Engagement, Task Evaluation, and Vocabulary Acquisition Through Extensive Listening

Satoko Suzuki

Ibaraki University

In extensive listening (EL) outside the classroom, students select listening materials based on their proficiency level and interests; however, studies of its effectiveness are scarce. This study explored engagement, task evaluation, and vocabulary acquisition through a one-semester EL intervention. Participants included 75 Japanese university students studying once a week for 15 weeks in a general English course and required to practice EL outside of class. The results showed that they engaged in EL for an average of one hour per week and had a relatively negative evaluation for EL. Nonetheless, the statistical analysis showed significant aural vocabulary improvement for all groups, with greater improvements seen among students with a lower initial vocabulary level. This study concludes that EL homework is effective, but adjustments are necessary to facilitate vocabulary learning.

多聴活動では学生が自らの英語力と興味にあった教材を選ぶが、その効果に関する研究は少ない。本研究では、1学期間の多聴への取り組み、多聴に対する評価、多聴後の語彙習得について調査した。参加者は週1回15週間の一般英語クラスを受講する75名の日本人大学生で、授業外で多聴に取り組んだ。その結果、彼らは平均週1時間の多聴に取り組んだが、彼らの多聴への評価は比較的否定的であった。一方、語彙テストスコアの向上は全てのグループで確認され、語彙力の低いグループのスコアが語彙力の高いグループより向上した。この研究結果から、授業外での多聴で語彙の学習効果はあったが、その効果をより高めるためには調整をする必要があると分かった。

In an English as a foreign language (EFL) environment such as Japan, where English-language exposure is usually limited to the classroom, students need to listen to English outside the classroom to improve their listening ability. One promising method is extensive listening (EL), a task to increase students' exposure to spoken English by selecting listening materials based on their proficiency and interests. While previous studies have found the effectiveness of listening practice based on vocabulary acquisition (Vidal, 2011; van Zeeland & Schmidt, 2013a), few have investigated whether vocabulary acquisition occurs as a result of EL and whether the acquisition varies depending on initial vocabulary knowledge. Also, few studies used aural vocabulary tests such as the Listening Vocabulary Levels Test (LVLT) to measure vocabulary acquisition. Furthermore, because student perception of EL might affect task engagement, psychological factors also need to be examined. Therefore, this study investigated the effect of EL on aural vocabulary acquisition among students with different initial aural vocabulary levels as well as students' opinions toward EL. In the following sections, first, studies related to vocabulary, listening, and psychology as concerns EL are reviewed. Then, the online listening materials used for EL in this study, and two instruments (an aural vocabulary test and a task evaluation questionnaire) are explained along with procedures. Finally, the results are examined to generate suggestions for an improved method to conduct EL.

Literature Review

Vocabulary Acquisition Through Listening

To date, studies on whether listening practice promotes vocabulary acquisition are still scarce. Previous research attempted to measure the effect of listening practice on vocabulary acquisition and found promising results. Although the gains were not large compared to those of reading practice, vocabulary acquisition occurred as a result of listening practice (e.g., Brown et al., 2008; Vidal, 2011; van Zeeland & Schmitt, 2013a).

In addition, considering individual differences in proficiency, Vidal (2011) conducted a study in an ESP course in Spain and found that vocabulary learning differs depending on students' proficiency: Lower-proficiency students learn better by reading rather than

listening, but fewer differences in conditions have been found among higher proficiency students. These differences might be derived from the learning strategy. Learning strategy studies have found listening strategy differences depending on proficiency: Lower-proficiency students focused on each word and relied on translation, while higher-proficiency students could focus on sentences and phrases to understand meaning using background knowledge and inferring from context (see review for Oxford, 2017).

The findings above suggest that initial vocabulary knowledge, or which frequency word level students have mastered, is an important factor that influences and affects vocabulary acquisition during EL because learning strategies may differ depending on proficiency. Learners' vocabulary knowledge, particularly aural vocabulary knowledge, especially early on, might impact the acquisition process because EL requires them to choose their listening materials and use strategies to solve listening difficulties. In the next section, how much vocabulary is necessary to understand spoken language and the importance of aural vocabulary are introduced.

Vocabulary for Listening

Vocabulary research on listening has found that language learners need to know 95% of the vocabulary when listening to and understanding informal narratives (van Zeeland & Schmitt, 2013b). Also, a vocabulary of around 3,000-word families has been found to cover 95% of the coverage of spoken language in informal situations such as narratives, movies, and television programs (Adolphs & Schmitt, 2003, 2004; Webb & Rodgers, 2009a, 2009b). Thus, acquiring 3,000 frequently used words is crucial for listening comprehension. However, these results are based on vocabulary tests that measure reading vocabulary knowledge (e.g., Vocabulary Levels Test, Vocabulary Size Test).

When chunking of the stream of speech fails during listening, learners may not recognize words they know in the written form, which impairs language processing and listening comprehension (Goh, 2000). Thus, for listening success, learners need to recognize the spoken form of the word. A test to measure aural vocabulary knowledge such as Listening Vocabulary Levels Test (LVLT) should be used to examine the effect of listening practice. When there are too many unknown words during listening, learners are overwhelmed and their motivation to listen can be impeded (see Vandergrift, 2005; Vandergrift & Goh, 2012). In the following section, the ideal conditions that are important for learners to engage in EL are reviewed.

Extensive Listening and Its Influencing Factors

EL is one way to incidentally learn vocabulary and natural language use through listening. Since the focus is on meaning, learners can learn new words by guessing from context (Nation, 2013). Recently, even in an EFL environment such as Japan, the development of the Internet as well as the widespread use of mobile phones allows language learners to easily access a wide variety of both written and spoken English; thus, anyone who has an internet connection can experience EL. However, whether language learners take advantage of those ample resources depends on psychological factors such as individual motivation, feeling competent, and the autonomy to complete the tasks (Ryan & Deci, 2000).

For learners to engage in EL, the following three psychological factors may be important. First, they need to be intrinsically motivated to learn English phonology. English phonology has special characteristics such as reduced forms, delivery speed, and stress-timed prosody, making it difficult to understand spoken English (Brown & Lee, 2015). Learners' desire to listen and communicate in English is important to overcome difficulties. Second, they need to feel competent in completing EL tasks. To do so, they should be given comprehensible listening materials and an appropriate workload. Difficult materials and a heavy workload should be avoided because learners may think it is too excessive and beyond their capacity. It can, thus, negatively affect their perceived competence (Ryan & Deci, 2008). Third, they have the autonomy to choose listening materials. They need to choose materials that are interesting to them. To avoid autonomy loss, learners should not be placed under too much pressure or behavior control (Brown & Lee, 2015). For successful EL engagement, conditions such as learners' intrinsic motivation toward English phonology, appropriate listening materials, and a reasonable workload that encourages learners to feel competent and autonomous when choosing EL materials are necessary.

When learners engage in EL, they listen to many materials at their proficiency level and guess unknown words in context. As a result, aural vocabulary acquisition would occur, but how much a given student acquires may vary depending on their initial vocabulary knowledge. To examine aural vocabulary learning and related psychological factors, the following questions were examined in this study:

1. To what extent do learners engage in EL? How many hours per week do they listen on average?

2. What are the learners' task evaluations (intrinsic motivation, perceived competence, perceived choice, and pressure) toward EL? What is the relationship among these variables?

3. Does a course incorporating EL improve learners' aural vocabulary? If so, does aural vocabulary improvement differ depending on the learners' initial aural vocabulary level?

Methods

Participants

The participants were 75 Japanese university students: 22 third-years from the humanities department (12 males, 10 females), 21 third-years (21 males), and 32 second-year students (31 males, 1 female) from the engineering department. Their TOEIC scores were between 350 and 545. All participants signed the consent form to participate in this study. The participants attended a general English course once a week for 15 weeks in the 2019 spring semester and used a textbook intended for listening and reading sections of the TOEIC test. The course also required engaging in 30 hours of EL at home. At the end of the course, participants submitted an EL report that indicated when, for how long, and what content they engaged in listening.

Materials

For the EL task, participants mainly used three websites: ELLLO (http://www.elllo.org/index.htm), TED (https://www.ted.com/talks), and ALC Net Academy 2 (online TOEIC preparation course). Although more websites were recommended, the participants selected the above websites for listening practice. These websites offered audio and corresponding text files; therefore, the participants could refer to the text files, in case they had difficulty understanding the listening materials.

Instruments

Listening Vocabulary Levels Test

To investigate whether there were any gains in how many spoken forms of the most frequent words of English are known, an LVLT was conducted in weeks 1 and 11 of the course. The audio files for LVLT were recorded by a general American English speaker, which is the most familiar English accent for Japanese learners of English. The complete LVLT (McLean et al., 2015) measures the first 5,000 word-frequency levels of the British National Corpus/Corpus of Contemporary American English with 24 items per level, and the Academic Word List with 30 items. For each item, a word and a sentence that includes the word are read, and the participants choose from four multiple choice options in Japanese. For adequate comprehension of most listening materials, students need to know the first 3,000 word-frequency levels (van Zeeland & Schmitt, 2013a),

therefore, this study used those three levels only. When participants have a score of at least 23 out of 24, they have passed the level (McLean & Kramer, 2016).

Task Evaluation Questionnaire

To check how students perceive EL, a task evaluation questionnaire, which is a shorter version of a modified version of the Intrinsic Motivation Inventory (see Ryan, 1982; Self-Determination Theory, n.d.; Suzuki, 2019), was used. The questionnaire consisted of 20 items (5 items for each of 4 constructs *intrinsic motivation*, *perceived competence*, *perceived choice*, and *pressure*). Examples of items include *I enjoyed this task very much* for intrinsic motivation, *I am satisfied with my task performance* for perceived competence, *I had enough opportunities to prepare while completing this task* for perceived choice, and *I felt very intense while doing this task* for pressure. A 6-point Likert scale was used for this inventory (1 = *strongly disagree*, 2 = *disagree*, 3 = *slightly disagree*, 4 = *slightly agree*, 5 = *agree*, 6 = *strongly agree*). At the end of the questionnaire, open-ended questions were asked: *What were the good points of the EL assignment? What were the bad points of the EL assignment?* These were asked to supplement the responses of the Likert scale questionnaire.

Procedure

In week 1, the participants completed the first 3,000 word-frequency levels of LVLT. Based on the test results, the participants were divided into three groups: below the 1000 word-frequency level (*below 1k*), 1000 word-frequency level (*1k*), and 2000 word-frequency level (*2k*). There were 22 students at below the 1k level, 45 students at the 1k level, and 8 students at the 2k level. In week 11, participants took the LVLT, Task Evaluation Questionnaire, and reported how long they engaged in EL at home.

Results

Research Question 1: To What Extent Do Learners Engage in EL? How Many Hours per Week Do They Have Time to Do EL on Average?

First, to answer research question 1, the participants completed a survey on their use of EL in week 11. The self-reported time of usage was analyzed using descriptive statistics on IBM SPSS Statistics version 26. The results showed that on average, students completed 12 hours ($Mdn = 10$, $Min = 0$, $Max = 26$) of EL in week 11: specifically, the below 1k group reported 11 hours ($Mdn = 10$, $Min = 2$, $Max = 26$), the 1k group 12 hours ($Mdn = 10$, $Min = 0$, $Max = 25$), and the 2k group 12 hours ($Mdn = 12$, $Min = 9$, $Max = 18$). Each groups' hours engaging in EL are shown in Figure 1.
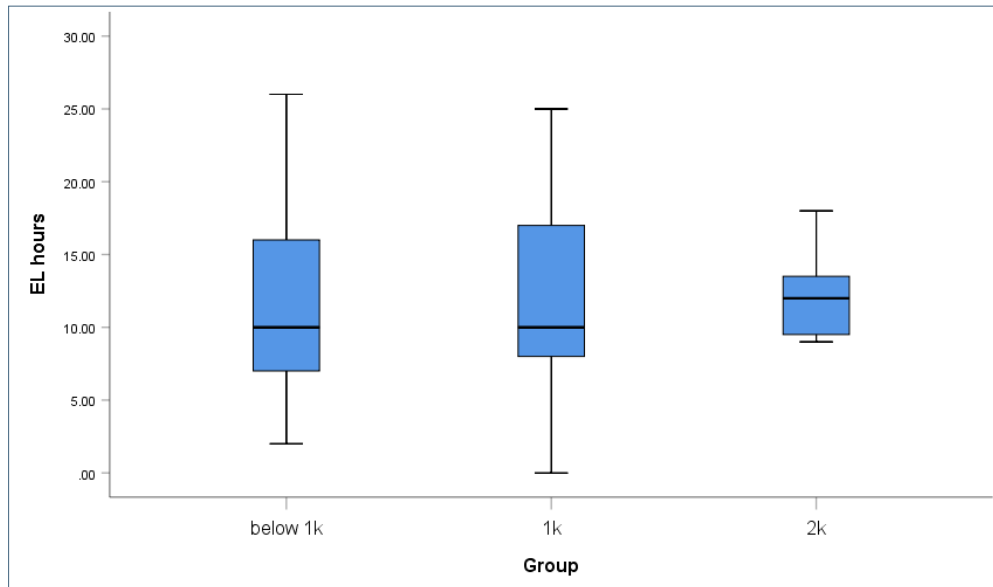
Figure 1
EL Hours for Three Listening Vocabulary Level

Research Question 2: What Are the Learners' Task Evaluation (Intrinsic Motivation, Perceived Competence, Perceived Choice, and Pressure) Toward EL? What Is the Relationship Among These Variables?

Second, to examine research question 2, a task evaluation questionnaire was conducted. Following Agnesia (2010), the reliability of all four variables was checked first running a Cronbach's alpha test in IBM SPSS Statistics version 26. Except for perceived choice, all the constructs met the acceptable threshold: One item in the perceived choice category was deleted to achieve reliability. The results showed that the overall reliability was good (= .83). Regarding each of the four constructs, intrinsic motivation was the most reliable (= .90), followed by pressure (= .69), perceived competence (= .62), and perceived choice (= .62). Descriptive statistics for the four variables showed that the mean score was 3.33 for intrinsic motivation (*Mdn* = 3.40, *Min* = 1.00, *Max* = 5.20), 3.17

for perceived competence (*Mdn* = 3.20, *Min* = 1.60, *Max* = 4.80), 3.65 for perceived choice (*Mdn* = 3.75, *Min* = 1.25, *Max* = 5.25), and 2.48 for pressure (*Mdn* = 2.60, *Min* = 1.20, *Max* = 4.60).

The quantitative results of the task evaluation questionnaire above can be supported by the responses to the open-ended questions. While most students regarded EL homework as a good chance to listen to and familiarize themselves with English, several negative points were reported: 30 hours of required EL time was too much (*n* = 18), the recommended listening materials were difficult (*n* = 9), concentrating on listening to English at home was difficult (*n* = 6), writing the EL report was troublesome (*n* = 7), and working on EL constantly was hard (*n* = 4).

Correlation analysis was also conducted to examine the relationships among the four variables. As presented in Table 1, intrinsic motivation was significantly correlated with two variables: A weak positive correlation was found with perceived competence (*r* = .34, *p* < .01), and a moderately strong positive correlation was found with perceived choice (*r* = .77, *p* < .001). In addition, perceived competence was found to have a weak positive correlation with perceived choice (*r* = .41, *p* < .001).

Table 1
Intercorrelations for Four Variables

| Variable	1	2	3	4
1. Intrinsic motivation	—			
2. Perceived competence	.34*	—		
3. Perceived choice	.77**	.41**	—	
4. Pressure	.19	-.04	.20	—

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

Research Question 3: Does a Course Incorporating EL Improve Learners' Aural Vocabulary? If So, Does the Improvement of Aural Vocabulary Differ Depending on Learners' Initial Aural Vocabulary Level?

To check research question 3, a two-way repeated measures of analysis of variance (ANOVA) was used to examine the statistical differences of the score gains. Before conducting a two-way repeated measure of ANOVA, univariate outliers and the normality of the data were checked. Regarding the posttest of 1k group, one outlier was

found and *z*-scores of skewness exceeded the absolute value of 1.96. Thus, to meet the assumption of normality, one participant in 1k group was excluded from the analysis and the total number of participants became 74.

To measure whether the score gain was statistically significant, a two-way repeated measure of ANOVA was conducted with *groups* (three levels: below 1k, 1k, 2k) as the independent variable and *vocabulary gains* (pretest, posttest) as the dependent variable. A statistically significant difference was found in terms of *vocabulary gains* [$F(1,71)$ = 16.97, $p < .01$, = .19] and *group* [$F(2,71)$ = 15.02, $p < .01$, = .30] but not in terms of *interaction* [$F(2,71)$ = .69, n.s., = .02]. This means that the score increase was found in all groups. The descriptive statistics showed that all the groups improved LVLT scores: An increase of 2.5 scores for the below 1k group, 2.84 for the 1k group, and 1.1 for the 2k group (see Figure 2).

Figure 2
Pretest and Posttest Results of LVLT for Three Listening Vocabulary Level Groups
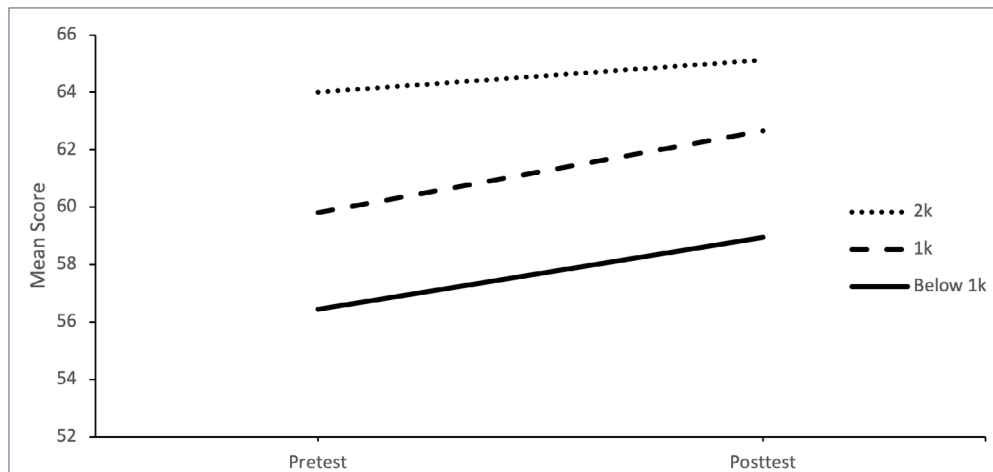
Discussion

First, students' EL engagement can be explained by the results of the Task Evaluation Questionnaire. Students reported they completed only about half of the required EL by week 11. Specifically, they should have accomplished 22 hours by week 11; however, they

reported only 12 hours on average (*Mdn* = 10, *Min* = 0, *Max* = 26). This might be because, as the results of the Task Evaluation Questionnaire for EL showed, students were not motivated to do EL, did not feel competent, and did not have autonomy in choosing listening materials. Also, according to the responses from the open-ended questions, which asked students to explain the good and bad points of EL, although they reported it was good for them to get used to listening to English, several kinds of disadvantages were reported: 30 hours of EL was too much and felt mandatory, it was hard to find listening materials at their level by themselves, it was difficult to concentrate on listening to English at home, it was hard to write down comments for each listening material, and they could not do EL constantly and eventually did long hours all at once. Overall, the fact that they did only half of the expected EL hours by week 11 was due to less intrinsic motivation, perceived competence, and perceived autonomy to choose listening materials.

The results of the correlation analysis of four variables (intrinsic motivation, perceived competence, perceived choice, and pressure) can suggest what we can do to conduct EL more effectively: Learners should be provided with autonomy to select listening materials that match their interests and levels. The correlation analysis found that perceived competence and perceived choice were related, and intrinsic motivation was related to students' perceived competence and autonomy. Therefore, students who perceived themselves to be competent might be able to choose listening materials on their own or when they could choose listening materials at their level, they might feel competent. In order to increase intrinsic motivation, students might need to feel competent and should be given the opportunity to choose listening materials that are both interesting and at their level.

Second, although students' EL engagement and evaluation for EL were not positive, they improved their aural vocabulary. Students only engaged in half the expected EL hours and three task evaluation constructs were perceived negatively: Students slightly disagreed with feeling motivated to do EL (*M* = 3.33), they slightly disagreed with feeling competent to do EL (*M* = 3.17), and they slightly disagreed with feeling autonomy in choosing listening materials (*M* = 3.65); however, they gained their aural vocabulary test scores. All three aural vocabulary groups improved their LVLT scores after 11 weeks, although the improvements were small. This means that 1 hour of EL a week helped students recognize the words they already knew in written form by continuing to match written scripts and audio recordings. In addition, the EL homework was more effective for students with a lower aural vocabulary because the below 1k and 1k groups improved more than the 2k group. Considering the small improvements in aural vocabulary in all groups and the lower vocabulary acquisition of students with higher vocabulary

knowledge, deliberate study of vocabulary is necessary to increase vocabulary and aural vocabulary understanding.

Limitations

The first limitation was that the EL hours are based on students' self-reports. To control this situation, a website or software that can monitor students' actual EL hours should be used in future research. The second limitation was the low reliability of most variables: three out of four variables in the task evaluation questionnaire. To capture students' reactions to EL better, items in these variables should be revised to improve the reliability.

Conclusion

This study investigated students' behavior toward EL as well as aural vocabulary acquisition as a result of EL homework. Although the results were based on three classes of university students (75 students), this study showed useful implications when conducting EL effectively. The results showed that students engaged in EL 1 hour per week, although they were expected to do 2 hours. The task evaluation questionnaire revealed that they felt less intrinsically motivated, less competent to do EL, and less autonomous to choose listening materials. In order to improve this situation, the correlation analysis suggested students be given the freedom to choose interesting listening materials that are at their level to make them feel competent. Nonetheless, EL helped students in all group levels (below 1k, 1k, 2k) acquire aural vocabulary, though the improvements were small, and it was more effective for students who had not reached 2k aural vocabulary knowledge. For vocabulary acquisition, more focused types of learning, such as deliberate vocabulary learning, might be necessary along with EL. When conducting EL homework, we need to consider students' competence and assist them in selecting listening materials from a wide variety of choices so that they feel autonomy. EL itself may not be effective enough for vocabulary learning; thus, the combination of deliberate or intensive learning is important.

Bio Data

Satoko Suzuki is a lecturer at Ibaraki University. She obtained her Ed.D. at Temple University, Japan Campus. Her current research interests include listening, pronunciation, and global English language teaching. <satoko.suzuki.lang@vc.ibaraki.ac.jp>

References

Adolphs, S., & Schmitt, N. (2003). Lexical coverage of spoken discourse. *Applied Linguistics, 24*(4), 425-438. https://doi.org/10.1093/applin/24.4.425

Adolphs, S., & Schmitt, N. (2004). Vocabulary coverage according to spoken discourse context. In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language: Selection, acquisition, and testing* (pp. 39-49). Benjamins.

Agnesia, R. H. (2010). Features affecting task-motivation in English for academic purposes online learning. *Second Language Studies, 29*(1), 1-34. Retrieved from http://www.hawaii.edu/sls/wp-content/uploads/2014/09/Agnesia3.pdf

Brown, H. D., & Lee, H. (2015). *Teaching by principles: An interactive approach to language pedagogy* (4th ed.). Pearson Education.

Brown, R., Waring, R., & Donkaewbua, S. (2008). Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories. *Reading in a Foreign Language, 20,* 136-163. Retrieved from http://www.robwaring.org/papers/various/Brown_Waring_Donkaewbua.pdf

Goh, C. (2000). A cognitive perspective on language learners' listening comprehension problems. *System, 28*, 55-75. https://doi.org/10.1016/S0346-251X(99)00060-3

McLean, S., & Kramer, B. (2016). The development of a Japanese bilingual version of the new vocabulary levels test. *VERB, 5*(1), 2-5. Retrieved from https://brandonkramer.net/pdf/2016_VERB_BilingualNVLT.pdf

McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. *Language Teaching Research, 19*(6), 741-760. https://doi.org/10.1177/1362168814567889

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

Oxford, R. L. (2017). *Teaching and researching language learning strategies: Self-regulation in context* (2nd ed.). Routledge.

Ryan, R. M. (1982). Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory. *Journal of Personality and Social Psychology*, *43*, 450-461. https://doi.org/10.1037/0022-3514.43.3.450

Ryan, R., & Deci, E. (2000). Intrinsic and extrinsic motivations: Class definitions and new directions. *Contemporary Educational Psychology, 25,* 54-67. https://dx.doi.org/10.1006/ceps.1999.1020

Ryan, R. M., & Deci, E. L. (2008). Self-determination theory and the role of basic psychological needs in personality and the organization of behavior. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 654–678). The Guilford Press.

Self-Determination Theory. (n.d.). Intrinsic motivation inventory-complete packet. *Center for Self-Determination Theory.* Retrieved from https://selfdeterminationtheory.org/intrinsic-motivation-inventory/

Suzuki, S. (2019). Factors predicting motivation for and engagement in production tasks among Japanese university students. 茨城大学全学教育機構論集大学教育研究 [Bulletin of Institute for Liberal Arts Education, Ibaraki University, Studies of University Education], 2, 43-55. Retrieved from https://roseibadai.repo.nii.ac.jp/?action=pages_view_main&active_action=repository_view_main_item_detail&item_id=18091&item_no=1&page_id=13&block_id=21

Vandergrift, L. (2005). Relationships among motivation orientations, metacognitive awareness and

proficiency in L2 listening. *Applied Linguistics*, *26*(1), 70-89.

https://doi.org/10.1093/applin/amh039

Vandergrift, L., & Goh, C. M. (2012). *Teaching and learning second language listening: Metacognition in action.* Routledge.

Vidal, K. (2011). A comparison of the effects of reading and listening on incidental vocabulary acquisition. *Language Learning, 61*, 219-258. https://doi.org/10.1111/j.1467-9922.2010.00593.x

van Zeeland, H., & Schmitt, N. (2013a). Incidental vocabulary acquisition through L2 listening: A dimensions approach. *System, 41*, 609-642. https://doi.org/10.1016/j.system.2013.07.012

van Zeeland, H., & Schmitt, N. (2013b). Lexical coverage in L1 and L2 listening comprehension: The same or different from reading comprehension? *Applied Linguistics, 34*(4), 457-479. https://doi.org/10.1093/applin/ams074

Webb, S., & Rodgers, M. P. H. (2009a). The lexical coverage of movies. *Applied Linguistics, 30*(3), 407-427. https://doi.org/10.1093/applin/amp010

Webb, S., & Rodgers, M. P. H. (2009b). The vocabulary demands of television programs. *Language Learning, 59*(2), 335-366. https://doi.org/10.1111/j.1467-9922.2009.00509.x