests, smaller non-significant correlations were found for the fourth 1,000, fifth 1,000 and Academic levels of the test.

Table 5 Summary of correlations between L2 listening tests and word frequency level sections (1K-5K and Academic) of the Listening Vocabulary Levels Test (measuring aural vocabulary knowledge)

Listening Vocab Levels Test frequency level sections	TOEIC L	Eiken Pre-2
1K	.48**	.42**
2K	.47**	.44**
3K	.33*	.30*
4K	.11	.25
5K	.03	.20
Academic	.08	.21

Note. 1K to 5K refers to sections of the Listening vocabulary Levels Test which assess knowledge of the first 1,000-word frequency level up to the fifth 1,000-word frequency. The section labelled *Academic* assesses knowledge of vocabulary included in the Academic Word List.

4.2. Research question 2: To what degree do the variables measured predict L2 listening?

As presented in Table 5, a medium to strong relationship was found between aural vocabulary knowledge of the first 1,000, second 1,000 and third 1,000-word levels (as measured by the Listening Vocabulary Levels Test) and L2 listening ability (as measured by TOEIC listening section and Eiken Pre-2 listening section). To provide a clearer picture of the relationships and relative predictive capacities these variables have on listening, hierarchical multiple regression analysis was used. The regression modelling used Listening Vocabulary Levels Test scores (1K, 2K and 3K) and Paused Transcription Test scores as predictor variables, to predict the outcome variables, TOEIC Listening and Eiken Pre-2 scores. All analyses entailed entering the Listening Vocabulary Levels Test scores before the Paused Transcription Test scores. The underlying logic of this order entry was that knowledge of single words (as measured by the Listening Vocabulary Levels Test) is fundamental to lexical segmentation ability for multi-word chunks (as measured by the Paused Transcription Test). In essence, the Listening Vocabulary Levels Test assesses both knowledge of the target words' phonology as well as their semantics, while the Paused Transcription Test is focused on phonological (i.e.,

segmental and suprasegmental) issues and arguably does not directly measure semantic knowledge. When constructing each of the regression models, the entry order of the Listening Vocabulary Levels Test scores was as follows: first 1,000-word level, second 1,000-word level, and then the third 1,000-word level. The underlying logic for this decision was that knowledge of higher frequency vocabulary is likely to be more fundamental to L2 listening than knowledge of lower frequency words (Adolphs & Schmitt, 2003).

The first model (see Table 6) sought to determine the degree to which aural vocabulary knowledge of the first 1,000, second 1,000 and third 1,000-word levels and lexical segmentation ability predicted variance in TOEIC listening scores. Aural vocabulary knowledge of the first 1,000-word level and lexical segmentation ability were the only two statistically significant variables in the model. The first 1,000-word level could account for 22% and lexical segmentation ability accounted for an additional 12% of variance in the TOEIC.

Table 6 Hierarchical Multiple Regression Model 1 - Aural vocabulary knowledge for first, second, and third 1,000-word level (AVK) and lexical segmentation ability and as predictors of TOEIC listening

Predictor	<i>R</i>	<i>R</i> ²	<i>R</i> ² change
1: First 1,000-word level AVK	.47**	.22**	.22**
2: Second 1,000-word level AVK	.53	.28	.06
3: Third 1,000-word level AVK	.54	.29	.004
4: Lexical segmentation ability	.63*	.40*	.12*

Note. * *p* < .01. ** *p* < .001

In the second model (see Table 7) again aural vocabulary knowledge of the first, second and the third 1,000-word levels and lexical segmentation ability, were used to predict the outcome variable Eiken Pre-2 listening scores. Similar to Model 1, the first 1,000-word level of the Listening Vocabulary Levels Test accounted for 21% and lexical segmentation ability accounted for an additional 17% of the variance in the Eiken Pre-2 scores, with both predictive contributions being statistically significant.

Table 7 Hierarchical Multiple Regression Model 2 – Aural vocabulary knowledge for first, second, and third 1,000-word level (AVK) and lexical segmentation ability and as predictors of Eiken Pre-2 listening

Predictor	<i>R</i>	<i>R</i> ²	<i>R</i> ² change
1: 1st 1,000-word level AVK	.46*	.21*	.21*
2: 2nd 1,000-word level AVK	.48	.23	.02
3: 3rd 1,000-word level AVK	.48	.23	.001
4: Lexical segmentation ability	.63*	.40*	.17*

Note. * *p* < .01. ** *p* < .001

Results from Model 1 (see Table 6) indicated that the first 1,000-word level aural vocabulary knowledge scores, and not the second or third, achieved statistical significance in the model and could predict 22% of the variance in TOEIC scores. In addition, Paused Transcription Test scores could predict an additional 12% of the variance in TOEIC scores, with the two lexical measures offering a combined predictive capacity of 34% to the model. Results from Model 2 (see Table 7) also revealed similar results in that the first 1,000-word level aural vocabulary knowledge scores and lexical segmentation ability could predict 38% of variance observed within Eiken Pre-2 scores. In summary, aural vocabulary knowledge of 2K, 3K, 4K, 5K and Academic word levels added no predictive capacity in regression models for predicting the variance in TOEIC and Eiken Pre-2 listening scores. However, a combination of the first 1,000-word level of the Listening Vocabulary Levels Test and the Paused Transcription Test significantly predicted variance observed in TOEIC listening scores and Eiken Pre-2 listening scores.

5. Discussion

Perhaps the most notable finding from this study was the significant predictive capacity that high-frequency aural vocabulary knowledge at the first 1,000-word level contributed to regression models for two tests of listening. Scores from the first 1,000-word level of the Listening Vocabulary Test could independently predict 22% of variance in TOEIC listening scores and 21% of variance in Eiken Pre-2 listening scores. Aural vocabulary knowledge at the 1,000-word level had more predictive power than any other predictor variable used in the models. This finding is surprising because the Listening Vocabulary Levels Test is not a test of listening comprehension and was designed to assess phonological recognition and semantic knowledge of individual words. Correlations between total scores for the Listening Vocabulary Levels Test and the two tests of listening used in this study were weak in magnitude (i.e., $r = .15$ and $r = .12$). However, when correlations were investigated separately by 1,000-word frequency level the first 1,000, second 1,000 and third 1,000-word levels of the Listening Vocabulary Levels Test had medium to large correlations with the listening tests (see Table 5). Upon further investigation with hierarchical multiple regression analysis, it was determined that only scores from the first 1,000-word level of the test contributed significant predictive capacity to both models. This finding highlights the important association that aural knowledge of high-frequency vocabulary has with listening ability. In addition, the consistency in the predictive capacity for the two different standardized tests of listening used in the regression models supports the validity of the claim that aural vocabulary knowledge of the first 1,000-word level is associated with listening ability. Furthermore, these results corroborate previous research

demonstrating that knowledge of high-frequency vocabulary is an important foundation for comprehending authentic listening texts and performance on L2 listening tests (Matthews, 2018; Matthews & Cheng, 2015; Webb & Rodgers, 2009).

Another notable finding of the current study was the strength of association between lexical segmentation ability and L2 listening. Firstly, this association was evident from correlations between Paused Transcription Tests and the two listening tests ($r = .39$ and $r = .51$ respectively). Secondly, and potentially more importantly, this strength of association was also observed in the regression analyses. In each instance, lexical segmentation ability added a significant predictive capacity beyond that offered by aural vocabulary knowledge at the first 1,000-word level (i.e., an additional 12% and 17%, see Table 6 and Table 7). This is important as, although it is clear that knowledge of the 1,000 most frequent words in the aural modality provides a foundation for L2 listening, the capacity to segment clusters of words in the aural modality adds something extra. The current study also speaks to the relative additional importance of the learners' lexical segmentation ability in the prediction of their L2 listening scores.

Stronger correlations were found between lexical segmentation ability and L2 listening scores as compared to those found between L2 listening and aural vocabulary knowledge. This result is likely due to the format of the Paused Transcription Test which measures lexical segmentation ability and more closely resembles listening processes by utilizing both bottom-up and top-down processing. It is also important to recall that the target items and contextual language used for the Paused Transcription Test consisted of very high-frequency words (0-1K). This in turn emphasizes the importance of the capacity to segment words in the first 1,000-word frequency range, which cover approximately 89% of spoken discourse (Adolphs & Schmitt, 2003). This suggests that a learner's capacity to fluently process the highest frequency words in connected speech is likely to be strongly facilitative of L2 listening comprehension.

This investigation demonstrates that better listeners had a stronger capacity to recognize the phonological form of high-frequency words and could associate these forms with an appropriate semantic representation. Further, better listeners could also more effectively segment clusters of three very high-frequency words that were presented in connected speech.

6. Pedagogical implications

This study found that aural vocabulary knowledge of the first 1,000-word level and lexical segmentation ability together could predict approximately 30% of the variance in scores for two of the most widely used tests of listening ability in Japan. These findings suggest that developing aural knowledge of high-frequency

vocabulary as well as lexical segmentation ability may be effective for improving listening performance. In terms of recommendations for classroom practice, pedagogical activities that build the capacity to aurally recognize and understand high-frequency vocabulary should be prioritized. Although there is a need to be somewhat speculative due to the limitations of the correlational research paradigm used here, a general rule of thumb based on the evidence at hand would be to ensure learners have a solid grounding in high-frequency aural vocabulary before explicitly addressing vocabulary beyond the second 1,000-word range. As almost 90% of the vocabulary used in typical spoken discourse is from the first 1,000-word level (Adolphs & Schmitt, 2003) it seems very important that L2 listeners develop fluent recognition of these most frequently occurring words.

These findings also support the assertion that helping learners build knowledge of words as they occur in speech is an important strand of vocabulary knowledge development, and that such endeavors are likely to result in positive language learning outcomes (Siegel, 2016). Here we hypothesize that such interventions are likely to be especially impactful in learning contexts within which vocabulary knowledge development has been traditionally addressed through reading and writing largely without also presenting the target words in contextualized speech. Rather than only judging vocabulary to be “known” when a learner can establish a form-meaning link for written words, educators are encouraged to reexamine vocabulary learning in terms of learners’ aural recognition and comprehension of words in connected speech as well. Limited development of aural vocabulary knowledge and lexical segmentation ability could result in poorer listening ability for even high-frequency words (Carney, 2020). In addition, instructional approaches should be developed that improve learners’ familiarity with the phonological form of words as they occur in connected speech and which also enhance learners’ ability to comprehend chunks of lexis under time constraints.

Regular use of test formats that require the learner to process lexis through the aural modality is suggested, especially those that target the highest frequency words (e.g., Matthews, 2018; McLean et al., 2015). As shown by the results of this study, combining semantic assessment of high-frequency vocabulary via the Listening Vocabulary Levels Test with assessment of form recognition via the Paused Transcription Test may be more predictive of actual listening ability. Such testing is likely to be useful in enabling teachers to stay abreast of the aural vocabulary knowledge status of their students and their ability to comprehend and segment that vocabulary in connected speech. If used as a diagnostic tool, as recommended by Field (2003), such testing will provide data that can be used to inform pedagogical decisions aimed at developing learners’ capacity to better handle lexis mediated through the aural modality. Keeping records on the types of segmentation errors that occur amongst learners is also

suggested, and regular use of paused transcription tests such as those used in this study is likely to be a valuable way of doing this. Such data may be used to assess and facilitate the development of aural vocabulary knowledge and lexical segmentation skills necessary for listening development.

These pedagogical recommendations are particularly important in the Japanese EFL educational context (and others like it) as an inordinate amount of effort from students and teachers is focused on learning increasingly lower frequency vocabulary in preparation for university entrance exams (Kobayashi, 2001). However, this can result in a substantial difference between aural and written vocabulary sizes for learners (Mizumoto & Shimamoto, 2008). An increased focus on evaluating aural vocabulary knowledge and lexical segmentation ability through formats such as the Paused Transcription Test and Listening Vocabulary Levels Tests could help to emphasize the importance of these skills as well as to diagnose listening difficulties for learners. Such a focus on the development of skills for listening proficiency is needed to promote more balanced aural/oral English skills for Japanese learners. Further, such a focus may encourage a cultural change within EFL pedagogy in Japan towards assessment for learning (Davison & Leung, 2009), namely increased use of assessment modes that inform ongoing teaching and learning decisions. Additionally, finding time to facilitate verbalized introspection, especially in the student's L1, immediately after individual learners engage with paused transcription tests can provide an even deeper insight into the origins of segmentation errors. Such information could help to inform other bespoke classroom-based interventions aimed at promoting lexical segmentation (e.g., Field, 2003; Siegel & Siegel, 2015).

7. Limitations and future research

One possible limitation is that the relatively low proficiency level of the participants indicates they may have been unfamiliar with much of the low-frequency vocabulary from the 1,000 word-level and above. Possible floor effects for sections of the Listening Vocabulary Levels Test containing low-frequency vocabulary may have diluted the value of the aural vocabulary knowledge data. However, the participants' mean score for the test overall was roughly 67.8% and therefore did not indicate excessively low scores.

A central objective of this study was to provide a preliminary snapshot of the relationships between scores from test instruments measuring lexical capacities and L2 listening among a cohort of Japanese EFL learners. An important area for future research will be to expand the scope of similar studies both within larger cohorts of Japanese EFL students, as well as with learners with different L1 backgrounds and linguistic proficiency levels. Of interest in this regard

is to determine the degree to which the generalized trends observed as part of the current study are mirrored or contrasted among other cohorts of learners.

A further suggestion for future research is to investigate the efficacy of interventions aimed at enhancing learners' capacity to handle lexis from the aural modality. Longitudinal studies that involve tracking the development of aural vocabulary knowledge and lexical segmentation ability as targeted pedagogical interventions are of particular interest. Further, verifying the validity of the assertion that improvements in lexical segmentation ability and aural vocabulary knowledge can directly improve L2 listening comprehension is key. Confirming or refuting such assertions will require the implementation of quasi-experimental research paradigms.

The development of a broader array of tests that measure lexical capacities mediated through the aural modality is another important future research direction. In particular, the development of tests that measure the capacity to handle multiple sequential words is warranted. This seems especially important in light of the specific and robust relationship between L2 listening comprehension and the capacity to segment, recognize and understand lexis mediated through the aural modality.

8. Conclusion

Overall, our findings suggest that greater learner familiarity with high-frequency vocabulary, at the first 1,000-word level in particular, may contribute more to overall listening proficiency than aural knowledge of lower frequency words. Further, it seems clear that lexical segmentation ability is significantly associated with L2 listening ability. Measurements of lexical segmentation ability derived through paused transcription testing provide the opportunity to assess aural recognition of chunks of lexis within connected speech. The listener's ability to establish form-meaning links between high frequency aural vocabulary, and the capacity to recognize phonologically modified chunks of lexis are very useful indicators of general listening comprehension.

References

- Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics*, 24(4), 425-438. <http://doi.org/10.1093/applin/24.4.425>
- 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, 49-78. <http://doi.org/10.1111/j.1467-9922.2012.00706.x>
- Carney, N. (2020). Diagnosing L2 listeners' difficulty comprehending known lexis. *TESOL Quarterly*. Advance online publication. <http://doi.org/10.1002/tesq.3000>
- Cheng, J., & Matthews, J. (2018). The relationship between three measures of L2 vocabulary knowledge and L2 listening and reading. *Language Testing*, 35(1), 3-25. <http://doi.org/10.1177/0265532216676851>
- Cobb, T. Compleat Web VP v.2 [computer program]. <https://www.lex tutor.ca/vp/comp/>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159. <http://doi.org/10.1037//0033-2909.112.1.155>
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98-104. <http://doi.org/10.1037/0021-9010.78.1.98>
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34(2), 213-238. <http://doi.org/10.2307/3587951>
- Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly*, 43(3), 393-415.
- Educational Testing Service. (2015). Mapping the TOEIC tests on the CEFR. Princeton, NJ: Educational Testing Service. https://www.ets.org/s/toEIC/pdf/toEIC_cef_mapping_flyer.pdf
- Eiken Foundation of Japan (2016). *Comparison table*. <http://www.eiken.or.jp/eiken/en/research/comparison-table.html>
- Field, J. (2003). Promoting perception: Lexical segmentation in L2 listening. *ELT Journal*, 57(4), 325-334. <http://doi.org/10.1093/elt/57.4.325>
- Field, J. (2008a). *Listening in the language classroom*. Cambridge, UK: Cambridge University Press. <http://doi.org/10.1017/CBO9780511575945>
- Field, J. (2008b). Revising segmentation hypotheses in first and second language listening. *System*, 36, 35-51. <http://doi.org/10.1016/j.system.2007.10.003>
- Field, J. (2008c). Bricks or mortar: Which parts of the input does a second language listener rely on? *TESOL Quarterly*, 42(3), 411-432. <http://doi.org/10.1002/j.1545-7249.2008.tb00139.x>

- Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research*, 26(3), 499-510. http://doi.org/10.1207/s15327906mbr2603_7
- Hulstijn, J. H. (2003). Connectionist models of language processing and the training of listening skills with the aid of multimedia software. *Computer Assisted Language Learning*, 16(5), 413-425. <http://doi.org/10.1076/call.16.5.413.29488>
- Institute for International Business Communication. (2018). *TOEIC program data & analysis 2018*. Tokyo: IIBC. http://www.iibc-global.org/library/default/toEIC/official_data/pdf/DAA.pdf
- Kim, H. (2013). Statistical notes for clinical researchers: Assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry and Endodontics*, 38(1), 52-54. <http://doi.org/10.5395/rde.2013.38.1.52>
- Kobayashi, Y. (2001). The learning of English at academic high schools in Japan: Students caught between exams and internationalization. *The Language Learning Journal*, 23(1), 67-72. <http://doi.org/10.1080/09571730185200111>
- Lange, K. (2018). Analyzing difficulties in aural word recognition for Japanese English learners: Identifying function words in connected speech. *CASELE Research Bulletin*, 48, 63-73.
- Lange, K., & Matthews, J. (2020). Paused Transcription Test (Lange & Matthews, 2020), Mendeley Data, V1. <http://doi.org/10.17632/g278w62zpg.1>
- Matthews, J. (2018). Vocabulary for listening: Emerging evidence for high and mid-frequency vocabulary knowledge. *System*, 72, 23-36. <http://doi.org/10.1016/j.system.2017.10.005>
- Matthews, J., & Cheng, J. (2015). Recognition of high frequency words from speech as a predictor of L2 listening comprehension. *System*, 52, 1-13. <http://doi.org/10.1016/j.system.2015.04.015>
- Matthews, J., & O'Toole, J. M. (2015). Investigating an innovative computer application to improve L2 word recognition from speech. *Computer Assisted Language Learning*, 28, 364-382. <http://doi.org/10.1080/09588221.2013.864315>
- Matthews, J., O'Toole, J. M., & Chen, S. (2017). The impact of word recognition from speech (WRS) proficiency level on interaction, task success and word learning: Design implications for CALL to develop L2 WRS. *Computer Assisted Language Learning*, 30(1-2), 22-43. <http://doi.org/10.1080/09588221.2015.1129348>
- McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research*, 19(9), 741-760. <http://doi.org/10.1177/1362168814567889>
- Milton, J. (2013). Measuring the contribution of vocabulary knowledge to proficiency in the four skills. In C. Bardel, C. Lindqvist, & B. Laufer (Eds.), *L2 vocabulary acquisition, knowledge and use: New perspectives on assessment*

- and corpus analysis* (pp. 57-78). European Second Language Association. <http://www.eurosla.org/monographs/EM02/EM02home.php>
- Mizumoto, A., & Shimamoto, T. (2008). A comparison of aural and written vocabulary size of Japanese EFL university learners. *Language Education and Technology*, 45, 35-51. http://doi.org/10.24539/let.45.0_35
- Nation, I. S. P. (n.d.). The BNC/COCA headwords lists. [PDF files]. <http://www.victoria.ac.nz/lals/about/staff/paul-nation>
- Nation, I. S. P., & Beglar, D. (2007). A vocabulary size test. *The Language Teacher*, 31(7), 9-13.
- Pellegrino, F., Coupé, C., & Marsico, E. (2011). Across-language perspective on speech information rate. *Language*, 87, 539-558. <http://doi.org/10.1353/lan.2011.0057>
- Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing*, 10(3), 355-371. <http://doi.org/10.1177/026553229301000308>
- Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. Kunnan (Ed.), *Validation in language assessment* (pp. 41-60). Mahwah, NJ: Erlbaum. <http://doi.org/10.4324/9780203053768>
- Rost, M. (2002). *Teaching and researching listening*. Essex: Longman. <http://doi.org/10.4324/9781315833705>
- Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behavior of two new versions of the Vocabulary Levels Test. *Language Testing*, 18(1), 55-89. <http://doi.org/10.1191/026553201668475857>
- Sheppard, B., & Butler, B. (2017). Insights into student listening from paused transcription. *CATESOL Journal*, 29(2), 81-107.
- Siegel, J. (2016). Listening vocabulary: Embracing forgotten aural features. *RELC Journal*, 10(3), 377-386. <http://doi.org/10.1177/0033688216645477>
- Siegel, J., & Siegel, A. (2015). Getting to the bottom of L2 listening instruction: Making a case for bottom-up activities. *Studies in Second Language Learning and Teaching*, 5(4), 637-662. <http://doi.org/10.14746/ssllt.2015.5.4.6>
- Stæhr, L. S. (2009). Vocabulary knowledge and advanced listening comprehension in English as a foreign language. *Studies in Second Language Acquisition*, 31(4), 577-607. <http://doi.org/10.1017/s0272263109990039>
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston: Allyn & Bacon/Pearson Education.
- Vandergrift, L., & Baker, S. (2015). Learner variables in second language listening comprehension: An exploratory path analysis. *Language Learning*, 65(2), 390-416. <http://doi.org/10.1111/lang.12105>
- Wallace, M. P. (2020). Individual differences in second language listening: Examining the role of knowledge, metacognitive awareness, memory, and attention. *Language Learning*. Advance online publication. <http://doi.org/10.1111/lang.12424>

- Webb, S., & Rodgers, P. (2009). The lexical coverage of movies. *Applied Linguistics*, 30(3), 407-427. <http://doi.org/10.1093/applin/amp010>
- Wong, S. W. L., Mok, P. P. K., Chung, K. K., Leung, V. W. H., Bishop, D. V. M., & Chow, B. W. (2017). Perception of native English reduced forms in Chinese learners: Its role in listening comprehension and its phonological correlates. *TESOL Quarterly*, 51(1), 7-31. <http://doi.org/10.1002/tesq.273>

APPENDIX A

A partial sample of the dialogue used for the first two target phrases for Paused Transcription Test 1

Where did you grow up?

I grew up in St Louis Missouri it's in the center of the United States and it's on the Mississippi River it's a fairly big city.

What was it like?

So growing up in St Louis was fun I lived in a neighborhood with a few kids so we could play. We usually just played sports or rode our bicycles, it was... it was a good childhood.

What were your parents like?

My parents were a little strict I guess. I couldn't stay out very late I guess you know I had to come home when the ... when it began to get dark ... dinner time but they didn't pressure me to do homework. On the weekends I usually had to do a lot of housework and there was always washing the dishes or vacuuming or cleaning something so my friend said my parents were strict.

(Note: partial sample only)

APPENDIX B

Target words used in each section of the Paused Transcription Test

Item #	PTT 1	PTT 2	PTT 3	PTT 4	PTT 5
1	we could play	lots of pictures	a few weeks	sort of thing	how important listening
2	me to do	interested in Japanese	walk to work	know them as	skill that helps
3	we had together	she made sure	still in a	the hard part	be able to
4	quite happy for	to learn more	some books about	I grew up	you go on
5	she got out	for learning another	why did you	in the evenings	not just yourself
6	when we were	interesting for me	I can do	would be fun	you can get
7	like to play	what it was	things to learn	get off the	found out later
8	visit those natural	eat my favorite	can learn about	helped him get	had to be
9	what I like	which is exciting	first you're	don't steal	near the castle
10	do a lot	it looked like	seasons are really	was turned away	when you walk
11	which are both	at the store	aren't allowed	most of the	kinds of unique
12	that's all	wash your hands	as a good	comes to mind	thousand years old

APPENDIX C

PTT scoring rubric with rationale and examples

General instructions for scoring individual words in target phrases

Score	Principle	Comments	Example answer → correct target word
1.0	The word is spelled correctly.	This answer is easy to score because there is no subjectivity involved.	unique → unique evenings → evenings favorite → favorite
1.0	The word is spelled incorrectly but its phonological form is acceptable according to English phonology.	<i>Subjectivity involved.</i> The test construct is aural decoding therefore slight spelling errors are accepted as long as the spelling approximates the phonological form of the target word.	unike/uniek/unieque → unique wosh → wash wark → work natral → natural heands → hands pictuer → picture turnd → turned thousan → thousand cathle/casltel/castl/casle/catsle → castle lestening/lisning → listening wuld → would allowd → allowed
0.75	A homophone of the target word is decoded instead of the target word	Although the target word was accurately decoded phonetically, the spelling of the word indicates the wrong word was decoded indicating difficulty with understanding the meaning of the input.	steel → steal witch → which aloud → allowed
0.5	The word is spelled incorrectly and has more ambiguity in the interpretation of its phonological form.	<i>Subjectivity involved.</i> The incorrect spelling results in an incorrect phonological form that does not approximate the target word. However, there is partial evidence of correct phonological recognition depending on the interpretation of the word's spelling.	unik/unic/unice/uniece/unecue → unique larn/laur → learn laurning → learning alaud/aroud → allowed yourself → yourself gat → got leastening/listeing → listening thousant → thousand excating → exciting watsh → wash seson → season reary → really rater → later wark → walk
0.5	Incorrect conjugation/ incorrect form of verb but clear evidence that the root word is recognized.	The core element of the target word is recognized correctly but there is an error in inflection or word form, such as tense or plurality.	played → play visiting → visit interesting → interested look → looked was → is make → made will → would can → could a → the can't → can come → comes are → aren't

Exploring the relationships between L2 vocabulary knowledge, lexical segmentation, and L2...

0	Significant spelling mistakes make interpretation of the phonological form difficult and its association with the target word tenuous.	The phonological form of the spelling represents a clearly different word from the target word. (Two or more incongruencies with phonological form.)	leran → learn leauning → learning sousend/thouthont/sousond → thousand unirk → unique araude/arowd → allowed
0	A different word, which may be phonologically similar is decoded.	The orthographic form represents a clearly different word from the target word. Despite the phonological similarities, the accurate spelling of the transcribed word demonstrates that a separate word from the target word was decoded.	quit/quiet → quite way → away national → natural pray → play a → are leaning → learning fan → fun mine → mind listing → listening waking → walking latter → later
0	No target word provided		

General instructions regarding deducting points for errors in the target phrases

Score	Principle	Comments	Example target word → example answer
- 0.25	0.25 points are deducted for mistakes of word order in the target phrase.	One of the words in the three-word target phrase is transcribed in an incorrect order relative to the other two target words.	to me do → me to do you hushed hands → wash your hands to walk → walk to you my why → why did you can I do → I can do the most → most of the important how listening → how important listening listening is important → how important listening helps skills → skill that helps
- 0.25	0.25 points are deducted for every extra word in the target phrase transcription.	An extra word is contained within the target phrase. It must come between two of the words in the target phrase.	We have time together → we had together a few day on weeks → a few weeks comes to the mind → comes to mind comes to my mind → comes to mind grow them up → I grew up of the thing → sort of thing kind of the unique → kinds of unique you are going → you go on