(Research Article)

Are Language Aptitude and Vocabulary Listening Levels Predictive Variables of TOEIC Listening and Reading Scores?

Brett Collins

Abstract

English as a second language is an important skill to have for university graduates in Japan. Higher standardized test scores on institutional English tests such as the Test of English for International Communication – Listening and Reading (TOEIC) can signal language ability and can increase the appeal of graduates to employers (Bresnihan, 2014). This paper aims to describe two variables, language aptitude and vocabulary listening, as predictors of second language (L2) learner ability measured by a practice TOEIC post-test provided with the textbook, which was given at the last meeting of a TOEIC course. In this study, language aptitude was measured by the Modern Language Aptitude Test - Elementary (MLAT-E), and vocabulary listening was measured using the Listening Vocabulary Levels Test (LVLT), both administered prior to the course. Exposing a connection between these variables might increase focus on implementing aural aspects of curriculum design (e.g., extensive listening input) into second language English learning courses in order to boost TOEIC scores within the university. Scores on the TOEIC Post-Test were correlated with scores from the MLAT-E and LVLT. Correlation analyses using raw scores showed that Hidden Words, which tests aural mapping onto consonants without vowels, and Number Learning, which required participants to learn new sound-number associations within a three-minute span, both correlated with TOEIC Post-Test scores after post hoc corrections were made. However, only one of the LVLT scores showed correlation with TOEIC Post-Test scores, which indicated that the predictive powers of the test were not, in general, very strong.

Key words: second language aptitude, listening vocabulary levels, TOEIC scores

Introduction

English as a second language is an important skill to have for university graduates in Japan. Higher standardized test scores on institutional English tests such as the Test of English for International Communication – Listening and Reading (TOEIC) can signal language ability and can increase the

appeal of graduates to employers (Bresnihan, 2014). Universities are likewise mindful of their graduates' language skills because the Japanese media reports job placement results of students graduating from each university, and the results are a key sales point for schools as they compete for a decreasing pool of candidates; due to the low birthrate, universities are competing for smaller numbers of prospective students (Uehara, 2016). Moreover, as Hongo (2013) stated, stipulating educational guidelines for Japan's future, "the government wants to improve English education at the primary and secondary levels and plans to require that all new government officials take an internationally recognized test of English" (para. 10). Furthermore, Japan's Prime Minister Abe's Revitalization Strategy (2014) claimed that the business world needs workers who are proficient in English, and that this is true not only for larger corporations but also for small- to medium-sized companies in Japan. Also, due to the economic vulnerability in Japan, companies want to recruit graduates who are already proficient in English, rather than to spend money on training them after they are hired. English language ability has become a required skill within the Japanese workforce, and university language programs are on the hook for making sure that graduates can fulfill the requirement.

If higher standardized test scores on institutional English tests are important for graduate employment, then it is important also to look at factors within the university English language curriculum that contribute to success on such tests. This paper aims to describe two variables, language aptitude and vocabulary listening, as predictors of second language (L2) learner ability measured by a practice TOEIC Post-Test provided with the textbook, which was given at the last meeting of a TOEIC course. In the study, language aptitude was measured by the Modern Language Aptitude Test -Elementary (MLAT-E), and vocabulary listening was measured using the Listening Vocabulary Levels Test (LVLT), both administered prior to the course. By showing that language aptitude and listening vocabulary levels were important variables to predict success of listening and reading test scores on the practice TOEIC Post-Test, I aim to highlight the connection between aural processing, which makes up the bulk of both language aptitude and listening vocabulary assessments, and overall language ability (i.e., TOEIC scores). Exposing a connection between these variables might increase focus on implementing aural aspects of curriculum design (e.g., extensive listening input) into second language English learning courses in order to boost TOEIC scores within the university. In the following section, I will discuss background literature for L2 spoken language processing, second language aptitude, and listening vocabulary levels.

Background Literature

This section contains an overview of the background literature for four key variables in the current study. I first give an overview of L2 Spoken Language Processing, which is an underlying proficiency variable. Second, I discuss second language aptitude. Third, I discuss listening vocabulary levels. Fourth, I look at two prior studies on TOEIC test scores in university programs.

L2 Spoken Language Processing

Rost (2016) identified two principle heuristics needed for perception: recognition and categorization. He suggested that listeners should maximize recognition in order to handle the challenges inherent in dynamic speech, such as speech spoken in dialog, and to minimize categorization in order to allow for speaker idiosyncrasies (e.g., utterance rate). Despite the difficulties involved in processing spoken language, Rost argued that the speech signal is redundant and therefore sampling from it, where "only limited extracts of a signal are perceived" (p. 301) is enough for listeners. Cutler and Broersma (2005) hypothesized that listeners' first language (L1) always dictates how other languages are processed, thereby adding more challenges to the role of non-native listeners than simply dealing with a speech stream and speaker differences. However, Rost (2016) has argued that the phonological disconnect between listeners' L1 and L2 can be compensated for by social reasoning and vocabulary networks. Thus, in either an L1 or an L2, listeners need to already have acquired the phonological and lexical content of the speech signal they hear. In other words, if a listener does not understand what they are hearing, they cannot then be successfully assessed on their comprehension of what they heard.

In addition, the target language can create confusion for learners because of its phonological range. Phonological elements of the target language, such as the type of timing of the language, can be conducive or adverse to learning. Limitations for Japanese L2 listeners of English, for example, include making phonological distinctions in isochrony (Bybee, Chakraborti, Jung, & Scheibman, 1998), like interpreting the L2 from the phonological range of a *mora*-based L1 (i.e., the timing of a vowel or consonant-vowel pair in Japanese) or limitations such as word-level difficulties or dialectical changes (Baese-Berk, Bradlow, & Wright, 2013). The dynamic relationship between the phonological range of the L1 and L2 could be described as a *reciprocal causation*, meaning that confusion runs in both directions (Bandura, 2001).

The ability of learners to adjust to an L2's aural makeup is not limited to the ability to understand the semantic information in the target audio. Some languages are stress-based languages (e.g., Indo-European-based), some are tonal languages (e.g., Afro-Asiatic), and some are both stress-based and tonal languages (e.g., Sino-Tibetan). In stress-based languages such as English, stressed and unstressed sounds often demark boundaries in the speech stream. In tone-based languages, tonal placement on

words can change meaning. Learners moving from one language's phonological sound system to another need to come to terms with the differing phonological range of the L2, which is one of many subtle yet powerful contrasts between an L1 and the L2. In the next subsection, I discuss second language aptitude.

Second Language Aptitude

For L2 learners, second language aptitude is generally defined as the ability to learn a second language relative to other learners and conditions. Second language aptitude has been shown to predict second language achievement (Li, 2016) and learning rate in instructed contexts (Granena & Long, 2013); it has also been argued to be significant when learners are untutored, such as when learning occurs in an extensive listening curriculum (Li, 2016). Moreover, Li (2016) suggested that second language aptitude is "a set of abilities of central importance in the preliminary stages of L2 development" (p. 804). The suggestion of preliminary stages hinges on the lack of studies of aptitude in advanced-proficiency L2 learners. Finally, second language aptitude, because of its effect on success in second language acquisition, has been a component in decisions to invest (e.g., money) in learning at all stages of L2 development (Doughty, 2013).

Li (2016) conducted a meta-analysis of second language aptitude research involving a population of 13,035 L2 learners. His study consisted of 66 prior studies of aptitude, including 34 studies that used the MLAT as an instrument. Li used the meta-analyses to investigate how aptitude is related to individual difference variables (e.g., motivation), working memory, and general L2 proficiency. Between-group effects were examined using a fixed-effects model, and within-group effects were examined using a random-effects model. He found that aptitude predicted L2 proficiency, L2 knowledge, and L2 skills, but not vocabulary learning. Moreover, he found that aptitude for phonetic coding was associated with general L2 proficiency, but it was unimportant for listening comprehension. Li's explanation was that meaning takes precedence over form for L2 learners, so the participants relied more on contextual clues and schematic knowledge to compensate for what Juffs and Harrington (2012) described as linguistic deficiency in nonnative listeners. L2 learners can rely on top-down processing more at earlier stages of learning, which suggests that bottom-up processing is weaker and therefore needs more attention in the earlier stages of L2 development. For learners studying for institutional English proficiency tests such as the TOEIC test, the study might suggest that a more foundational approach (e.g., four skills and some grammar) might be better for their proficiency and, therefore, their success on such institutional English proficiency tests, than studying materials designed only for the scope of the English on such tests. In the following section, I discuss listening vocabulary levels as they relate to overall assessed proficiency.

Listening Vocabulary Levels

In contrast to a written vocabulary test, a listening vocabulary test, necessitates knowledge of English phonology, linking, rhythm, and stress patterns (McLean, Kramer, & Beglar, 2015; Rost, 2016), which are prosodic elements of spoken language. Vocabulary acquisition through listening is an important skill given the amount of oral input in communicative language teaching, especially in classes where native teachers teach students in a target language, such as English. In Japan, there is a move made by the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) towards a more communicative approach (MEXT, 2010), underscoring also the importance of investigating how vocabulary can be best developed through oral input. However, this type of on-line phonological processing of lexical forms can place greater demand on working memory, so thorough assessment of learner knowledge of lexical forms necessitates an aural assessment in order to move toward understanding the nature of aural processing and general language proficiency (Trofimovich, 2008).

University TOEIC Test Scores

In the study, TOEIC Post-Test scores are used to measure proficiency predicted by language aptitude scores and vocabulary listening scores. TOEIC scores measure proficiency because test questions are unknown to test takers prior to the test. Using TOEIC scores to gauge proficiency and to place learners into proficiency streams is common at the university level. However, affording TOEIC scores such importance remains unchecked. One study looked at learner progress marked by TOEIC scores. Bresnihan (2014) studied the reasonability for non-English majoring learners at university level in Japan to expect gains on their TOEIC scores. He concluded that, although a practice effect could be expected if participants took more than one TOEIC test, overall TOEIC test scores should not be used to evaluate university students because the test is unrelated to common course content. Moreover, the consistency of low scores on TOEIC tests was demotivating for the test-takers. Robson (2011) studied pre- and post-tests on a four semester TOEIC course intervention. He found that, at the end of the fourth semester, participants showed mixed results. Lower proficiency groups in his study showed more improvement than the higher proficiency groups, and participants mainly improved their reading and not their listening scores. Robson concluded that the reading done by participants preparing for listening and speaking activities might have been a boost to their TOEIC reading scores.

The current study aims to find learner variables that might mark success on a TOEIC. By finding such markers, it is perhaps possible to tailor classroom language input to better serve the learners.

Research Questions for the current study were as follows:

- 1. Do mean MLAT-E scores have significant correlation with mean scores on the TOEIC Post-Test for participants?
- 2. Do mean LVLT scores have significant correlation with mean scores on the TOEIC Post-Test for participants?

This is a post-test only quasi-experimental design using multiple measures. In the study, the independent variables are (second) language aptitude and vocabulary listening level. The dependent variable is the Post-Test scores (i.e., Listening and Reading) on the TOEIC test. In the next section, I will outline the methodology used in the current study. I will discuss the educational setting, participants, instrumentation, including the MLAT-E and LVLT.

Methodology

In this section I discuss the methodology used in the current study. I first outline the Educational Setting, including the type of class the participants were taking. Next, I discuss the Participants. I then outline the Instrumentation used in the current study, which include the TOEIC, the MLAT-E, and the LVLT.

Educational Setting

The setting for the study was a private, coeducational university located in Chiba Japan, specifically an intensive Test of English for International Communication – Listening and Reading (TOEIC) course that was part of a four-year, tourism curriculum. The students in the tourism program were expected to graduate from the program and go to work in the tourism field (e.g., as front desk clerks in hotels). The course was a two-credit elective. The course met for six hours, once per week, for eight and a half weeks (i.e., for one quarter). At the first meeting, students were given a TOEIC pre-test, a listening vocabulary test, a language aptitude test, and were introduced to the course and textbook. From the second to the eighth meeting, a quiz was administered to assess the learners' comprehension of the textbook content and more specifically the vocabulary taught in the previous meeting. Next, regular lessons were given in accordance to the textbook. The quizzes, along with the Pre- and Post-test scores constituted the course grades. All the students successfully passed the course based on these grades.

The class was an intermediate level TOEIC course. Participants included in the current study had successfully taken a beginner-level course, which had a score-target of 400, and had received a minimum letter grade of B (i.e., a de facto proficiency prerequisite was used to gatekeep participation). Also, participants used the TOEIC textbook *Winning Formula for the TOEIC Listening and Reading Test* (WFT), by Akaida and Bruce (2018) as the content of an eight-week intensive course and had taken both a TOEIC-like Pre-Test and the Post-Test supplied by the textbook manufacturer National Geographic Learning. The textbook had 12 units with themes from TOEIC tests, for example, shopping. Within each unit, activities were divided into sections labeled Vocabulary Building, Grammar, and Tactics. The Vocabulary Building section focused on commonly used words. The Grammar section focused on using grammatical points in a variety of ways. The Tactics section focused on listening and

reading tactics, and further elaborated on the seven different parts of TOEIC tests, which are (1) Photographs, (2) Question-Response, (3) Conversations, and (4) Talks in the Listening section, and (5) Incomplete Sentences, (6) Text Completion, and (7) Reading Comprehension in the Reading Section. The listening tactics section gave advice for kinds of information that test takers should listen for during the test and included warnings for common problems that might be encountered. The reading tactics section introduced the idea of time management and question types.

Participants

The participants' (N = 30) ages ranged from 18 to 23 (M = 19.6). They were selected initially by having enrolled in a TOEIC for Careers 600 course and included in the study by completing the requirements of the course. The participants had from 6 to 10 years (M = 7.2) of English education, with the majority having a minimum of six years of formal English education in secondary school (junior high and high school) and university. The participants were from three different countries, Japan (10), China (8), and Vietnam (12). Participants included first- through fourth-year students, five of whom were returning to their countries to begin their careers in tourism and needed a TOEIC score to apply for work. Only four of the 30 students had taken an official TOEIC exam prior to the course. My own sense of the group was that they were varied in their proficiency from intermediate to advanced levels.

Instrumentation

Three instruments were used in the current study to gauge aspects of language proficiency in terms of aural processing. The Modern Language Aptitude Test – Elementary (MLAT-E) was used to assess the participants' English language aptitude. Although language aptitude is different from proficiency, aptitude has been shown to predict general L2 proficiency (Li, 2016). A Listening Vocabulary Levels Test (LVLT) was used to test the participants' knowledge of the first five 1000-word frequency levels and the Academic Word List (AWL) (Coxhead, 2000) in English. Finally, a Test of English for International Communication (TOEIC) Pre- and Post-Test were administered to gauge general language ability at the beginning and end of the course. The MLAT-E and the LVLT were validated by converting the item results to Rasch logits (Collins, 2018). The TOEIC Pre- and Post-Tests were not validated.

Modern Language Aptitude Test-Elementary

The results of the MLAT-E (Carroll & Sapon, 2010) were used to assess the participants' language aptitude (i.e., phonemic coding). Phonemic coding ability concerns the effective auditory processing of input (Skehan, 2002) and allows the participants to analyze and code auditory material for the purpose of online retention. Chan, Skehan, and Gong (2011) defined phonemic coding ability as "the

ability to analyze sound so that it can be retained for more than a few seconds" (p. 56). According to Skehan (2002), MLAT-E scores can estimate aptitude for foreign language learning.

One disadvantage to using the MLAT-E was that the directions were given in the L2 instead of the participants' L1, which presented a possible confound to the participants' L2 proficiency (Doughty, 2013). However, instructions and tasks on the MLAT-E were worded simply, as they were designed for elementary school-aged learners. Instructions were broken into parts and practice examples were given for each part. For example, the Matching Words section of the MLAT-E had 25 examples and eight practice tasks in the instructions before the participants attempted the actual task. In addition, prior to each section, the proctor checked with the participants to be sure that they had understood the tasks.

The MLAT-E consisted of four parts designed to assess the test-takers' ability to associate sounds and symbols, sensitivity to grammatical structure, ability to hear speech sounds, auditory alertness, and ability to remember. Participants' raw scores were used. Part 1 of the test was labeled Hidden Words and was designed to measure English vocabulary knowledge and sound-symbol association ability. An example of a test item from the Hidden Words section of the test was as follows:

rivr



| a part of the body |
|--------------------|
| a dog's name |

For this item, the participants chose and marked the box next to the word or group of words that mean the same as the word representation on the left. For the above example, the correct answer is large stream of water, as it is a definition of river (rivr). There were 30 items in this part of the test. Part 2 of the test was labeled Matching Words and was designed to test the participants' understanding of target word usage in example sentences. An example Matching Words test item is as follows:

Henry THREW the heavy stone.

| Sally | rides a b | oicycle. |
|-------|-----------|----------|
| | | |

For this item, participants chose and marked the box under the word in the sentence that had the same function as the word in all capital letters in the sentence directly above it. For the above example, the correct answer was *rides* because it is a transitive verb, just as *threw* is a transitive verb. There were 30 items in this part of the test. Part 3 of the test was labeled Finding Rhymes and was designed to test the participants' ability to match speech sounds by selecting words that rhyme. An example of an item was as follows:

DOOR..... car..... four..... mayor..... our

For this item, participants chose and marked the box next to the word that rhymes with the word on the left. For the above example, the correct answer is *four*, as it rhymes with *door*. This type of question tested the participants' familiarity with the English sound system. There were 45 items in this part of the test. Part 4 of the test was labeled Number Learning and asked the participants to learn nonce names for numbers and then recognize the nonce names aurally. The participants wrote the nonce numbers they heard. An example of an item on the test is *chall* [\mathfrak{f} al] (spelling and transcription is mine), which was taught as equal to the number one. Participants heard the word and wrote the number that the word represented. There were six base numbers, 1, 2, 3, 10, 20, and 30. There were 25 items in this part of the test, and all but four numbers, 3, 30, 1, 31, and 10, were repeated twice.

Listening Vocabulary Levels Test

The LVLT (McLean, Kramer, & Beglar, 2015) was designed to assess knowledge of the first five 1,000-word frequency levels and the Academic Word List (Coxhead, 2000) (Appendix A). An example test item from the LVLT, translated from Japanese, is as follows:

A test taker hears:

(Man)

Waited. I waited for a bus.

The test taker reads in Japanese (L1):

| A. 食べた | (translation: <i>ate</i>) |
|--------|-------------------------------|
| B. 待った | (translation: <i>waited</i>) |
| C. 見た | (translation: <i>saw</i>) |
| D. 寝た | (translation: <i>slept</i>) |

The results of the LVLT were used to assess the participants' readiness to handle the lexical load of the aural targets. McLean et al. (2015) set comprehension of aural texts made up of the first through fifth 1,000 high frequency words of English at 23 - 24 correct answers out of 25 items. However, no benchmark was used as a minimum cut-off score for participation because frequency levels are not revealed by the Educational Testing Service (ETS) for the TOEIC test. In other words, lexical frequency levels could not be controlled in one instrument, so they were not controlled in either instrument. All six sections of the test were administered to the participants, with raw scores from each of the sections used in the analyses.

TOEIC

The results of the TOEIC Pre- and Post-Tests (Akaida & Bruce, 2018) were used to assess the participants' language proficiency. According to the ETS the TOEIC Listening and Reading test

consists of two, timed sections of 100 questions each. The test takes around 2.5 hours to complete. Within the total time, the Listening section (Section 1) is 45 minutes, and the Reading section (Section 2) is 75 minutes. Another 30 minutes is provided for participants to answer biographical questions (ETS.org). The specific parts of each section mirror the sections of the textbook used in the study, which were mentioned above (see Educational Setting), so they will not be written here. The TOEIC, a norm-referenced test (NRT), is designed to classify students into groups of proficiencies (Brown, 2005). Because test takers do not know exactly what is on the test before taking it, the test can arguably measure overall proficiency. Test takers' scores are measured on a percentile, instead of as raw scores, which places students on a continuum of scores of all the test takers who took a unique TOEIC test form. Moreover, the percentile of individual test takers is to a curve, meaning that, as percentile scores of test takers move away from the mean, the number of test takers attaining these scores is smaller (Robson, 2011). The TOEIC standard error of measurement is 25 for both the reading and listening sections (ETS.org).

In the following section I look at the descriptive statistics and the correlation coefficient computation results from the dependent and independent variables. Results are given for descriptive statistics and the correlation coefficients.

Results

This section looks at the results of analyses done for the two research questions. First, descriptive statistics are given for the MLAT-E, LVLT, Pre-Test, Quiz scores, and Post-Test scores. Next, correlation coefficients were computed among the four aptitude scales from the MLAT-E (i.e., Hidden Words, Matching Words, Finding Rhymes, and Number Learning), the LVLT frequency levels (i.e., First-1000, Second-1000, Third-1000, Fourth-1000, Fifth-1000, and Academic Word List), the Pre-and Post-Test, and the course quiz scores.

Table 1 shows the descriptive statistics for items from the four MLAT-E aptitude scales. Although mean statistics vary slightly, the total number of items for each of the aptitude scales were different. Hidden Words consisted of 30 items. Matching Words consisted of 30 items. Finding Rhymes consisted of 45 items. Number Learning consisted of 25 items. In relation to the total number of items for each aptitude scales, the mean for Matching Words (M = 23.00, SD = 4.06) was greatest at .77 of the total points possible. The mean for Finding Rhymes (M = 23.26, SD = 7.48) was smallest at .52 of the total points possible.

| | Hidden Words | Matching Words | Finding Rhymes | Number Learning |
|----------|----------------|----------------|----------------|-----------------|
| М | 18.17 | 23.00 | 23.26 | 15.74 |
| SE | 1.13 | .85 | 1.56 | 1.76 |
| 95% CI | [15.83, 20.52] | [21.25, 24.75] | [20.02,26.50] | [12.09,19.39] |
| SD | 5.42 | 4.06 | 7.48 | 8.45 |
| Skewness | .48 | -1.48 | .42 | 51 |
| SES | .48 | .48 | .48 | .48 |
| Kurtosis | .21 | 4.14 | 62 | 72 |
| SEK | .94 | .96 | .94 | .94 |

Table 1. Descriptive Statistics for MLAT-E Items (N = 30)

Note. All statistics were based on raw scores.

Table 2 shows the descriptive statistics for items from the six frequency levels of the LVLT. Each 1000-word frequency sub-test consisted of 24 items, and the AWL sub-test consisted of 30 items. Differences in subsequent 1000-word frequency levels can be seen in between the first 1000-word frequency level (M = 22.52, SD = 1.68) and the fifth level (M = 13.91, SD = 3.25). The AWL sub-test (M = 17.83, SD = 8.45) had the greatest standard deviation among participant scores.

Table 3 shows the descriptive statistics for assessments within the course. The Pre-Test and Post-Test scores were divided into section scores for listening and reading. Both Pre- and Post-Test Listening and Pre- and Post-Test Reading sections are divided evenly (i.e., 25 questions for each section in the Pre-Test and 50 total, and 50 questions for each section in the Post-Test and 100 total). The Quizzes used for the course are included in Table 3 and given as Quiz Totals; quiz scores are not a variable tested in this study but do provide insight into the participants' general abilities within the TOEIC course. There were six quizzes given in the course with a communitive point total of 210 points. The descriptive statistics for Post-Test Listening (M = 26.57, SD = 9.14) and Reading (M = 25.09, SD = 9.01) show greater mean scores when compared to the Pre-Test Listening and Reading means, though both Post-Test means had high standard deviations.

| Table 2. Des | criptive Statistics for | r LVLT Items (N = 3 | 0) | | | |
|--------------|------------------------------|--------------------------------|------------------------------|--------------------------------|------------------------------|-----------------------|
| | First 1000-Word Frequency | Second 1000- Word Frequency | Third 1000-Word Frequency | Fourth 1000- Word Frequency | Fifth 1000-Word Frequency | Academic Word List |
| W | 22.52 | 17.78 | 14.61 | 13.00 | 13.91 | 17.83 |
| SE | .35 | .84 | .76 | .76 | 68 | 1.03 |
| 95% CI | [21.80, 23.25] | [16.04,19.53] | [13.03,16.19] | [11.39,14.61] | [12.51,15.32] | [15.69,19.96] |
| SD | 1.68 | 4.03 | 3.65 | 3.73 | 3.25 | 4.94 |
| Skewness | -2.08 | 91 | 56 | 49 | .86 | 14 |
| SES | .48 | .48 | .48 | .48 | .48 | .48 |
| Kurtosis | 5.03 | .16 | .11 | 2.29 | 1.70 | 14 |
| SEK | .94 | .94 | .94 | .94 | .94 | .94 |

Note. All statistics were based on raw scores.

| | Pre-Test Listening | Pre-Test Reading | Quizzes Total | Post-Test Listening | Post-Test Reading |
|----------|-----------------------|---------------------|---------------------|------------------------|----------------------|
| М | 9.96 | 12.87 | 135.10 | 26.57 | 25.09 |
| SE | .88 | .94 | 8.07 | 1.91 | 1.89 |
| 95% CI | [8.12, 11.79] | [10.93, 14.81] | [118.34, 151.83] | [22.61, 30.52] | [21.19, 28.98] |
| SD | 4.24 | 4.50 | 38.72 | 9.14 | 9.01 |
| Skewness | .88 | .30 | 51 | .42 | .43 |
| SES | .48 | .48 | .48 | .48 | .48 |
| Kurtosis | 1.11 | 63 | .19 | -1.08 | 91 |
| SEK | .94 | .94 | .94 | .94 | .94 |

Table 3. Descriptive Statistics for Assessments Used in the TOEIC Course (N = 30)

Note. All statistics were based on raw scores.

Next, correlation coefficients were computed among the four aptitude scales from the MLAT-E (i.e., Hidden Words, Matching Words, Finding Rhymes, and Number Learning), the LVLT (i.e., first through fifth 1000-word frequency and AWL), and the TOEIC Post-Test sections (i.e., Listening and Reading). Raw scores from each of the four sections of the MLAT-E, the six sections of the LVLT, and the two sections of the TOEIC Post-Test were used in the analysis. *Post hoc* correction was then calculated to correct the level of significance for each test.

First, using the Bonferroni approach to control for Type I error across the 6 variables of the four MLAT-E aptitude scales and the two TOEIC Post-Test scores, a *p*-value of less than .008 ($.05 \div 6 = .008$) was required for significance. The results of the correlational analyses presented in Table 4 show that 4 out of 15 correlations were statistically significant and were greater than .32. The correlations of Number Learning were significant for both TOEIC Post-Test scores. In general, the results with post hoc correction indicate that language aptitude for Matching Words, Finding Rhymes, and Hidden Words did not correlate with the TOEIC Post-Test scores.

Second, using the Bonferroni approach to control for Type I error across the 8 variables of the six LVLT scales and the two TOEIC Post-Test scores, a *p*-value of less than .006 ($.05 \div 8 = .006$) was required for significance. The results of the correlational analyses presented in Table 5 show that 2 out of 28 correlations were statistically significant and were greater than .43. The correlations of the fourth 1000-word frequency level were significant for the TOEIC Post-Test Reading scores. In general, the

results with *post hoc* correction indicate that vocabulary listening levels did not strongly correlate with the TOEIC Post-Test scores. The slight correlation of the fourth 1000-word frequency level may have indicated a lexical link to the TOEIC test (i.e., an item that was used in both tests). In the next section I discuss the results of the analyses in relation to the research questions.

| Table 4. Correlations Among th | e MLAT-E and | TOEIC Post-To | est Scores (N = | 30) | | |
|--------------------------------|--------------|-------------------|-------------------|--------------------|-------------------------------|-----------------------------|
| Hid | dden Words | Matching Words | Finding Rhymes | Number Learning | TOEIC Post- Test Listening | TOEIC Post- Test Reading |
| Hidden Words | I | | | | | |
| Matching Words | 20 | I | | | | |
| Finding Rhymes | .51 | .34 | I | | | |
| Number Learning | .16 | .29 | .29 | I | | |
| TOEIC Post-Test Listening | | 90. | .41 | .32* | I | |
| TOEIC Post-Test Reading | .48* | 14 | .51 | .59* | .89* | I |
| *p < .008 | | | | | | |

| |) | | | | | | | |
|-------------------------------|----------------------------------|----------------------------------|----------------------------------|-----------------------------------|----------------------------------|-----------------------|----------------------------------|--------------------------------|
| | First 1000- Word Frequency | Second 1000-Word Frequency | Third 1000- Word Frequency | Fourth 1000- Word Frequency | Fifth 1000- Word Frequency | Academic Word List | TOEIC Post- Test Listening | TOEIC Post- Test Reading |
| First 1000-Word Frequency | | | | | | | | |
| Second 1000-Word Frequency | 12 | I | | | | | | |
| Third 1000-Word Frequency | .86 | .54 | I | | | | | |
| Fourth 1000-Word Frequency | .68 | .55 | .70* | I | | | | |
| Fifth 1000-Word Frequency | .39 | .24 | .31 | .32 | | | | |
| Academic Word List | 14 | .24 | .40 | .15 | .56 | I | | |
| TOEIC Post-Test Listening | 12 | .11 | .21 | .36 | .37 | .31 | I | |
| TOEIC Post-Test Reading | 07 | .15 | .22 | .43* | .34 | 80. | 68. | I |
| | | | | | | | | |

Table 5. Correlations Among the LVLT and TOEIC Post-Test Scores (N = 30)

*p < .006

Discussion

In this section I discuss the results of the analyses in relation to the research questions. I then discuss possible reasons for the results and problems with the current study. Finally, I make suggestions for further research.

Research Question 1 asked if mean MLAT-E scores have significant correlation with mean scores on the TOEIC Post-Test for participants. The MLAT-E scores and TOEIC Post-Test correlations of Hidden Words and Number Learning were significant for both TOEIC Post-Test scores, Reading and Listening. Hidden Words tests aural mapping onto consonants without vowels, meaning that success is a measurement of lexical representations of sound. Number Learning required participants to learn new sound-number associations within a three-minute span, and perhaps this process is like the expedient processing required on the TOEIC. Both scales required advanced meta-knowledge of aural processing. In general, the results with *post hoc* correction indicate that language aptitude for Matching Words and Finding Rhymes did not correlate with the TOEIC Post-Test scores.

Research Question 2 asked if mean LVLT scores have significant correlation with mean scores on the TOEIC Post-Test for participants. The fourth 1000-word frequency level were significant for the TOEIC Post-Test Reading scores. Perhaps lexical items in the fourth 1000-word frequency level on the LVLT were similar to the lexical items within the Reading section of the TOEIC instrument used in the current study, and those participants who had the lexical items were able to use them to answer TOEIC questions. In general, the results with *post hoc* correction indicate that vocabulary listening levels did not strongly correlate with the TOEIC Post-Test scores. Perhaps the items on the LVLT were not related to the lexical requirements of the TOEIC Post-Test.

The results indicate a small predictive ability in the case of the MLAT-E, but only a weak predictive ability for the LVLT on this type of TOEIC post-test. However, both aural processing ability of the type required on the MLAT-E, and lexical knowledge of the type needed for success on the LVLT are also needed for success on any TOEIC. The results of the current study do not suggest that language aptitude and lexical knowledge are not necessary or do not correlate with TOEIC success. The results indicate that they do not warrant the label of strongly predictive variables in this case.

Because the current study was able link some aspects of TOIEC and language aptitude which required a strong sense of aural processing, classroom input in TOEIC classes might benefit from aural-focused exercises. For example, a simple exercise that might enhance learner awareness is to have them spell target vocabulary words. By spelling target items, learners make connections when rendering speech sounds into words.

Several drawbacks to the design of this study were *N*-size and the use of a mock TOEIC. First, the *N*-size was too small to gain optimum understanding of how language aptitude and vocabulary listening

levels might relate to TOEIC scores. Second, the TOEIC Post-Test was not an official TOEIC test, so the results cannot be said to predict an official TOEIC test. However, the test included the seven parts of an official TOEIC test, and the test questions were culled from past official tests.

Future research should investigate L1 differences in relation to TOEIC scores after participants complete a course. In the current study, the participants were from three different countries, but these differences were not used as variables within the analyses. Analyzing L1 differences with language aptitude and listening vocabulary levels might reveal significant differences in language processing.

Conclusion

This paper aimed to describe two variables, language aptitude and vocabulary listening, as predictors of L2 learner ability measured by a practice TOEIC Post-Test provided with a textbook, which was given at the last meeting of a TOEIC course. In the current study, language aptitude was measured by the MLAT-E, and vocabulary listening was measured using the LVLT, both administered prior to the course. By examining whether language aptitude and listening vocabulary levels were important variables to predict success of listening and reading test scores on the practice TOEIC Post-Test, I aimed to highlight the connection between aural processing, which makes up the bulk of both language aptitude and listening vocabulary assessments, and overall language ability. Exposing a connection between these variables might increase focus on implementing aural aspects of curriculum design (e.g., extensive listening input) into second language English learning courses in order to boost TOEIC scores within the university. Results of correlation coefficients showed that scores on the TOEIC Post-Test were correlated with scores from the MLAT-E and LVLT. Correlation analyses using raw scores showed that Hidden Words, which tests aural mapping onto consonants without vowels, and Number Learning, which required participants to learn new sound-number associations within a three-minute span, both correlated with Post-Test TOEIC scores after post hoc corrections were made. However, only one of the LVLT scores showed correlation with Post-Test TOEIC scores, which indicated that the predictive powers of the test were not, in general, very strong.

References

Akaida, T., & Bruce, J. (2018). Winning Formula for the TOEIC L&R Test (Revised Edition). Tokyo: Cengage Learning.
Baese-Berk, M., Bradlow, A., & Wright, B. (2013). Accent-independent adaptation to foreign accented speech. The Journal of the Acoustical Society of America, 133(3), 174-180. doi:10.1121/1.4789864

Bandura, A. (2001). Social cognitive theory: An Agentic perspective. Annual Review of Psychology, 52(1), 1-26. doi:10.1146/annurev.psych.52.1.1

- Bresnihan, B. (2014). Should Japanese university students' TOEIC scores be expected to increase? 人文論集 /Journal of Cultural Science, 53, 1-10.
- Brown, J. D. (2005). *Testing in Language Programs: A Comprehensive Guide to English Language Assessment*. New York: McGraw Hill.
- Bybee, J., Chakraborti, P., Jung, D., & Scheibman, J. (1998). Prosody and segmental effect: Some paths of evolution for word stress. *Studies in Language*, 22, 267-314. doi:10.1075/sl.22.2.02byb
- Carroll, J. B., & Sapon, S. (2010). *Modern Language Aptitude Test Elementary: Manual, 2010 Edition*. Rockville, MD: Second Language Testing.
- Coxhead, A. (2000). A new Academic word list. TESOL Quarterly, 34, 213-238. doi:10.2307/3587951
- Chan, E., Skehan, P., & Gong, G. (2011). Working memory, phonemic coding ability and foreign language aptitude: Potential for construction of specific language aptitude tests—the case of Cantonese. *Ilha do Desterro, A Journal of English Language, Literatures in English and Cultural Studies, 60*, 45-73. doi:10.5007/2175-8026.2011n60p045
- Collins, B. (2018). *Sandhi-variation and the comprehension of spoken English for Japanese learners* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 13294)
- Cutler, A., & Broersma, M. (2005). *Phonetic precision in listening*. In W. J. Hardcastle, & J. M. Beck (Eds.), A figure of speech: A festschrift for John Laver (pp. 63-91). Mahwah, NJ: Erlbaum.
- Doughty, C. (2013). *Assessing aptitude*. In A. Kunnan (Ed.) The companion to language assessment (pp. 23-46). Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781118411360
- Educational Testing Service. (2015). The research foundation for the TOEIC® tests: A compendium of studies. Princeton, NJ: Author.
- Granena, G, & Long, M. (2013). Age of onset, length of residence, aptitude and ultimate L2 attainment in three linguistic domains. *Second Language Research*, 29, 311-43.
- Hongo, J. (July 2013). *Japan Ranks 40th of 48 Countries in TOEIC Scores*. The Wall Street Journal Online: http://on.wsj.com/1wCt9ig.
- Juffs, A., & Harrington, M. (2012). Aspects of working memory in L2 learning. *Language Teaching*, 44(2), 137-166. doi:10.1017/S0261444810000509
- Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language* Acquisition, 38(4), 801-842. doi:10.1017/S027226311500042X
- McLean, S., Kramer, B., & Beglar, D. (2015). The creation and validation of a listening vocabulary levels test. Language Teaching Research,19(6), 741-760. doi:10.1177/1362168814567889
- MEXT. (2010). 高等学校学習指導要領解説:外国語編・英語編 [The guide to course of study: Foreign language and English]. Tokyo: Kairyudo.
- Robson, G. (2011). Analysis of TOEIC scores of tourism department students 2009-2010. *Journal of Tourism Studies*, 10, 113-126.

Rost, M. (2016). Teaching and researching listening. London, England: Longman. doi:10.4324/9781315833705

- Skehan, P. (2002). Theorizing and updating aptitude. In P. Robinson (Ed.), Individual differences and instructed language learning (pp. 69-93). Amsterdam, The Netherlands: Benjamins. doi:10.1075/lllt.2
- Trofimovich, P. (2008). What do second language listeners know about spoken words? Effects of experience and attention in spoken word processing. *Journal of Psycholinguistic Research*, 37, 309-329. doi:10.1007/s10936-008-9069-z
- Uehara, Y. (2016). How Japanese students hunt for jobs. Nippon.com. Retrieved October 1, 2019.