

# Assessing the Efficacy of Dictation Exercises to Improve SLA Listening in Japan

**Christopher Edelman**  
Ritsumeikan University, Kyoto, Japan  
ORCID ID: 0000-0002-0177-2059

**Robert McClung**  
Kansai University, Osaka, Japan  
ORCID ID: 0000-0002-0930-3695

**Peter Ferguson**  
Nara Prefectural University of Education, Nara, Japan  
ORCID ID: 0000-0001-6635-6331

Although dictation is seen as being a useful exercise for assessment and progress in language learning, there has been little research into the use of spaced dictation. The focus of this research was assessing the effect of limiting cognitive burden with spaced dictation exercises as classroom activities. Participants attended pre- and post-tests to evaluate gains in listening ability from a four-week intervention activity of spaced dictation exercises. Results were measured by using the number of content words, function words, and affective state of participants at the beginning and end of the intervention. The results demonstrated that spaced dictation was efficacious for improving listening ability, motivation, and participant feeling of achievement. The authors concluded that spaced dictation was effective in increasing student ability to decode and identify linguistic features due to the decrease of cognitive burden, which allowed greater processing of aural input. Contrary to the perception of dictation activities as a grueling task for learners, a large majority of participants reported being greatly satisfied with the tasks that were conducted. These findings aid educators and researchers interested in the efficacy and results of dictation exercises in second language acquisition (SLA).

**Keywords:** Cognition, Comprehension, Segmentation, Suprasegmentals

Some of the difficulty of learning a second language lies in the inability to decode, segment, differentiate, and identify lexical items in natural spoken language. This is apparent in Japan, where many students study English for a minimum of six years before entering the higher education system but are only able to understand simple greetings or phrases when spoken at a very slow pace. Listening is built upon a series of interconnected cognitive processes that are created and reinforced through exposure and practice.

Attention and training of suprasegmentals is often neglected in language classrooms (Gilakjani, 2011). One reason might be the highly intrinsic development of the listening system, and the impracticality of teaching prosodic features of every word combination possible. Listening ability occurs with the development of linguistic knowledge and reinforcement from repeated exposure - with the acquisition apparatus itself being intrinsic - it is a skill that cannot be forced, but it can be aided.

Audio recordings used in English language classrooms can be viewed as slow, simplified, stilted, and unnatural. While this may aid in creating a simple phonetic representation, lexis in naturally spoken speech is quite different. In the context of naturally spoken English, learners may suddenly find themselves unable to comprehend a never-ending stream of connected sounds, or identify and segment it into chunks of comprehensive input. Thus, the rationale for this study was to investigate whether spaced dictation can be a positive aid in listening comprehension. By decreasing the cognitive burden in listening by providing more time for processing, spaced dictation would aid learner recognition and acquisition of suprasegmentals and other prosodic information, thereby increasing overall listening comprehension and identification of lexis.

### **Literature Review**

The definition of dictation is typically summarized as “a technique where the learners receive some spoken input, hold this in their memory for a short time, and then write what they heard” (Nation, 1991, p.12). Dictation has been used in the classroom as a learning activity viewed as efficacious in developing listener comprehension while making a connection with the target language in its written form (Pappas, 1977; Whitaker, 1976). In addition, it has been viewed as an all-purpose tool that is accurate in not only teaching but also testing language learners (Morris, 1983). In a recent study by Wong and Leeming (2014), the authors reported strong correlation between a measure of dictation and standardized tests, which highlights the parsimonious advantage offered by using dictation to measure learner proficiency.

The smallest units of sound are phonemes, then morphemes, and then words, which are then grouped into lexical chunks and connected through prosodic features, suprasegmentals, intonation, and inflection (Rost, 2011). Activities that incorporate focused attention while listening have been found to raise learner awareness and ability to identify suprasegmental features (Gorsuch, 2001). Suprasegmentals are vocal effects that occur within a combination of more than one sound and characterize the way words blur and blend together (Lindfield, Wingfield, & Goodglass, 1999). Whether words are accurately retrieved from these sounds depends on the phonetic representation retained within the listener’s lexis. Allophonic variations due to co-articulation, assimilation, reduction, and elision occur naturally for native speakers, and make listening easier due to the decreased processing demands; however, they are problems for non-natives, particularly if they have only acquired the form by learning it detached from natural discourse (Rost, 2011).

Some studies (e.g. Martin & Ellis, 2012) have been conducted that focus on the connection between the cognitive burden placed upon working memory and phonological loop. However, there is still a gap in the research concerning the longitudinal results of dictation exercises that

are broken into lexical units to reduce cognitive load. With his Cognition Hypothesis, Robinson (2005) asserts that increasing complexity is essential to activate maximum cognitive involvement in order to achieve linguistic gains; on the other hand, this same input must be within the range of ability for learner linguistic processing if there is to be comprehension of the input received. In summary, if language learners are overwhelmed with incomprehensible input, there is no opportunity for reflection, identification, or understanding to occur. Limited capacity and selective attention are two important factors that influence listening (Rost, 2011). Selective attention describes the ability to focus on only one source of information at any given time. Listeners are limited to how much information they are able to maintain in their phonological loop, which acts as a temporary storage area for new audio input (Baddeley, 1986). If learners are overwhelmed with incomprehensible input and unable to focus their attention on features of the language, they will be unable to acquire them. It is important for working memory not to be overloaded if learning is to occur (Just & Carpenter, 1992). The studies cited above have examined varied styles of implementing dictation (e.g., group listening, communicative, multiple listenings of the same recording). However, to the best knowledge of the three researchers involved in this study, spaced dictation, which is based on psycholinguistic theory, has not been examined. With this in mind, the current study sought to test the efficacy of spaced-dictation exercises for improving listener ability to identify and comprehend suprasegmentals in spoken messages. We defined spaced dictation exercises as dictation exercises that include intervals between lexical chunks by breaking up the sentences into noun, verb, and adverbial phrases. By not over-loading the working memory capacity, it was assumed that listeners would be able to attend the features of the input, primarily the suprasegmentals.

This study was guided by four main hypotheses:

1. Increased ability to identify suprasegmentals will result in an increased written function word count.
2. Overall comprehension will increase parallel to an increase in more accurate decoding of audio input.
3. Working memory will positively correlate with increased comprehension and word count.
4. The difference in confidence on the affective scores will show systematic positive improvement.

In addition, there was the overall question of which measure would be the greatest predictor of future outcomes.

## **Method**

### **Participants**

A single intact class at a private university in Kansai, Japan in the second semester of a one-year English course participated in the study. There were 32 students included in the data at the beginning of the study, although seven were later discontinued from the study due to absence. Gender was equally divided with 15 females and 17 males. Data from seven students were dropped because they had missed two or more interventions, post-test, or pre-test leaving

a total of 25 (females = 12, males =13). While individual student Test of English for International Communication (TOEIC) scores were unavailable, participants were assigned to classes according to TOEIC results before the start of the school year. Placement within this class indicated a TOEIC score of less than 400. Learners were not involved in any other English classes, and previous interviews indicated that none were involved in any extracurricular events or study groups that would influence linguistic development in any way.

### **Procedure and Materials**

A pre-test was administered during the first week of instruction at the beginning of class. Immediately after the dictation exercises, students completed comprehension questions related to the listening, then filled out the affective measurement survey (see Appendix A). The three activities took no longer than ten minutes. The class was given The Vocabulary Level Listening test (LVLT) within the last 15 minutes of the same class period.

Spaced dictation activities were carried out at the beginning of each class from Week 2 to Week 5 (four sessions), with administration of the digit span test at the end of class in Week 2. In the spaced dictation sessions, students were given a handout with blank lines on the front. They were then instructed to listen to the recording and write down as much as possible. After students finished the spaced dictation task they turned the paper over and completed the comprehension questions on the back (see Appendices). The transcript was then displayed in large font on the overhead projector and answers to the comprehension questions were reviewed with the first author of this paper. Students were also given a few minutes to compare their notation with the transcript. During the review time, students were instructed not to write anything down or change their answers. After the exercise, all papers were collected by the instructor. Suprasegmental and phonological speech features were not explicitly taught so as not to influence learner development. The total time for the activity was roughly 15 minutes each week.

In the last session in week six, the post-test and extended affective measurement surveys were administered within the first fifteen minutes (See Appendix B). All tests and surveys were retrieved and scored separately by two teachers.

### **LVLT**

The LVLT is a vocabulary level listening test, created by McLean, Kramer, and Beglar (2015) based on word frequency. This was used as a benchmark for learner knowledge of the vocabulary incorporated into activities and tests. This listening test was used as opposed to a written vocabulary test to make sure an accurate representation of what learners were able to comprehend only through auditory input.

The LVLT consists of six different levels ranging from the first one thousand most frequent words to the sixth one thousand most frequent according to the Academic Word List (AWL) (McLean et al., 2015). Within each level are 24 test items, each with four multiple-choice answers. For this study, the Japanese version was used; learners had to match the item heard in English to its Japanese equivalent. The test component was played over the classroom audio

system during administration. The recording consisted of an item word spoken once, then repeated in a context that would not define its meaning (e.g., shoe, he has a shoe).

For the purpose of this study, only the first two levels were administered because finding participant maximum vocabulary ability was not the main interest. Rather, it was used to evaluate the assumption that listeners had sufficient knowledge of the first one thousand most frequent words to alleviate the issue of vocabulary knowledge confounding the test scores. The total administration time was roughly eight minutes, four minutes for each level.

### **Japanese Digit Span Test**

The digit span test is made up of multiple lists ranging from two digits long to ten. For each subset, six lists of similar length are read separately with a one-second interval in between. The listener must repeat back or write down the digits in the order they were heard after the last digit in the list is read. Successfully repeating back five or more of the lists for one subset receives a passing score. The longest subset successfully completed indicates a corresponding digit span score. If a listener is able to accurately repeat back five 6-digit lists, then their digit-span score is six. In order not to confound proficiency with working memory, the Japanese language version was administered.

The test was administered as a measurement of the capacity of the phonological loop, which underlies one of the verbal measures of working memory (Baddeley & Hitch, 1974). This measure was included to test the hypothesis of the study that learners with a higher working memory would be able to maintain and reflect upon a larger section of information, achieving greater long-term gains. It was also chosen because measuring constructs involved in cognition can be difficult, and the measurement tool must account for other influencing variables, such as language, in order to achieve a pure score (Declerck, Philipp, & Koch, 2013). Total time for the test was roughly ten minutes.

### **Dictation Activities and Tests**

All tests and activities were created using the same guidelines to maintain equivalency. Only lexical items within the first one thousand most frequent words were utilized. The reasons for this were a) for measurement purposes, it was important that unknown vocabulary was not a confounding variable, and b) struggling with new unknown words would further burden listeners' cognitive faculties, pulling attention away from focusing on suprasegmentals and depressing the ability to integrate awareness of new sounds. In addition, no relative clauses or perfect aspect verbs were used. Every dictation script was similar in length: around 105 words including function and content words.

Tests were reviewed separately by two different English professors to confirm equivalency before being administered. On the pre- and post-tests, there were 15-second intervals after every full stop. Listeners were allowed to write during any point of the exercise, though not allowed to speak to each other or work with another person. Dictation audio recordings were made using average speaking speed (120 wpm). This is less than the average speaking speed of 165 to 180 wpm (Rubin, 1994); however, the speed was deemed appropriate by the researchers for the target audience.

The dictation activities were made with the same parameters as the tests in terms of lexis, syntactic complexity, and word count. However, they differed in the length of interval and speech. All sentences were broken into lexical phrases, with a 5- to 7-second interval between them. A concern was that the pauses mid-sentence would break up the natural rhythm during reading, altering the type of onsets and stressed words in the sentence. To account for this concern, all scripts were recorded using full-length sentences. Stresses, intonation, and accents were marked and used as a point of reference when recording the spaced phrase version. Both recordings were later compared for equivalence by a rater before being administered in class. Every session consisted of five multiple choice comprehension questions given on the back of the dictation handout. All questions were information style questions with distractors that were plausible but incorrect only according to content.

Both the pre- and post-tests had surveys measuring the affective state of the listener after they completed the dictation and questions sessions. These surveys used 4-point Likert scales to measure listener a) perceived difficulty of activity, b) confidence, c) anxiety, and d) effort. In addition, the post-test included two additional items; one related to the level of enjoyment the participant received from doing the task, and the other to their perceived value of the dictation exercises for improving their English language ability.

Tests were constructed using a similar theme: relating a time of activity in the recent past. The pre-test, Week 1, was based around how the instructor spent time over the summer (keeping in mind this class was in the fall semester), while the post-test, Week 6, was centered on how the instructor spent the previous weekend. Many of the basic sentence structures were kept the same, while changing most content words. In the pre-test, the final sentence is "I'm looking forward to spring break already, because I want to travel more if I have the chance." This is highly similar in structure and length of the last sentence of the post-test "I'm looking forward to winter break already, because I want to rest more if I have the chance." This was done to mitigate any possibility of grammatical difficulty creating a confound in learner ability between the two different tests. This also helped ensure that the pre- and post-tests were equivalent in terms of difficulty.

The four dictation exercises used as the intervention were built around a continuing story line, a narration of Tom and Susan Smith and their lives in a small town. Exercises were not dependent upon having fully comprehended the previous one. If students were unable to completely grasp the information in an exercise, were absent, or simply forgot, it would not affect their ability to comprehend succeeding exercises (cite research supporting this concept). In order for listeners to direct their attention mainly on listening and decoding word forms, the storyline was kept simple for each exercise, with the only linear theme being the fact that Tom and Susan live in a small town.

## Results

### Digit Span and LVL

Descriptive statistics for the vocabulary measure were conducted. For the first one-thousand word level, the mean was 22.68 ( $SD = 1.02$ ) demonstrating coverage of 94.5% of the items. The second one thousand words level, the mean was 19.92 ( $SD = 2.38$ ). Because the listening

activities were created using only lexis from the first one thousand words level, the large coverage of the first one-thousand word level ensured knowledge of the vocabulary contained in the tasks. On the measure of working memory represented by the digit span test, an average length of 4.2 digits was observed ( $SD = .87$ ).

### Dictation

The analysis of dictation notation on both tests was conducted by counting the number of content words, function words, and morphological deviances in each test. Content words were defined as being any noun, adjective, adverb, or verb (excluding auxiliary verbs, e.g., *do, is, have*). Function words were anything that lay outside of these parameters, e.g., pronouns, prepositions, and articles. Because orthographic accuracy was not a consideration, words that were misspelled were not counted as deviant. If a word was written in a way that its phonetic representation could be understood it was included in the data set as long as it was properly placed. Morphological mistakes were counted separately, and not included with the number of accurately scored function or content groups. Only words that were spoken in the dictation in roughly the same linear order they occurred were included. All data was entered into a Microsoft Excel file then transferred to SPSS (version 23).

Paired samples *t* tests were used to compare the relationship for the measures of the number of function words, content words, morphological deviations, and comprehension. The results for this can be found in Table 1. For function words, there was a decrease from the pre-test that was significantly less than the post-test,  $p < .001$ . While there was no hypothesis regarding the content words, there was a significant increase between the pre-test and the post-test counts,  $p < .000$ . Looking at the word count averages, the relationship of the two scores was almost completely inverse between the two tests. There was also a significant decrease in the number of morphological errors as well,  $p < .021$ . Finally, the comprehension scores on the pre-test were significantly lower than the post-test,  $p < .000$ .

**Table 1**  
**Comparison of Pre- and Post-Test Results**

	Pre <i>M</i>	Pre <i>SD</i>	Post <i>M</i>	Post <i>SD</i>
Morphological Errors	2.08	1.55	1.24	1.12
Content Words	12.28	3.73	17.8	4.2
Function Words	15.16	5.09	12.6	5.22
Total # of Words	27.44	8.04	30.4	8.89
Correct Answers	1.08	0.9	2.84	0.99

Even though there was a decrease in the mean for function words and an increase on content words on the post-test compared to the pre-test scores, this change did not sacrifice one for the other. The total word count, including content and function words, was significantly higher on the post-test compared to the pre-test,  $p < .009$ .

A comparison of the scores from the comprehensive listening questions attached to each test was done using paired sample *t* tests. Compared to listener ability to comprehend the narration

on the pre-test ( $M = 1.08$ ,  $SD = .91$ ) large significant improvement was shown on the post-test ( $M = 2.84$ ,  $SD = .99$ ),  $t(25) = -7.33$ ,  $p < .000$ .

**Affective Tests**

Paired samples  $t$  tests were also used to compare the affective state of the listeners at the beginning and end of the intervention. It was hypothesized that confidence would be significantly improved at the end of the intervention, and that learners' perceived difficulty of the task would significantly decrease as they became more familiar with it.

In measures of learner confidence, there was a significant increase between the pre-test ( $M = 3.6$ ,  $SD = .65$ ) and the post-test ( $M = 2.92$ ,  $SD = .91$ ),  $t(25) = 4.654$ ,  $p < .000$ . Although there was a decrease in the mean score for confidence, the scoring for this measure was reversed, with a score of 1 indicating higher confidence, and a score of 4 indicating low confidence. This was done to keep all Likert items moving in similar relation, with a higher number indicating a higher negative affective score.

This change in confidence was an expected result, particularly when compared to the change in perceived difficulty on the pre-test ( $M = 3.56$ ,  $SD = .58$ ) that was significantly less than the post-test ( $M = 3.12$ ,  $SD = .73$ ),  $t(25) = 2.86$ ,  $p < .009$ . On the measure of learner anxiety, results from the pre-test ( $M = 3.0$ ,  $SD = .65$ ) were significantly lower than the post-test ( $M = 2.48$ ,  $SD = .82$ ),  $t(25) = 2.59$ ,  $p < .016$ . However, when comparing the measure of effort on the pre-test ( $M = 2.52$ ,  $SD = .82$ ) to the post-test ( $M = 2.64$ ,  $SD = .70$ ), there was no significant difference,  $t(25) = -.55$ ,  $p < .588$ .

Descriptive statistics were used to look at the last two items on the post-test survey: learner enjoyment of the activity and perceived usefulness of the activity for improving skills in English. On the measure of enjoyment, there was a mean score of 2.20 ( $SD = .76$ ). The stem for this item was "I enjoy this type of practice," with answers of "1 = Yes, a lot, 2 = Yes, a little, 3 = No, 4 = I don't like it." Table 2 gives a more accurate representation of the frequency of scores.

**Table 2**  
**Level of Learner Enjoyment**

Score	Frequency	Percent	Cumulative Percent
1	4	16	16
2	13	52	68
3	7	28	96
4	1	4	100
Total	25	100	

The mean score of 2.20 places participant opinion just over the divide from not enjoying the exercise, to enjoying it to a small degree. Looking at the frequencies provides a more detailed view. Only one student chose a score of 4, indicating dislike of the task, and seven chose a score of 3, indicating they did not like it. However, because they did not choose the option of dislike, this score might be considered a neutral point as far as interest is concerned. Overall,

17 students out of 25 ranked scores indicating a measure of enjoyment, with four choosing the highest score of enjoyment.

All students reported positive scores regarding their perception of the usefulness of the task ( $M = 1.56$ ,  $SD = .50$ ). The stem for this item was "This type of practice is good to improve my English skill," with answers of "1 = Yes, 2 = A little, 3 = Not really, 4 = Not helpful at all." Frequencies for these results are shown in Table 3, showing the class roughly divided between perceiving the tasks having a small degree of usefulness and a large degree of usefulness.

**Table 3**  
***Usefulness of the Task***

Score	Frequency	Percent	Cumulative Percent
1	11	44	44
2	14	56	100
Total	25	100	

### **General Linear Model**

In order to assess which variables would be the best predictors for future ability, a multivariate general linear model was used. The results are displayed in Table 4. Word count averages for content words and function words, the vocabulary listening test score (LVLT), working memory measure (digit-span), and the four affective measures were used as covariates; the dependent variables were word counts on the post-test.

**Table 4**  
**Relation Between Pre-Test Variables and Writing Outcomes**

Pre-test	DV	<i>MS</i>	<i>F</i>	<i>Sig.</i>	Partial E2
Corrected Model	PostCon	40.74	6.64	0.001	0.77
	PostFun	58.69	5.09	0.003	0.72
Intercept	PostCon	14.08	2.3	0.149	0.13
	PostFun	0.18	0.02	0.903	0
Content Words	PostCon	3.18	0.52	0.482	0.03
	PostFun	1.36	0.12	0.736	0.01
Function Words	PostCon	56.82	9.27	0.008	0.37
	PostFun	160.23	13.9	0.002	0.47
Digit-span	PostCon	6.85	1.12	0.306	0.07
	PostFun	11.93	1.04	0.324	0.06
LVLT	PostCon	50.95	8.31	0.011	0.34
	PostFun	4.39	0.38	0.546	0.02
Difficulty	PostCon	0.26	0.04	0.839	0
	PostFun	18.27	1.59	0.226	0.09
Confidence	PostCon	5.59	0.91	0.354	0.05
	PostFun	0.04	0	0.955	0
Anxiety	PostCon	6.49	1.06	0.319	0.06
	PostFun	14.99	1.3	0.271	0.08
Effort	PostCon	4.21	0.69	0.42	0.04
	PostFun	27.17	2.36	0.144	0.13

a R Squared = .769 (Adjusted R Squared = .653)

b R Squared = .718 (Adjusted R Squared = .577)

PostCon = Post-test content words.

PostFun = Post-test function words.

Partial E2 = Partial Eta Square.

None of the affective variables shared any systematic relationship with either outcome. This lack of positive correlation can be seen with the measure of working memory as well, looking at outcomes for content words  $F(1,25) = 8.31$ ,  $p = .306$ , partial eta square = .07, as well as function words  $F(1,25) = 1.04$ ,  $p = .324$ , partial eta square = .06. This was counterintuitive and ran against the third hypothesis of the study. One possibility is that the sample was not large enough, limiting the variance of the score, or that there were problems in administration of the test.

The LVLT was seen to share a significantly positive relationship with content words  $F(1,25) = 1.12$ ,  $p = .011$ , with a robust effect size (partial eta square = .34). There was no strong correlation between it and function words  $F(1,25) = .38$ ,  $p = .546$ , partial eta square = .02;

however, this is not surprising, as the words measured in the LVLT are primarily content words (nouns, adjectives, adverbs).

Pre-test count average of function words was the greatest predictor for either dependent variable: content words  $F(1,25) = 9.27, p = .008$ , partial eta square = .37, and function words  $F(1,25) = 13.90, p = .002$ , partial eta square = .47. While it might seem obvious that function words would at least be an excellent predictor of future outcomes of itself, the same was not true with the measure of content words, which has no systematic relationship with either dependent variable. Out of all pre-test variables, learner ability to recognize and notate function words was the greatest predictor of future ability to accurately notate the dictation activity.

### Discussion

The findings of the study were in line with our hypothesis that dictation practice incorporating the segmentation of full sentences would decrease the cognitive burden and allow time for reflection and analysis by participants. Over time, this practice resulted in an increase in positive gains of lexical recognition by the participants in the post-test, which did not employ segmentation but used full sentences. The results of this study show spaced dictation to be a highly efficacious activity for increasing language learner ability to process and comprehend spoken messages of English aural input received at a natural speaking speed, with normal intonation.

Concerning the hypothesis that comprehension would increase with listener ability to process suprasegmentals, the data appears to point towards the null hypothesis. Although listener comprehension increased, the number of written function words in the post-test decreased. However, this decrease does not necessarily show a similar decrease in ability to process suprasegmental features. Notation is not a pure representation of cognitive functions; it is volitional, with the writer choosing what information to write down. Complex cognitive functions, such as the decoding of prosodic information are non-volitional, being carried out automatically as soon as input is received. Although listeners were not recording the function words in their notation on the post-test, it is entirely possible they heard the words but did not write them down.

With this in mind, it is possible that even though learners were decoding more information, they were prioritizing words with the most value: content words. The previous dictation exercises were reviewed, showing that participants were not prioritizing in this same manner during the tasks, rather writing down full phrases of information. These tasks were conducted differently, and had a 5- to 7-second interval between lexical phrases. When listeners encountered the longer streams of text in the post-test, it is likely that they changed their tactic, and prioritized content words because these words were the most vital to understanding the meaning of the text. This was also found to be true in a study by Field (2008), which investigated prioritization of function words versus content words by L2 listeners; the study concluded that function words, although highly frequent, were given much less focus by listeners because of their lack of meaning. In the current study, because of the sudden increase of complexity and cognitive burden imposed, listeners prioritized meaning over form.

The lack of any relationship between working memory with either of the writing measures was unexpected, but could be an issue with the lack of variance within the working memory measure. Digit span scores had a range of 3 (*Min* = 3, *Max* = 6), were content words had a range of 14 (*Min* = 12, *Max* = 26), and function words a range of 23 (*Min* = 2, *Max* = 25). Another additional issue could be the small size of this sample for running parametric tests. This leads to the largest limitation of this study: the small sample size. Conducting the treatment with a larger number of participants across a wider range of language abilities would add strength to the results and give greater variance to the subsets of measures.

The hypothesis that learner confidence would increase proved to be true. The affective scores indicated that there was a systematic increase in learner confidence on the post-test, while learner perception of task difficulty and feelings of anxiety decreased. This implies that although learners perceived their overall affective state to be more positive and the difficulty of the task to be lower, they did not perceive any change in their own effort when completing the task.

The high level of achievement as well as an increased listening ability most likely contributed to the positive trend of the scores on the affective survey. Slightly surprising was the level of enjoyment students reported receiving from the exercise. It is also possible that this is a culture specific issue as well. Japanese language students are often resistant to speak English in the classroom, or engage in communicative tasks. In a study by King (2013) of nine different universities, nearly 48 English language classes, and almost 1,000 students, student initiated speech accounted for less than one percent of total class time in each class – approximately 54 seconds. Part of this hesitance to speak may be due to cultural expectations in Japan of a teacher-centered class and reluctance to make mistakes in front of peers (Cutrone, 2009). Because of this apprehension to communicate, participants in our study might have enjoyed the solitary nature of the activity and the structure of the lesson because it was not dependent on interaction. All students reported the task as being beneficial for their improvement, indicating a level of satisfaction in its own.

As shown in the general linear model in Table 3, the function word count from the pre-test proved to be the strongest significant predictor for outcomes on both dependent variables. At first glance this appears an oddity, but retrospectively viewing the aims of the research make the reason quite clear: greater ability to accurately decode suprasegmental features applies to function words, because they are the least salient auditory input. They are unstressed, altered by the prosodic features by the words around them, and often elided with increased speed of speech (Rost, 2011). Properly identifying function words on the pre-test could be interpreted as a proxy of listening ability for suprasegmental features. The ability to decipher this input is indicative of a higher capability to decode phonological input resulting in greater listening ability overall.

### **Conclusion**

This study found spaced dictation exercises to have great utility in the language classroom. Even though dictation activities may carry with them negative connotations among language educators, this study found spaced dictation to be effective in improving learner ability to

recognize lexis, comprehend naturally spoken input, and build learner confidence. Future research comparing the outcome of spaced dictation with other dictation techniques is necessary to provide an understanding of the differences between in order to improve student outcomes.

### **Author note**

#### **Chris Edelman**

Chris Edelman is currently a lecturer at Ritsumeikan University in the department of Science and Engineering. His research interests include psycholinguistics, L1 influence on language acquisition, phonological representation creation and retrieval, content-based instruction, and using rhythm in the language classroom.

#### **Peter Ferguson**

Peter Ferguson teaches second language teacher education at Nara University of Education, focusing on elementary and junior high school English. His research interests cover elementary school EFL curriculum and pedagogy, pre- and in-service teacher education, language policy, sociolinguistics, early literacy, and comparative education.

#### **Robert McClung**

Robert McClung is a full-time contract teacher at Kansai University. He has taught for over 10 years in Japan. He is currently pursuing a doctorate in education with a specialization in college teaching and learning. He is focused on increasing student positivity and intrinsic motivation in technology-based, student-centered classrooms.

### References

- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8, pp. 47-90). San Diego, CA: Academic Press.  
[https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Baddeley, A.D. (1986). Working memory. New York: Oxford University Press.
- Cutrone, P. (2009). Overcoming Japanese EFL learners' fear of speaking. *Language Studies Working Papers*, 1, 55-63.
- Declerck, M., Philipp, A. M., Koch, I. (2013). Bilingual control: Sequential memory in language switching. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(6), 1793-1806. <https://doi.org/10.1037/a0033094>
- Field, J. (2008). Bricks or mortar: Which parts of the input does a second language listener rely on? *TESOL Quarterly*, 42(3), 411-432. <https://doi.org/10.1002/j.1545-7249.2008.tb00139.x>
- Gilakjani, A. P. (2011). A study on the situation of pronunciation instruction in ESL/EFL classrooms. *Journal of Studies in Education*, 1(1), 1-15.
- Gorsuch, G. J. (2001). Testing textbook theories and tests: The case of suprasegmentals in a pronunciation textbook. *System*, 29(1), 119-136. [https://doi.org/10.1016/S0346-251X\(00\)00049-X](https://doi.org/10.1016/S0346-251X(00)00049-X)
- Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99(1), 122-49.  
<https://doi.org/10.1037/0033-295X.99.1.122>
- Lindfield, K.C., Wingfield, A., & Goodglass, H. (1999). The contribution of prosody to spoken word recognition. *Applied Psycholinguistics*, 20, 395-405.  
<https://doi.org/10.1017/S0142716499003045>
- Martin, K. I., & Ellis, N. C. (2012). The roles of phonological short-term memory and working memory in L2 grammar and vocabulary learning. *Studies in Second Language Acquisition*, 34(3), 379-413. <https://doi.org/10.1017/S0272263112000125>
- McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Testing*, 1-20.  
<https://doi.org/10.1177/1362168814567889>
- Morris, S. (1983). Dictation – a technique in need of reappraisal. *ELT Journal*, 37, 121-126.  
<https://doi.org/10.1093/elt/37.2.121>
- Nation, P. (1991, October). Dictation, dicto-comp and related techniques. *English Teaching Forum*, 29(4), 12-14.
- Pappas, G.S. (1977). You mean you still give dictations? *Language Arts*, 54, 936-939.
- Robinson, P. (2005). Cognitive complexity and task sequencing: Studies in a componential framework for second language task design. *IRAL-International Review of Applied Linguistics in Language Teaching*, 43(1), 1-32. <https://doi.org/10.1515/iral.2005.43.1.1>
- Rost, M. (2011). *Teaching and researching listening* (2nd ed.). Harlow, UK: Pearson Education.
- Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, 78(2), 199-221. <https://doi.org/10.1111/j.1540-4781.1994.tb02034.x>
- Whitaker, S.F. (1976). What is the status of dictation? *Audio-Visual Language Journal*, 14, 87-93.

Wong, A. & Leeming, P. (2014). Using dictation to measure language proficiency. *Language Education in Asia*, 2014, 5(1), 160-169.

[https://doi.org/10.5746/LEiA/14/V5/I1/A13/Wong\\_Leeming](https://doi.org/10.5746/LEiA/14/V5/I1/A13/Wong_Leeming)

**Appendix A**

**Week 1 Pre-Test Transcript, Comprehension Questions, and Affective Survey**

This summer, I stayed at home most days. Usually, I prepared for my classes this fall. It takes a lot of time, but it's very important. I need to prepare lots of things for my students because I'm the teacher.

Of course, I didn't stay in my house for all of summer. I went to the park with my friends, but the weather was too hot. I also went to a party a couple weeks ago. The weather was cooler that day, so I had a good time. I'm looking forward to spring break already, because I want to travel more if I have the chance.

- |   |  |
|---|--|
| 1. This summer, your teacher:<br>A) Travelled a lot<br>B) Travelled sometimes<br>C) Never travelled<br>D) Felt cold in the party                          | 4. Which activity matches the weather?<br>A) Park - hot<br>B) Party - hot<br>C) Park - hot   Party - hot<br>D) Park - hot   Party - a little too hot   |
| 2. Your teacher:<br>A) Always stayed home<br>B) Usually was scared of work<br>C) Did some work<br>D) Didn't do any work                                   | 5. During the next break, your teacher wants to:<br>A) Go to the park again<br>B) Take a trip if possible<br>C) Travel to an exciting place<br>D) Travel to a cool place because they don't like hot weather |
| 3. Your teacher feels:<br>A) Their job is important<br>B) Happy when students get high scores<br>C) Sorry to work again<br>D) The weather was hot at work |  |

---

Please answer honestly.

- |  |   |
|--|---|
| 1. The difficulty was...<br>1 – Not difficult<br>2 – A little difficult<br>3 – Difficult<br>4 – Very Difficult | 3. My anxiety was...<br>1 – No anxiety<br>2 – A little anxiety<br>3 – Anxious<br>4 – Very anxious |
| 2. My confidence was...<br>1 – Very confident<br>2 – Confident<br>3 – A little confidence<br>4 – No confidence | 4. My effort was...<br>1 – Very high<br>2 – High<br>3 – Low<br>4 – Very low                       |

### Appendix B

#### Week 6 Post-Test Transcript, Comprehension Questions, and Affective Survey

Last weekend, I enjoyed time with my family. Friday, I cooked a nice dinner at home. It takes a little time, but it's not difficult. I went to the store and bought the food I needed before I went home to cook.

On Saturday, the weather was really nice. I went with my family and some friends to a park near the lake for lunch. We also went to the lake on Sunday. The weather was cooler that day, so we dressed in warm clothes. I'm looking forward to winter break already, because I want to rest more if I have the chance.

- |                                  |  |
|----------------------------------|--|
| 1. Last weekend, your teacher:   | 4. Which day matches the weather?                |
| A) Went to the city              | A) Saturday – Warmer than Friday                 |
| B) Stayed home and relaxed       | B) Friday and Saturday – Warm                    |
| C) Spent time with his family    | C) Friday – Warm   Sunday – A little warm        |
| D) Doesn't like cold weather     | D) Saturday - Warm   Sunday – Cool               |
| 2. Your teacher:                 | 5. During the next break, your teacher wants to: |
| A) Always stayed home            | A) Go to the lake again                          |
| B) Didn't enjoy the weekend      | B) Not do anything exciting                      |
| C) Cooked for his family         | C) Travel to an exciting place                   |
| D) Worked on tests               | D) Travel to a warm place                        |
| 3. Your teacher:                 |  |
| A) Cooks every Friday            |  |
| B) Bought food before going home |  |
| C) Always buys food on Friday    |  |
| D) Is a good cook                |  |

---

Please answer honestly.

- |                          |   |
|--------------------------|---|
| 1. The difficulty was... | 4. My effort was...   |
| 1 – Not difficult        | 1 – Very high   |
| 2 – A little difficult   | 2 – High  |
| 3 – Difficult            | 3 – Low   |
| 4 – Very Difficult       | 4 – Very low  |
| 2. My confidence was...  | 5. I enjoy this type of practice                              |
| 1 – Very confident       | 1 – Yes, a lot  |
| 2 – Confident            | 2 – Yes, a little   |
| 3 – A little confidence  | 3 – No  |
| 4 – No confidence        | 4 – I don't like it   |
| 3. My anxiety was...     | 6. This type of practice is good to improve my English skill. |
| 1 – No anxiety           | 1 – Yes   |
| 2 – A little anxiety     | 2 – A little  |
| 3 – Anxious              | 3 – Not really  |
| 4 – Very anxious         | 4 – Not helpful at all  |