

## **The Effects of Study Abroad Duration and Predeparture Proficiency on the L2 Proficiency of Japanese University Students: A Meta-Analysis Approach**

Akiyo HIRAI  
*University of Tsukuba*

### **Abstract**

The present study aims to clarify the effects of study abroad (SA) duration and predeparture proficiency on the second language (L2) progress of Japanese students of English. As a first step toward this goal, studies on SA of one month or less (short-term), of more than one month to less than six months (middle-term), and of six months or more (long-term) were reviewed extensively. Next, 31 studies, all of which reported SA students' pre- and post-test scores, were selected, and effect sizes of the students' L2 gains were generated to allow for further comparisons among the three lengths of SA and among three proficiency levels based on their pre-test scores that were carried out by means of a meta-analysis method. The results showed that the magnitude of the effect of long-term SA was more than twice as great as that of middle-term SA and more than four times as great as that of short-term SA. The second factor analyzed in this study, students' predeparture proficiency, did not seem to be an influential predictor of L2 gains. However, further analysis revealed that there was an interaction between the two factors, and low proficiency students tended to attend shorter-term SA programs.

**Keywords:** meta-analysis, study abroad duration, predeparture proficiency, effect size, L2 proficiency

In this era of globalization, the number of young people studying abroad has steadily increased; this is due to the belief that the study abroad (SA) experience has positive effects on various aspects of intellectual and personal growth. In fact, many studies report improvement of second language (L2) proficiency as well as particular psychological enhancements such as willingness to talk and self-confidence (e.g., Furuya, 2005; Iida, 2013). However, the degree of L2 improvement corresponding to the length of SA and predeparture proficiency level remains unclear. This is because administering pre- and post-tests to students participating in SA programs of various lengths and keeping track of them during their respective SA periods is difficult within the scope of a single study. It is also difficult to compare studies on a single length of SA since in each study, participants at different proficiency levels have studied abroad under different conditions, and their L2 improvements have been reported in different ways.

Thus, the present study adopts a method of meta-analysis, i.e., a quantitative statistical method of synthesizing relevant studies using a common metric, and thereby makes the comparison of L2 gains carried out in these studies possible. Since meta-analysis requires a significant number of studies, the author first undertook an extensive review of relevant literature not only to select a large body studies for meta-analysis but also to uncover attributes of length of SA and predeparture L2 proficiency which cannot be found by the meta-analysis method alone.

## Literature Review

### Japanese Study Abroad Students

In recent years, the Japanese government has encouraged more students to study abroad partly because Japanese companies wishing to strengthen their overseas operations are struggling to recruit a sufficient number of workers with the language abilities and international experience they require. To help address this shortfall, the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) launched a financial support plan called *Tobitate!* (Leap for Tomorrow) to increase the number of Japanese students studying abroad (MEXT, 2017). Thanks to support from the government and the private sector, there has been a growing interest in studying abroad, particularly among Japanese university students. A survey published by the Japan Student Services Organization (JASSO) shows that the number of Japanese students studying abroad has been steadily increasing (see Figure 1).

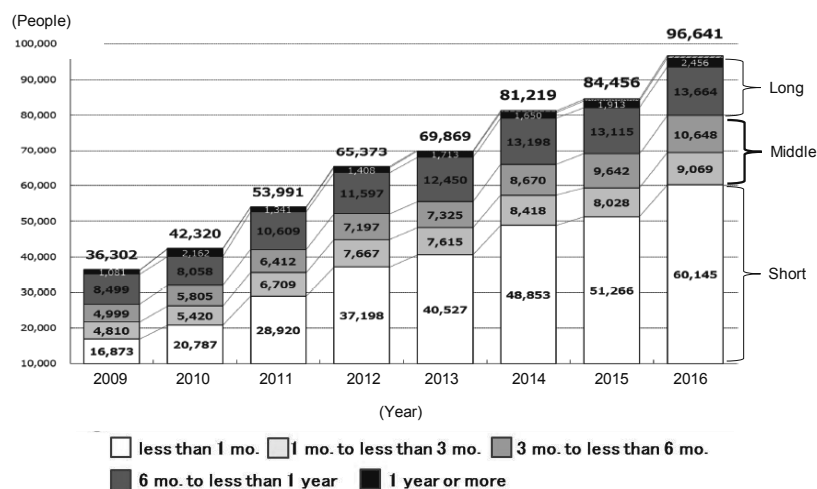


Figure 1. The number of Japanese students who studied abroad from 2009 to 2016 (MEXT, 2017). The segmentation terms on the right side were added by the author of this study.

Regarding the categorization of SA duration, the Institute of International Education (2016), a U.S. organization, categorizes one summer program, or eight weeks or less, as

short-term; one or two quarters, or one semester, as middle-term; and an academic or calendar year as long-term. However, as shown in Figure 1, the JASSO survey shows five lengths of SA. Among them, the most popular SA programs with Japanese university students are short-term programs lasting less than one month. Considering the current interest in one-month SA programs, the author classified one month or less as short-term, more than one month to less than six months as middle-term, and six months to one year as long-term (see the segmentations added to the right side of Figure 1).

### **Study Abroad Duration**

The correlation of SA duration with L2 gains has not been much investigated (e.g., Baró & Serrano, 2011; Hoffman-Hicks, 2000; Ife, Vives, & Meara, 2000; Sasaki, 2011; Seki, Nonaka, & Sumida, 2009). Since most SA studies focus on the effects of a single duration, let us review the studies according to the durations of SA as defined above, including studies directly comparing multiple SA durations.

**Research on short-term study abroad.** Most short-term SA studies reported student listening gains (e.g., Joto & Numoto, 1989; Kimura, 2006, 2011; Suzuki & Hayashi, 2014). Kimura (2006) investigated the effects of short-term SA on the English proficiency of 19 Japanese university students in comparison with that of a control group of 48 students who did not study abroad. The participants stayed in New Zealand for three weeks. Their English proficiency was measured before and after they went abroad by means of a test that consisted of EIKEN 3rd- and pre-2nd-grade test items. The results revealed that participants made significant gains on their listening and total test scores and outperformed the control group. However, their scores on the grammar, reading, and composition sections neither significantly improved nor excelled those of the control group. Kimura (2011) confirmed the same tendency in her study (i.e., Kimura, 2006) with another group of 14 participants. Besides improvement of listening skills, she found improvement of participants' writing fluency (not accuracy) in just three weeks.

On the other hand, studies demonstrating participants' L2 gains on reading or grammar sections of assessment instruments were few. Among them, Kobayashi (1999), using the TOEFL ITP for 31 students who stayed abroad for about three weeks, reported score gains on all three test sections (listening, vocabulary & grammar, and reading). This favorable result may be attributable to the relatively stringent selection of SA participants and the grading system. Participants received grades based on their post-test scores; this arrangement may have given them a strong motivation and clear purpose while studying abroad.

However, Nonaka (2008) cautioned that results may differ depending on which English proficiency test is used to measure students' L2 improvement. Nonaka (2008) found a significant gain in students' listening skills when he used the TOEIC IP, while his previous study (Nonaka, 2005) did not show any L2 gains of participants on any sections of the TOEFL ITP, even though students in both studies had similar proficiency levels and participated in similar SA programs. To address this point, a meta-analysis can synthesize a

large number of studies that use various L2 proficiency tests and thus reduce the differences attributable to L2 proficiency measurement.

Overall, students in short-term SA programs are likely to improve their listening skills but are unlikely to improve other skills. However, the degree of L2 gains seems to be small or non-significant. Kuno (2011), comparing the English scores of students who studied abroad and students who studied in Japan, found the same level of L2 progress for both groups.

**Research on middle-term study abroad.** The lengths of middle-term SA programs reported in the studies varied from five weeks to five months. Tanaka and Ellis (2003) administered the TOEFL ITP test to 116 Japanese university students who participated in a 15-week SA program and found significant gains for all three sections of the test. Le Pavoux and Tsuchihira (1998) investigated students who studied in New Zealand for 15 weeks and observed significant L2 gains on both the listening and reading sections of the TOEIC IP. The results suggest that not only listening but also other receptive skills and linguistic knowledge may progress, although listening may be the first skill to improve.

Related to this point, Iwakiri (1993) compared an experimental group of 96 students who participated in a five-week program in Australia with a control group of the same number of students who did not study abroad. The results showed that the experimental group's TOEFL ITP scores on the listening and structure sections were significantly higher than those of the control group, although the scores on the reading section were similar. Furthermore, Chiba (2005) administered the TOEIC Bridge to students attending an eight-week program and only found significant improvement of their listening scores.

Thus, according to the results of the short-term SA studies herein reviewed, middle-term SA students tended to significantly improve at least their listening skills. This is reasonable if we consider the environmental necessities and the natural development of language acquisition when living in another country. In English-medium schools and in students' off-campus lives, the amount of phonetic input usually exceeds the amount of written input, and aural comprehension is of the utmost importance in communicating with others. Another reason for the significant gains made in listening could be that conversation, rather than reading and writing, may be emphasized in classes in SA programs (Muta, 2007).

Some middle-term SA studies focusing on productive skills reported improvement in students' oral fluency, noting fewer and shorter pauses and faster speech rates (e.g., Iwakiri, 1993; Segalowitz & Freed, 2004). For example, Iwakiri (1993) conducted interview and role-play tests with students who studied for five weeks and found improvement in their oral fluency. However, unlike oral fluency, accuracy and complexity seem to be difficult to improve in two or three months, as reported by Baró and Serrano (2011), who examined the oral production of Spanish students of English.

**Research on long-term study abroad.** Due to the difficulty of keeping track of participants for six months to one year and of conducting pre- and post-tests with them, only a few studies met the long-term criteria. Seki et al. (2009) investigated three-month ( $n = 6$ ) and six-month ( $n = 16$ ) SA programs, using the TOEIC IP test. They found that L2 gains for participants in the six-month program were significantly greater than gains obtained by

participants in the three-month program. Furthermore, the follow-up investigation found that those who had completed the long-term SA program showed greater memory retention and an increased tendency to study English independently after returning from abroad. Therefore, in addition to the expectation of greater L2 improvement, long-term SA programs appear to have positive cognitive and psychological effects.

Positive effects were also pointed out by other long-term SA studies. Hoffman-Hicks (2000) reported that students' pragmatic abilities in L2 French improved significantly after two or three months abroad and continued to improve for 10 months. Sasaki (2011, 2018) focused on the effect of the length of SA by measuring Japanese students' English proficiency and writing ability over ranges of SA duration of 1.5 months ( $n = 9$ ), 4 months ( $n = 7$ ), and 8 to 11 months ( $n = 12$ ). She found that longer periods abroad seemed to be more beneficial and that the benefits continued even after the study period. Thus, by completing long-term SA programs, students tended not only to improve their L2 productive abilities but also to gain more confidence in the L2 and other aspects of their lives, which may have enhanced their motivation to continue to study after the SA programs.

However, not all the students obtain significant L2 gains from a long-term SA. Iida (2013) investigated 92 participants who were assigned to four different classes in a year. The highest proficiency group belonged to regular university academic courses (AC); the second-highest group attended a combination of AC and English as a second language (ESL) courses; the third group attended ESL courses and later moved to AC; and the lowest group attended ESL courses (Table 1). The results indicated that the highest proficiency group (AC) improved on neither the listening nor reading sections significantly, the second-highest group improved only on the listening section, while the lowest two groups improved both skills significantly. To this point, Iida claims that initial L2 proficiency and different learning styles and attitudes toward SA might have influenced the outcome. This initial proficiency factor is discussed in the next section.

### **Predeparture L2 Proficiency**

Initial L2 proficiency can also be a potential predictor for L2 progress in SA contexts. Based on several SA studies (e.g., Brecht, Davidson, & Ginsberg, 1995; Hernandez, 2016; Iida, 2013; Otsu & Satake, 2016), a trend can be noted that low-proficiency learners tend to make greater L2 gains than do higher-proficiency learners. Hernandez (2016) investigated American students studying in Spain for four weeks and reported that students with lower pre-program scores achieved greater gains than did students with higher pre-program scores. Otsu and Satake (2016) also observed that lower-proficiency Japanese students of English achieved greater gains than did higher-proficiency students in a four-week stay. The same tendency was observed among Japanese students who studied abroad for 15 weeks (Le Pavoux & Tsuchihira, 1998). One reason for this tendency may be that it is easier for low-proficiency students to increase their test scores because they have more to learn, while it may require greater effort for high-proficiency learners to match the score gains of low-proficiency learners (e.g., Baró & Serrano, 2011; Ife et al., 2000).

Another interpretation derives from the difference in the content of the SA programs students enroll in. In the case of Iida's (2013) study on one-year SA mentioned above, the largest TOEIC score gains of four different proficiency groups were made by the lowest group with gains gradually decreasing up to the highest group. Members of the highest group, who had already possessed sufficient L2 proficiency, attended regular academic courses of which the content and requisite skills were very different from those of the ESL courses the other groups attended. Consequently, not only the learning styles but also the language skills each group obtained may have differed (Iida, 2013). The knowledge and skills that the lower-proficiency groups acquired probably matched those required for success on TOEIC tests.

With regard to how learning styles differ between low- and high-proficiency learners, high-proficiency learners may not obtain advanced language gains from conversations in or outside of class. Freed (1990) investigated the effect of six weeks of study in France on American students and reported that low-proficiency students improved their grammar and reading abilities through increased interactions with native speakers of French. On the other hand, high-proficiency students benefited more from non-interactive exposure to the L2.

However, there may be a negative element that prevents low-proficiency learners from making progress in their L2 in an SA context. Low-proficiency learners tend to face more language problems, particularly at the initial stage of their SA. Furthermore, they have strong language-related anxiety and may be afraid of cross-cultural contact (e.g., Wang, 2013). Harada (2013) examined, in non-Japanese students studying in Japan, the relationship between learners' L2 proficiency and their degree of adaptation. The participants were from various countries, such as the U.S., the U.K., and France, and had stayed with a Japanese host family. The results of Harada's study revealed a significant correlation between the students' L2 speaking proficiency and their satisfaction in staying with their host families. The study also suggested that the students with low Japanese proficiency felt themselves to be in a state of crisis in their situation abroad and had difficulty acquiring interpersonal social skills. In other words, when learners' initial L2 proficiency is low, they seem to be more likely to have difficulty adapting to a host country due to repeated communication failures.

From these results, it is not clear whether initial proficiency is a crucial factor; thus, a larger-scale comparison is necessary. In this regard, the meta-analysis undertaken in the current study may produce more generalizable results.

### **Summary of the Literature Review and Research Questions for Meta-Analysis**

Reviewing the literature, SA duration seems to be an important predictor of L2 improvement. Long-term SA students made greater L2 gains than did middle- or short-term SA students. Among the L2 skills, listening seems to improve first—even participants in short-term SA programs may improve their listening skills. Oral and written fluency may improve more easily than accuracy and complexity, which even middle-term SA may not be sufficient to improve. As for predeparture L2 proficiency, many studies have reported the advantage of lower-proficiency students at their initial stage of any length of SA. However,

psychological factors may work against students of a lower proficiency level due to communication failures in SA contexts. Thus, the results are not consistent.

The literature review yielded general tendencies attributable to the factors of SA duration and initial proficiency. However, the following points have not been made clear: first, the degree to which L2 improvement in long-term SA is greater than that in short- or middle-term SA; second, whether low-proficiency students can gain the same effects as higher-proficiency students; and third, whether the factors of students' SA duration and predeparture proficiency interact with each other. To clarify these points, this study addresses the following three research questions (RQs):

- RQ1. To what degree does the effect on students' L2 proficiency differ between short-, middle-, and long-term periods of study abroad?
- RQ2. To what degree does the effect of studying abroad differ between low-, middle-, and high-proficiency students?
- RQ3. How do the factors of students' SA duration and their predeparture proficiency impact the effect size?

This study is unique in that accumulated studies were compared via meta-analysis so that more accurate L2 effects could be suggested regarding the impact of SA duration and initial L2 proficiency on L2 improvement. Depending on the length of SA programs, the cost and effort that students invest in them differ. Therefore, clarifying the degree to which the duration of studying abroad affects L2 proficiency could provide valuable information for students who plan to study abroad. In addition, determining whether initial proficiency influences the effectiveness of studying abroad could help students decide how much predeparture preparation is necessary to get the maximum benefit of SA.

## **Method**

### **Data Collection**

To accurately determine the effects of the two factors (or moderator variables) on L2 proficiency using meta-analysis, we collected studies that met the following criteria: (a) in order to avoid effects of age and linguistic distance between L1 and L2, participants were Japanese university students in Japan; (b) the major purpose of students' going abroad was to study English; (c) SA programs that students participated in were held at educational institutions in host countries; and (d) the students took both pre- and post-SA tests.

In regard to (a), numerous studies in second language acquisition (SLA) research have suggested that the speed and final outcome of learning are different for children versus adults (e.g., Brecht & Robinson, 1995; De Keyser, 2000) and are related to the linguistic distance between the L1 and the L2 (e.g., Chiswick & Miller, 2004). Learning should be easier if the L2 is structurally similar to the L1 (e.g., between languages of the Indo-European language family) than in cases where the L2 is very different from the L1 (e.g., between Japanese and

English). Besides this theoretical reason, it was practically difficult to obtain a substantial number of studies on high school students or other L1 and L2 participants. Regarding (b) and (c), since motivation (i.e., a willingness to study English) is an important factor for SLA (e.g., Hayashi, 2010), we limited our analysis to students who had a similar purpose for going abroad. The last condition, (d), is essential to convert the effects reported in studies into a common metric for meta-analysis.

To collect studies that met these criteria, we used databases such as Citation Information by National Institute of Informatics (CiNii), the Educational Resources Information Center (ERIC), the Web of Science, and Google Scholar. Additionally, since the current study focuses on Japanese SA participants, domestic journals such as *The Japan Association of College English Teachers (JACET) Journal*, *Language Education & Technology (LET)*, *Annual Review of English Language Education in Japan (ARELE)*, *The Japan Language Testing Association (JLTA) Journal*, *Kanto-Koshinetsu Association of Teachers of English (KATE) Bulletin*, *JACET KANTO Journal*, and a number of other universities' bulletins were included in the search. As a result of the literature search, approximately 80 articles were examined. However, the number of empirical studies that reported L2 effects obtained using pre- and post-tests was much smaller. Most of the studies either lacked the necessary information for meta-analysis or used qualitative methods, such as case studies. After a careful review of the studies, a total of 31 studies from 21 articles (in which there were 923 participants) were selected (see Appendix A) and analyzed using Comprehensive Meta-Analysis (Ver. 3.0) software for the meta-analysis.

## Analysis

**Two categorical variables for grouping.** Regarding RQ1 and RQ2, two moderator variables, SA duration and predeparture proficiency, were converted into categorical data. For the first variable, the 31 studies were divided into three groups based on the length of the SA programs: one month or less as short-term ( $k = 14$ ); more than one month to less than six months as middle-term ( $k = 10$ ); and six months to one year as long-term ( $k = 7$ ).

To report predeparture proficiency, various proficiency tests were used in the studies (Appendix A). The EIKEN (2017) is one of the most widely used standardized tests in Japan; it contains both written and oral sections and has seven different grade tests. The Certificate of English Level Test (CELT), the TOEIC IP, and the TOEFL ITP are paper-and-pencil tests created by Educational Testing Service (ETS). The first two tests consist of listening and reading sections, and TOEFL ITP has three sections (listening, structure, and reading). The Computerized Assessment System for English Communication (CASEC, n.d.) is an online English proficiency test that has four sections: vocabulary, expression, listening for the main idea, and dictation. The General Tests of English Language Proficiency (G-TELP, n.d.) Level 3 measures basic English abilities in normal communication and has three sections: grammar, listening, and reading & vocabulary. These tests have the benefit of normative comparisons and psychometric investigations for reliability and validity (Savicki, 2011).



To compare scores across the different tests and with the Common European Framework of Reference for Languages (CEFR) levels, initially, reported pre-test scores were converted into TOEIC scores using the various conversion tables available on the websites of certain testing institutions (e.g., Educational Testing Service, n.d.) and MEXT (2016). Then, the 31 studies were divided into three groups: TOEIC scores less than 450 (equivalent to CEFR A1 to lower half of A2) as Low; TOEIC scores 450 to 550 (upper half of CEFR A2) as Middle; and TOEIC scores over 550 (CEFR B1) as High (see Table 1). However, a conversion table for the Michigan ELI, used in Otsu and Satake (2016), was not found. Thus, the participants' levels were predicted mainly according to the level classifications mentioned in their study. Also, another two studies (i.e., Joto & Numoto, 1989; Numoto & Joto, 1990) which used EIKEN 2nd and 3rd mock questions were categorized into the low-proficiency group because these studies reported that the percentage of correctly answered questions was less than 50%. In addition, when the participants' levels of studies were uncertain, the author obtained clarification from the studies' authors.

As a result of grouping in terms of the predeparture proficiency factor, 20 studies were categorized into the low-proficiency group, six into the middle-proficiency group, and five into the high-proficiency group. In other words, most students' initial proficiency level was low, either at the CEFR A1 or the lower half of the A2 level.

**Two continuous variables for meta-regression analysis.** To answer RQ3, the Pearson correlation was used to obtain the relationship between the two moderator variables and the effect size, and meta-regression analysis was used to determine the impact of these variables on the effect size (i.e., the dependent variable). It is preferable to use continuous data in these analyses to produce more accurate results. Thus, the length of each SA program was indicated by the number of weeks, and the predeparture proficiency was represented by the converted TOEIC score as mentioned above (see Appendix A).

### **Effect Sizes**

First, the effects of SA on participants' L2 proficiency, as measured by the pre- and post-SA tests of each study, were converted into effect size (*ES*) called Hedges' *g*. This is the correction of Cohen's *d* because the sample size of some of the data was small. Cohen's *d* may give a biased estimate of the population effect size (Hedges and Olkin, 1985), especially for small samples ( $n < 20$ ). In this study, if *g* becomes negative, it means that the post-test score is higher than the pre-test score.

The magnitudes of the effect size are commonly interpreted as small (0.2), medium (0.5), and large (0.8; Cohen, 1988). More recently, Plonsky and Oswald (2014) recommended adopting the L2 field-specific benchmarks of small (0.4), medium (0.7), and large (1.0) for between-group contrasts, and small (0.6), medium (1.0), and large (1.4) for pre/post or within-group contrasts. Since a pre/post experimental design generally produces a larger *g* value due to intragroup correlations, their new pre/post interpretation of *g* values was adopted in this study.

A meta-analysis calculates effect sizes under the fixed effects model and the random effects model. The assumption under the fixed model is that the studies used in the meta-analysis share a common true effect, allowing the model to estimate the common effect size. On the other hand, the assumption under the random effects model is that the true effects vary between the studies; consequently, the summary effect size is the weighted average of the effects in the studies. Since this model produces more conservative estimates and is recommended when heterogeneity is present (Borenstein, Hedges, Higgins, & Rothstein, 2009), the present study reports effect sizes calculated in accordance with the random model.

## Results

A meta-analysis produced the effect sizes of the 31 studies. Out of them, 17 were significant in their effect sizes at  $p < .05$ , which means that nearly half did not observe significant L2 gains. The overall effect size was significant, ranging from small to middle effect sizes, with  $g = -0.83$  (95% CI= -1.05, -0.61),  $z = -7.49$ , and  $p < .001$ .

However, the heterogeneity test yielded a Cochran's  $Q$  statistic that was significant at  $Q(30) = 130.49$ ,  $p < .001$ , and  $I^2 = 77.01$ , indicating that 77% of the observed variance across the studies was due to real heterogeneity rather than chance (Borenstein et al., 2009). This variance is relatively high according to the standard of Higgins, Thompson, Deeks, and Altman (2003) that  $I^2$  values of 25%, 50%, and 75% be considered as low, moderate, and high, respectively. Thus, in order to explain the large variance, the 31 studies were analyzed further based on SA duration and predeparture proficiency.

### Analysis of Study Abroad Duration

Table 1

*A Meta-Analysis by Study Abroad Duration*

Group	$k$	$g$	95% CI		$SE$	$Z$	$p$	$Q$	$df$	$p$	$I^2$
			Lower	Upper							
Fixed effect analysis											
Short	14	-0.34	-0.49	-0.19	0.08	-4.50	<.001	15.20	13	.295	14.48
Middle	10	-0.65	-0.78	-0.51	0.07	-9.33	<.001	19.85	9	.019	54.65
Long	7	-1.64	-1.91	-1.37	0.14	-12.08	<.001	25.98	6	<.001	76.91
Total within								61.03	28	<.001	
Total between								69.49	28	<.001	
Overall	31	-0.65	-0.74	-0.56	0.05	-13.56	<.001	130.49	30	<.001	77.01
Mixed effects analysis											
Short	14	-0.36	-0.53	-0.19	0.09	-4.15	<.001				
Middle	10	-0.82	-1.06	-0.57	0.12	-6.54	<.001				
Long	7	-1.77	-2.41	-1.14	0.32	-5.48	<.001				
Total within											
Total between								23.81	2	<.001	
Overall	31	-0.56	-0.70	-0.43	0.07	-8.13	<.001				

*Note.* A negative  $g$  value means that the post-test score was higher than the pre-test score.

The result of the meta-analysis for SA duration showed that out of 14 studies of the short-term group, only three were significant at  $p < .05$ , while the middle- and long-term groups had nine and six significant effect sizes, respectively. As shown in Table 1, the effect size of the short-term group in the mixed effects analysis (i.e., the random model) was very small at -0.36; that of the middle-term group was -0.82, which was between small and medium; and that of the long-term group was quite large at -1.77. Also, the difference in effect size between the groups was quite large and statistically significant, with  $Q(2) = 23.81$ ,  $p < .001$ .

Regarding the heterogeneity of these groups, the  $Q$  value of the short-term SA group was not significant, with  $Q(13) = 15.20$ ,  $p = .295$ , and  $I^2 = 14.48$ , but  $Q$  values for the middle- and long-term groups were significant, and  $I^2$  increased ( $I^2 = 54.65$  and  $74.91$ , respectively). This means that studies in the short-term SA group were quite homogenous and the factor of SA duration was an influential predictor; thus, the variance of this group was explained well by this factor. However, when the SA term is longer, other variables could be involved.

### Analysis of Predeparture L2 Proficiency

Next, the predeparture L2 proficiency factor was analyzed. As shown in Table 2, the effect sizes of the three proficiency groups were similar (-0.83, -0.75, and -0.96) and were not significantly different,  $Q(2) = 0.32$ ,  $p = .851$ . Thus, learners' initial L2 proficiency did not make a significant difference in the degree of L2 gains in SA environments. In other words, regardless of their predeparture proficiency, students on average had a similar chance to improve their L2 proficiency.

Table 2  
*A Meta-Analysis by Predeparture Proficiency*

Group	$k$	$g$	95% CI		$SE$	$Z$	$p$	$Q$	$df$	$p$	$I^2$
			Lower	Upper							
Fixed effect analysis											
Short	20	-0.57	-0.68	-0.47	0.05	-10.65	<.001	95.18	19	<.001	80.04
Middle	6	-0.73	-1.01	-0.46	0.14	-5.20	<.001	12.05	5	.034	58.52
Long	5	-1.19	-1.50	-0.88	0.16	-7.59	<.001	9.09	4	.059	55.97
Total within								116.32	28	<.001	
Total between								14.18		<.001	
Overall	31	-0.65	-0.74	-0.56	0.05	-13.56	<.001	130.49	30	<.001	77.01
Mixed effects analysis											
Short	20	-0.83	-1.10	-0.57	0.14	-6.16	<.001				
Middle	6	-0.75	-1.19	-0.30	0.23	-3.27	.001				
Long	5	-0.96	-1.55	-0.37	0.30	-3.20	.001				
Total within											
Total between								0.32	2	.851	
Overall	31	-0.83	-1.04	-0.62	0.11	-7.65	<.001				

*Note.* A negative  $g$  value means that the post-test score was higher than the pre-test score.

However, the heterogeneity of this factor displayed the reverse tendency of the SA duration. Studies in the low-proficiency group were quite heterogeneous,  $Q(19) = 95.18$ ,  $p < .001$ ,  $I^2 = 80.04$ , while the middle- and high-proficiency groups' heterogeneities were moderate (with  $I^2 = 58.52$  and  $55.97$ , respectively). This means that studies in the low-proficiency group were significantly more diverse than studies in the higher-proficiency groups. Overall, this factor did not partition the variance of the 31 studies well, which implies that the variable was not a crucial factor for L2 improvement.

### The Relationship Between Study Abroad Duration and Predeparture L2 Proficiency

Figure 2 shows that the relationship between SA duration and predeparture L2 proficiency. The middle- and high-proficiency groups were combined and re-named as the *upper group* because the data points of these groups across the three SA durations were too small. The original low-proficiency group was indicated as the *lower group*. Although the effect sizes of the two proficiency groups similarly increase from the short- to middle-term SA periods, the effect size of the lower group was greater than that of the upper group.

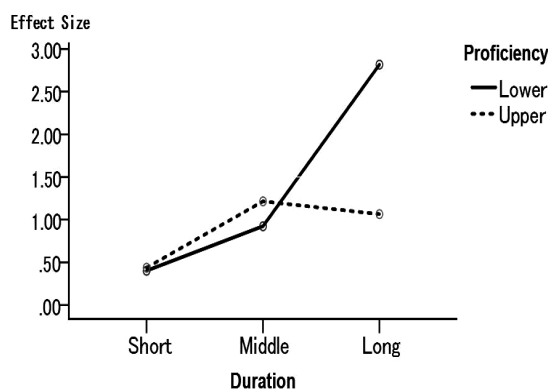


Figure 2. The relationship between duration and proficiency. Since the data points of the high and intermediate proficiency groups were small, they were combined as the upper proficiency group.

To further analyze the relationship between the two factors and the effect size, the Pearson correlation was calculated (see Table 3) using the two continuous variables of the number of weeks and the converted TOEIC score for each SA program (as explained in the Analysis section). The correlation between the effect size and the duration was moderately high at  $r = .52$ , whereas there was no correlation between the effect size and the initial proficiency. In addition, the correlation between the duration and the proficiency was moderately high at  $r = .58$ , which indicates that there was a moderately high interaction between the two factors. In other words, there was a tendency for students at a low initial proficiency to attend shorter-term SA programs.

Table 3

*The Correlations Between Effect Size, Duration, and Proficiency*

	<i>M</i>	<i>SD</i>	<i>1.</i>	<i>2.</i>	<i>3.</i>
1. ES ( <i>g</i> -value)	0.91	0.84	-		
2. Duration (Wks)	14.42	15.39	.52**	-	
3. Proficiency (TOEIC)	422.63	130.51	-.04	.58**	-

*Note.* Duration = the number of weeks in SA; Proficiency = converted TOEIC scores.

Since the two factors were found to be significantly related ( $r = .58$ ), an interaction factor was created and entered into the regression equation with the two moderator variables in the meta-regression analysis. To test the impact of the three independent variables accurately, they were recalculated to have a mean of zero (i.e., centered). The duration variable was named *duration\_c*; the predeparture proficiency variable was named *proficiency\_c*; and the interaction variable was named *duration\_c x proficiency\_c* (Borenstein, Hedges, Higgins, & Rothstein as cited in Borenstein & Hedges, 2015).

The results are shown in Table 4. The variables *duration\_c x proficiency\_c* ( $R^2$  change = .15,  $p < .001$ ) and *duration\_c* ( $R^2$  change = .49,  $p < .001$ ) were significant and explained 15% and 49% of the total variance of the effect size, respectively, whereas the variable *proficiency\_c* was not a significant predictor ( $R^2$  change = .05,  $p = .261$ ). All three variables together explained 69% of the total variance of the effect size.

Table 4

*Meta-Regression for the Influence of Duration, Proficiency, and the Interaction*

	Coefficient	<i>SE</i>	95% CI		<i>Z</i>	<i>p</i>	$R^2$ change
			Lower	Upper			
Intercept	-1.02	0.10	-1.21	-0.83	-10.39	<.001	
<i>Duration_c</i>	-0.05	0.01	-0.07	-0.04	-6.78	<.001	.49
<i>Proficiency_c</i>	0.00	0.00	0.00	0.00	1.12	.261	.05
<i>Duration_c x Proficiency_c</i>	0.00	0.00	0.00	0.00	3.80	<.001	.15

*Note.* Test of the model:  $Q(3) = 47.35$ ,  $p < .001$ ; Proportion of total between-study variance explained by the model:  $R^2 = .69$ .

**Publication Bias**

Lastly, to confirm whether the results produced in the meta-analysis were neutral, publication bias was checked for by producing a funnel plot. As shown in Figure 3, all the studies are plotted using white dots based on the standard errors along the Y-axis against the effect sizes (Hedges'  $g$  in the random model) along the X-axis. Larger studies, which in general have smaller standard errors, appear toward the top of the graph and near the mean effect size, while smaller studies tend to spread toward the bottom of the graph (Borenstein et

al., 2009). It is assumed that if the shape of the funnel plot is symmetrical, there is no publication bias. Figure 3 indicates that there was one study with a large effect size on the left, but overall, the dots were distributed on both sides of the center line.

The Trim and Fill procedure (Duval & Tweedie, 2000), which can impute missing studies and add to the funnel plot, produced two imputed missing studies on the funnel plot (indicated with black dots). In addition, this method re-computed the summary effect size,  $g = -0.88$ , (95% CI = -1.10, -0.66) as shown with the black rhombus on the bottom of the figure. The re-estimated effect size was not so different from the observed effect size,  $g = -0.83$  (95% CI = -1.05, -0.61). Therefore, the publication bias was not a problem.

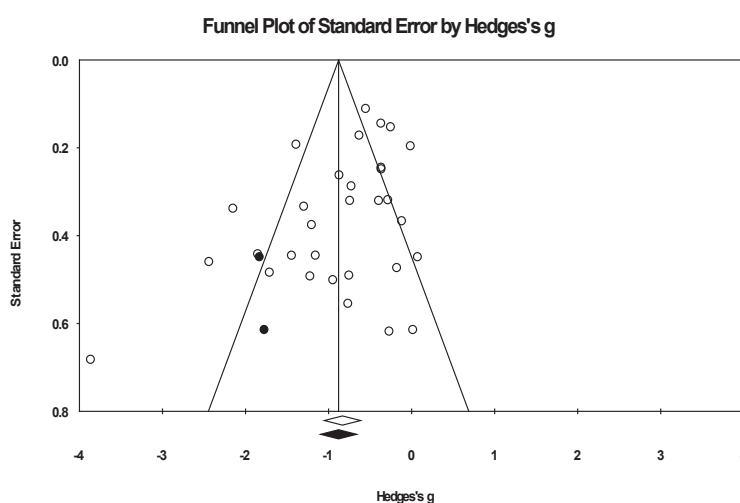


Figure 3. A Funnel plot with the Trim and Fill of 31 study abroad studies for publication bias.

### Discussion

This study synthesized 31 studies with a total of 923 students using a meta-analysis method and examined three RQs. Regarding RQ1 (*To what degree does the effect on students' L2 proficiency differ among short-, middle-, and long-term periods of study abroad?*), students who studied abroad longer showed greater improvement in their L2. The effect sizes of the short-, middle-, and long-term SA durations were -.36, -.82, and -1.77, respectively, and were thus distinctly different.

To illustrate the magnitude of these effect sizes in terms of L2 scores, Figure 4 displays the three groups' pre- and post-departure score distribution curves. Taking the predeparture scores of each group to have a mean of 50 and a standard deviation (*SD*) of 10, as shown in a dotted distribution curve, the mean post-test score of the short-term study group would become 53.6, the mean of the middle-term study group would become 58.2, and that of the

long-term group would drastically increase to 67.7, which is more than one standard deviation unit. This reveals that the effect of long-term study abroad was more than four times as great as that of short-term study and more than twice as great as that of middle-term study abroad.

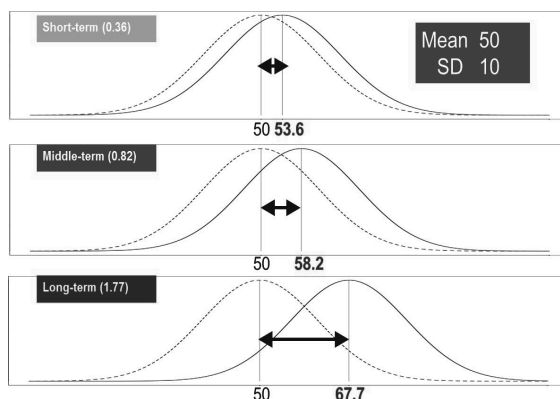


Figure 4. Image of effects by different study-abroad durations.

Such a distinctly small effect size of the short-term SA implies that the period of one month or less is too short for students to derive any benefit from their ESL environment in terms of L2 improvement. Therefore, a similar gain can be obtained without going abroad (e.g., Kuno, 2011). Another factor related to short duration of study abroad was pointed out by Katori (2016). She interviewed 15 students who had studied abroad for 4.8 weeks on average but had not experienced any significant L2 improvement ( $p < .05$ ) and found that only seven students had studied abroad in order to improve their English abilities. Other purposes of their going abroad included to have intercultural experiences ( $n = 12$ ); to test their limits ( $n = 7$ ); to have fun ( $n = 4$ ); and simply to go abroad ( $n = 4$ ). Thus, short-term study participants who did not experience significant L2 improvement may have a weak commitment to improving their English. In fact, as Kobayashi (1999) reported earlier, significant gains were obtained even in a three-week SA program, when students were highly motivated to study English and had a clear goal of studying, such as when grading was done based on their post-test score. Therefore, for short-term SA to be successful, it may be crucial that students have a strong motivation to study the target language and set clear goals for the duration of the SA program.

Unlike short-term SA, long-term SA groups were able to not only greatly improve participants' L2 proficiency, but also to increase participants' confidence in their L2 proficiency, with its attendant benefits. As a consequence, even after having left the SA programs, they tended to maintain a higher motivation to study further, and become independent learners. Thus, overall, students obtained the greatest benefit from long-term SA programs. However, when we examine individual long-term studies more closely, we see that

their L2 gains were diverse. This clearly indicates that other factors continue to influence participants over a significant period of time during the program. Based on the literature review, those factors include the content and strictness of the SA programs (e.g., Kobayashi, 1999) as well as participants' motivation (e.g., Harada, 2013; Sasaki, 2011).

Concerning RQ2 (*To what degree does the effect of studying abroad differ among low-, mid-, and high-proficiency students?*), it was found that learners' predeparture proficiency did not have a significant impact on the degree of L2 improvement, which implies that whatever their initial L2 proficiency, it is neither a hindrance nor an advantage for their language development. The results seem to contradict those of some previous studies. In Hernández's (2016) study, advanced American students of Spanish (CEFR level B2) improved less in speaking than did lower-proficiency students. Otsu and Satake (2016) also concluded that it would be more difficult for advanced learners to obtain the same gains as beginners. To this point, Engle and Engle (2004) mentioned that American students of French tended to stop attempting to progress further once they could communicate with reasonable efficiency in most social situations. This reason makes intuitive sense because in conversations, advanced speaking skills are not required and advanced learners may not feel it necessary to improve their speaking or other L2 skills further, rather preferring to concentrate on their own fields of study.

However, the high-proficiency Japanese learners of English in this study were quite different from these American students of Spanish or French in terms of their L2 skills and the linguistic distance between L1 and L2. The Japanese students in the high-proficiency group were still at an intermediate level (around CEFR B1) according to their scores on the TOEIC reading and listening tests. According to the results of a nationwide English test administered by the Ministry of Education (MEXT, 2016), only 11.0% and 17.9% of Japanese 3rd-year high school students had their speaking and writing skills, respectively, assessed at levels of A2 or above, while 31.9% and 26.2% had their reading and listening skills, respectively, assessed at those levels. Therefore, their L2 skills, in particular speaking skills, were highly likely to be much lower than those of the American students in Hernández's (2016) or Engle and Engle's (2004) studies, and they probably had plenty of room to improve their English proficiency while studying abroad.

Lastly, the inquiry into RQ3 (*How do the factors of students' SA duration and their predeparture proficiency impact the effect size?*) revealed two significant relations. One was the existence of a moderately high correlation between the SA duration and the effect size ( $r = .52$ ). The meta-regression analysis also revealed a large variance in the effect size ( $R^2 = .49$ ) was explained by the SA duration factor. The other was a moderately high relation between the SA duration and the predeparture proficiency ( $r = .58$ ), indicating that low-proficiency students had a tendency to participate in short-term SA programs.

Perhaps because of this tendency, low-proficiency students did not reveal greater L2 gains than did higher proficiency students. In other words, the length of SA was too short to make a difference in L2 gain between low- and high-proficiency groups. In fact, Figure 2 indicates that the effect size of the two proficiency groups was not different for short- and



middle-term SA, whereas for long-term SA, the effect size showed a clear tendency for low-proficiency students to make greater L2 gains. Furthermore, the results of the meta-regression analysis showed that the interaction factor was significant and accounted for the 15% variance of the effect size. Another reason why low-proficiency students may not have obtained a greater L2 gain than higher-proficiency students is that they tended to have more communication failures that led to more anxiety and fear of interacting with native speakers of L2 at the initial stage of their SA period (e.g., Wang, 2013).

In addition, the low- and high-proficiency students in this study may have had different advantages for improving their L2. The low-proficiency learners may have taken advantage of having more room to make greater L2 gains since their initial pre-test scores were low, although their gains may not have materialized in a short-term SA period. On the other hand, the higher-proficiency students were still at an intermediate level, that is, at a developing stage, and their advantage may have consisted in experiencing fewer communication failures at the initial stage of their SA period.

### **Conclusions and Future Research**

The present study examined the effects of study abroad on university students' English proficiency and revealed the following significant findings. First, the length of the study abroad period was a crucial factor for predicting L2 gains. Students who studied abroad for 6 to 12 months, on average, had L2 gains four times as large as those who studied abroad for one month or less, and more than twice as large as those who studied abroad more than one month to less than six months. Second, Japanese university students whose initial English levels are from the CEFR A1 level up to around the CEFR B1 level may have similar chances to improve their English proficiency. Third, low-proficiency students tend to attend short study abroad programs of one month or less. To generalize from these findings, if students' primary purpose for going abroad is to improve their L2 proficiency and they expect to make substantial L2 progress, more than six months' study abroad would be preferable. However, individual differences in the effect of SA duration on L2 improvement tend to become larger as students study abroad longer. Therefore, in addition to study abroad duration, factors such as motivation and clear goal setting for study abroad may be crucial.

One limitation of this study is that most of the participants were assigned to the low-proficiency group and the short-term group; thus, accurate analysis of predeparture proficiency was not possible. Another limitation is that it was not possible to show the degree of improvement in each subskill due to the small number of relevant studies. In particular, studies that reported pre- and post-tests on productive skills were scarce. Thus, if a large enough number of studies could be collected, it would be interesting to investigate the relationship between the length of study abroad and the progress of L2 productive skills. Additionally, due to a lack of description in each study examined, it was difficult to compare the content of study abroad programs. Thus, we need more studies that focus in greater detail on the characteristics of study abroad programs. For example, it would be valuable if a

comparative study could be undertaken to explore programs that have produced significant L2 gains versus those that have not.

This study focuses solely on L2 gains in order to make the use of meta-analysis possible, but it does not deny the value of short-term SA programs due to their limited effect in terms of L2 improvement. Throughout the literature review, many short-term studies reported that studying abroad seems to be worthwhile because participants often gain valuable benefits from their cross-cultural experiences. Lastly, the results for the effect sizes presented in this study can be generalized to other SA studies because no serious publication bias was detected. Therefore, it is hoped that this study will contribute to program organizers in planning SA programs and students in deciding which SA programs to attend.

### Acknowledgements

I would like to thank Mr. Keita Igarashi, Mr. Hiroki Maeda, Mr. Hideaki Oka, and Mr. Takeshi Kato for their cooperation in the earlier study. I also extend my gratitude to the two anonymous reviewers for their constructive feedback.

### References

- Baró, À. L., & Serrano, R. S. (2011). Length of stay and study abroad: Language gains in two versus three months abroad. *Revista Española de Lingüística Aplicada*, 24, 95–110.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: John Wiley & Sons, Ltd.
- Brecht, R. D., Davidson, D. E., & Ginsberg, R. B. (1995). Predictors of foreign language gain during study abroad. In B. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 37–66). Amsterdam, Philadelphia: John Benjamins.
- Brecht, R. D., & Robinson, J. L. (1995). On the value of formal instruction in study abroad: Student reactions in context. In B. F. Freed (Ed.), *Second language acquisition in a study abroad context* (pp. 317–334). Amsterdam, Philadelphia: John Benjamins.
- CASEC. (n.d.). CASEC. Retrieved from <http://global.casec.com/>
- \*Chiba, K. (2005). TOEIC Bridge ni miru 2 kagetsu ryugaku no eigoryoku heno kouka no kensho [The effect of 2-month study abroad program on English proficiency on the TOEIC Bridge test]. *Bulletin of Sakura no Seibo Junior College*, 29, 115–123.
- Chiswick, B. R., & Miller, P. W. (2004). Linguistic distance: A quantitative measure of the distance between English and other languages. *IZA Discussion Paper No. 1246*.
- DeKeyser, R. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, 22, 499–533.
- Comprehensive Meta-Analysis Ver. 3 [Computer software]. Englewood, NJ: Biostat.
- Duval, S. J., & Tweedie, R. L. (2000). A nonparametric "trim and fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, 95(449), 89–98.
- Educational Testing Service. (n.d.). *Mapping the TOEIC® and TOEIC Bridge™ Tests on the*

- CEFR. Retrieved from [https://www.ets.org/s/toeic/pdf/toeic\\_cef\\_mapping\\_flyer.pdf](https://www.ets.org/s/toeic/pdf/toeic_cef_mapping_flyer.pdf)
- Engle, L., & Engle, J. (2004). Assessing language acquisition and intercultural sensitivity development in relation to study abroad program design. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 10, 219–236.
- EIKEN. (2017). *EIKEN Grades*. Retrieved from <http://steipeiken.org/grades>
- Freed, B. (1990). Language learning in a study abroad context: The effects of interactive and non-interactive out-of-class contact on grammatical achievement and oral proficiency. In J. Atlatis (Ed.), *Georgetown University round table on languages and linguistics* (pp. 459–477). Washington D.C.: Georgetown University Press.
- \*Fujisawa, Y., & Komori, M. (2005). Chuki gogaku ryugaku program no eigo gakuryoku heno kouka ni kansuru kenkyu: Nihon no daigaku eigokyouiku heno teigen [Developing English language proficiency through a mid-term study abroad program]. *Osaka Shoin Jyoshidaigaku Ronsyuu*, 42, 35–47.
- Furuya, N. (2005). Tanki ryugaku ha koukateki de aruka? [Evaluating a short-term abroad program]. *Bunka Gakuen Univeristy Bulletin*, 13, 19–30.
- G-TELP. (n.d.). G-TELP™. Retrived from <http://www.g-telp.jp/about/a004.html>
- Harada, T. (2013). *Ryugakusei no douki to homestay* [Homestays and the motivation of international students]. Okayama, Japan: Fukuroshuppan.
- Hayashi, A. (2010). 4 kagetsukan ryuugaku program no kitai to seika ni tsuite – chuubu daigaku Ohio daigaku chouki kenshuu program sanku gakusei no baai [The expectations and the results of the 4-month study abroad program: in case of Chubu University students joined in Ohio University long-term study program]. *Gendai Kyouikugaku Kenkyuu Kiyuu*, 3, 19–29.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hernandez, T. A. (2016). Acquisition of L2 Spanish requests in short-term study abroad. *Study Abroad Research in Second Language Acquisition and International Education*, 1, 186–216. doi: 10.1075/sar.1.2.03her
- Higgins, J, Thompson, S, Deeks, J., & Altman, D. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327, 557–560.
- Hoffman-Hicks, S. D. (2000). *The longitudinal development of French foreign language pragmatic competence: Evidence from study abroad participants*. Unpublished doctoral dissertation. Bloomington, IN: Indiana University.
- Ife, A., Vives, B. G., & Meara P. (2000). The impact of study abroad on the vocabulary development of different proficiency groups. *Spanish Applied Linguistics*, 4, 55–84.
- \*Iida, T. (2013). Ryugaku to daigakusei no eigoryoku: Kaigai ESL tonu yoriyoi kankei wo motomete [Study abroad and English proficiency of university students: A consideration about the better relationship with ESL]. *JACET Kansai Journal*, 15, 106–124.
- Institute of International Education. (2016). Duration of U.S. Study Abroad, 2004/05-2014/15. *Open doors report on international educational exchange*. Retrieved

from <http://www.iie.org/opendoors>

- \*Iwakiri, M. (1993). Effects of a study abroad program on the English development of Japanese college students, *JACET Bulletin*, 24, 41–60.
- \*Jyoto, A., & Numoto, K. (1989). Homestay no eigoryoku he no kouka (I) [The effect of homestay on English proficiency (I)]. *Chuugoku Tankidaigaku Kiyou*, 20, 165–177.
- Katori, M. (2016). The trends of Japanese students' participation in study abroad programs. *US-China Foreign Language*, 14, 239–249.
- \*Kimura, K. (2006). Eigoken taizai ga gakusei no eigoryoku ni oyobosu eikyo – Tanki gogaku kenshu ni yori eigoryoku wa koujyo suruka [The influence of studying abroad on language proficiency: Do short-stay programs in English speaking countries affect students' English proficiency?]. *Bulletin of Policy and Management, Shobi-Gakuen University*, 12, 1–20.
- \*Kimura, K. (2011). Daigakusei no kaigai tanki kenshuu he no ichi kousatu: listening to writing ni shouten wo atete [How effective the short-term studying program in an English-speaking country is for Japanese university students: Focusing on the listening and writing proficiencies]. *Bulletin Policy and Management, Shobi-Gakuen University*, 21, 17–30.
- \*Kobayashi, T. (1999). Kaigai tanki ryugaku kenshuu de eigoryoku ha donokurai nobiru monoka [To what extent will Japanese EFL learners' English proficiency improve in intensive English programs abroad?]. *The review of liberal arts*, 97, 83–100.
- \*Kuno, H. (2011). Sanshuukan no tanki kaigai gogaku kenshuu ga daigakusei no eigonouryoku ni oyobosu eikyou ni tsuite [Effects of a short-term (3-week) study abroad program on the English proficiency of four-year college students]. *Journal of Hokkaido Bunkyo University*, 12, 127–145.
- Le Pavoux, M., & Tsuchihira, T. (1998). The New Zealand study abroad program of Shinshu Honan Women's Junior College: Overview and its effect. *Bulletin of Shinshu Honan Women's Junior College*, 15, 41–58.
- MEXT. (2017). *Nihonjin no ryugaku jokyō* [Status of Overseas Study by Japanese Nationals]. Retrieved from [http://www.mext.go.jp/a\\_menu/koutou/ryugaku/\\_icsFiles/afieldfile/2017/12/27/1345878\\_02.pdf](http://www.mext.go.jp/a_menu/koutou/ryugaku/_icsFiles/afieldfile/2017/12/27/1345878_02.pdf)
- MEXT. (2016). *Heisei 27 nendo eigoryoku chousa (Kokou 3 nensei) kekka no gaiyo* [The result of a survey about the English proficiency of senior students in high school in 2015]. Retrieved from [http://www.mext.go.jp/b\\_menu/shingi/chousa/shotou/117/shiryō/icsFiles/afieldfile/2016/05/24/1368985\\_7\\_1.pdf](http://www.mext.go.jp/b_menu/shingi/chousa/shotou/117/shiryō/icsFiles/afieldfile/2016/05/24/1368985_7_1.pdf)
- Muta, Y. (2007). Ryugaku de kakutoku dekiru eigo noryoku: Eigoryoku (hikki, listening) to gakushu kankyou heno tekiou noryoku ni kannsuru chousa [English proficiency gained by studying abroad: survey on student writing and listening ability and their adaptation to a new environment]. *Nagasaki Tankidaigaku Kenkyuu Kiyou*, 19, 1–12.
- \*Muta, Y. (2009). Ryugaku zengo no eigoryoku bunseki ni yoru kongo no eigoshido no kento: Canada, Australia 3 kagetsu ryugaku no gakushukouka wo saidaigen ni takameru [English proficiency before and after a 3-month study abroad program:

- Potential of the NJC integrated study abroad system]. *Nagasaki Tankidaigaku Kenkyuu Kiyuu*, 21, 1–10.
- Nonaka, T. (2005). Kaigaigogakukenshu no koukasokutei [Measuring the effect of studying abroad]. *Niigata Seiryuu Jyoshi Tankidaigaku Kenkyuu Houkoku*, 35, 7–12.
- Nonaka, T. (2008). Kaigaigogakukenshu no koukasokutei (2) [Measuring the effect of studying abroad (2)]. *Niigata Seiryuu Jyoshi Tankidaigaku Kenkyuu Houkoku*, 38, 43–49.
- \*Nonaka, T., & Seki, K. (2016). Kaigaigogaku kenshu·Tanki gogaku ryugaku niyoru eigo unyou noryoku no henka [Change in English proficiency through a homestay program and a short-term study abroad program]. *Niigata Seiryuu Jyoshi Tankidaigaku Kenkyuu Houkoku*, 46, 89–97.
- \*Numoto, K., & Jyoto, A. (1990). Homestay no eigoryoku heno kouka (II) [The effect of homestay on English proficiency (II)]. *Chuugoku Tankidaigaku Kiyuu*, 21, 135–142.
- \*Otsu, R., & Satake, M. (2016, August). How effective is a short-term study abroad program? A case of Tokiwa University. *Ryugaku Koryu*, 65, 16–24. Japan Student Services Organization.
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64, 878–912. doi:10.1111/lang.12079
- Sasaki, M. (2011). Effects of varying lengths of study-abroad experiences on Japanese EFL students’ L2 writing ability and motivation: A longitudinal study. *TESOL Quarterly*, 45, 1–105.
- \*Sasaki, M., Mizumoto, A., & Murakami, A. (in press). Developmental trajectories in L2 writing strategy use: A self-regulation perspective. *Modern Language Journal*.
- Savicki, V. (2011). Relationship of Foreign Language Proficiency to Study Abroad Outcomes and Inputs. *Frontiers: The Interdisciplinary Journal of Study Abroad*, 63, 63–85.
- Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition*, 26, 173–199.
- \*Seki, K., Nonaka, T., & Sumiida, A. (2009). Tanki gogaku ryuugaku program no kouka sokutei (3) [Assessment of a short-term study abroad program (3)]. *Research reports of Niigata Seiryuu University Junior College*, 39, 105–113.
- \*Suzuki, R., & Hayashi, C. (2014) Kaigai gogaku tanki ryugaku no kouka [The Effects of Short-term Study Abroad Programmes on Students’ English Proficiency and Affective Variables]. *KATE Journal*, 28, 83–96.
- \*Tajima, C., & Cookson, S. (2012). Factors affecting language gain during study abroad: A study on homesickness. *Oberlin Gengo Kyouiku Ronsou*, 8, 105–123.
- \*Tanaka, K., & Ellis, R. (2003). Study-abroad, language proficiency, and learner beliefs about language learning. *JLTA Journal*, 25, 63–85.
- Wang, L. (2013). *Daini gengo shuutoku ni okeru shinriteki fuan no kenkyuu* [The studies on psychological anxiety in second language acquisition]. Tokyo: Hitsujishobou.

*Note.* References marked with an asterisk indicate studies included in the meta-analysis.

## Appendix A

Study	<i>n</i>	Pretest	Pretest	Posttest	Posttest	Wks	English test	Duration	Prof-level
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Joto & Numoto (1989)	32	42.63	8.23	45.96	9.84	4	self-made	Short	Low
Kimura (2006)	19	28	7.77	31.37	9.01	3	EIKEN Pre2moc	Short	Low
Kimura (2011)	14	41.97	16.2	46.07	46.07	3	EIKEN Pre2moc	Short	Low
Kobayashi (1999)	31	459	31.4	488	34.4	4	TOEFL ITP	Short	Mid
Kuno (2011)	8	56.13	12.62	58.38	11.71	3	TOEIC moc	Short	Low
Nonaka (2005)	51	375.14	25.33	375.41	28.63	3	TOEFL ITP	Short	Low
Nonaka (2008)a	20	344.00	69.35	390.50	52.56	3	TOEIC IP	Short	Low
Nonaka (2008)b	9	498.89	52.78	494.44	58.71	3	TOEIC IP	Short	Mid
Nonaka & Seki (2016)a	85	383.9	85.1	404.1	75.9	3	TOEIC IP	Short	Low
Numoto & Joto (1990)	33	44.67	9.43	48.33	10.44	4	self-made	Short	Low
Otsu & Satake (2016)a	9	29.56	5.7	37.89	7.22	4	Michigan ELI	Short	Low
Otsu & Satake (2016)b	8	43.38	3.78	48.88	6.81	4	Michigan ELI	Short	Mid
Otsu & Satake (2016)c	4	61.75	9	61.5	13.72	4	Michigan ELI	Short	High
Suzuki & Hayashi (2014)	19	555	74.16	579	91.63	3	CASEC	Short	Mid
Chiba (2005)	70	125.43	15.29	134.97	14.92	8	TOEIC bridge	Mid	Low
Fujisawa & Komori (2005)	11	147.64	23.39	196.27	30.89	16	G-TELP Level3	Mid	High
Iwakiri (1993)	96	170.58	22.18	178.91	23.18	5	TOEFL ITPmoc	Mid	Low
LePavoux & Tsuchihira (1998)	21	268.57	72.42	355.95	59.43	15	TOEIC IP	Mid	Low
Muta (2009)	12	359.83	134.86	526.42	81.96	12	CASEC	Mid	Low
Nonaka & Seki (2016)b	11	352.7	55.3	433.2	77.1	12	TOEIC IP	Mid	Low
Sasaki (2018)a	16	90.38	14.66	117.13	27.00	11	CELT	Mid	Low
Seki, Nonaka & Sumida (2009)a	6	398.3	50.72	450.8	74.13	12	TOEIC IP	Mid	Low
Tajima (2013)	25	543.72	87	600.84	67.2	15	CASEC	Mid	Mid
Tanaka & Ellis (2003)	166	426.73	35.05	445.28	32.29	15	TOEFL ITP	Mid	Low
Iida (2013) ESL	14	471.4	95.8	646.4	87.8	48	TOEIC IP	Long	Mid
Iida (2013) ESL to AC	66	573	92	697.9	86.9	48	TOEIC IP	Long	High
Iida (2013) ESL & AC	8	673.8	128.5	764.4	97.7	48	TOEIC IP	Long	High
Iida (2013) AC	4	856.3	111.6	888.3	96.5	48	TOEIC IP	Long	High
Nonaka & Seki (2016)c	27	363.7	47.6	492.6	68.8	24	TOEIC IP	Long	Low
Sasaki (2018)b	12	86.75	15.66	144.17	12.92	38	CELT	Long	Low
Seki, Nonaka & Sumida (2009)b	16	378.4	39.64	513.4	65.28	24	TOEIC IP	Long	Low
Total	923								

*Note.* 1. Nonaka's (2008) data were divided into two proficiency groups. 2. Nonaka & Seki (2016) had three groups (a-c) of different SA lengths. 3. Otsu & Satake (2016) had three proficiency groups (a-c). 4. Sasaki's (2018) study had three groups (1.5 month, 4 month-, and 8 or more month-SA), but the first two groups were combined for this study (a, b). 5. Iida (2013) had four proficiency classes (ESL: English as a Second Language; ESL to AC: students changed from ESL class to AC class later. ESL & AC: students had both ESL and AC classes. AC: academic class).