The Effects of Study-abroad Experience on Speaking Performance:
The Case of a Three-week English Study Abroad Program for Kyushu International University Students

Yukiko Hosoki

1. Introduction

According to the White Paper of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) (2011), the number of Japanese students who studied abroad has been drastically decreasing from a peak of 82,045 in 2004 to 59,923 in 2009 (see Figure 1). It is often pointed out that this is greatly due to the influence of the "inward-looking" perspective of the younger generation in Japan in which individuals are satisfied with their lives in Japan and possess limited interest in things abroad. However, several other factors in the social system have also contributed to the current situation. Many students consider studying abroad as a disadvantage when it comes to job hunting, as they may miss out on recruitment and application opportunities in Japan. Furthermore, they feel that the costs of studying abroad are too economically burdensome. The university system itself also does not provide the institutional support needed to encourage students to study abroad. To some extent, all these conditions seem to be behind this decline (MEXT, 2012).

Figure 1. The number of Japanese students who studied abroad

(Source: UNESCO, OECD, IIE, Chinese Education Unit, etc. cited in MEXT, 2011)

（出典）　ユネスコ文化統計年鑑，OECD調べ，IIE（OPEN DOORS），中国教育部調べ等
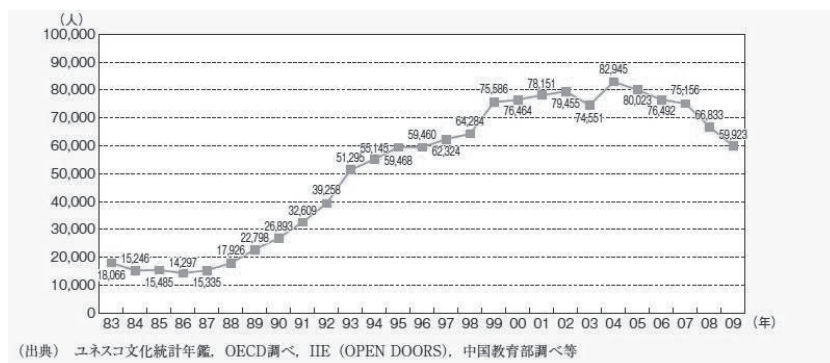
In its White Paper, MEXT（2011）expressed concern over the possibility that this lack of overseas experience among the young generation may have a severe impact on Japan's competitiveness in the international community. In 2009, the Ministry started a system for sending students overseas to promote globalization and the strengthening of Japan's global competitiveness. Under this system, in 2010, 1,231 Japanese students studied abroad for three to twelve months. The Ministry is planning a shorter overseas study program of less than three months, in order to promote two-way exchanges.

Looking at study abroad destinations, though the number of students for each country is on the decline and the destinations have become increasingly diverse, English-speaking destinations like the U.S., England, and Australia make up the majority, or almost 60 % of the total destinations (see Table 1). A follow-up study on those who studied abroad in 2011 shows disaggregated data by length studied abroad. According to the report, the length with the highest

percentage is between one year and two years (25.8%), followed by six months to one year (23.3%) and less than three months (17.4%) (JSSO, 2011, p. 14). Students choose program lengths depending on their purposes such as getting a degree, doing research in their specialized areas, engaging in cross-cultural exchange, etc. The purpose of students who study abroad for less than three months is especially clear, because 30.4% of them specifically set out to develop their language skills (JSSO, 2011, p.14).

Table 1. Main study abroad destinations and the number of Japanese students who went to those destinations (2009)

(Source: IIE, Chinese Education Unit, Taiwan Education Unit, OECD, etc. cited in MEXT, 2011)

Study Abroad Destinations	Number of Students
U. S. A	24,842
China	15,409
The United Kingdom	3,871
Australia	2,701
Taiwan	2,142
Germany	2,140
Canada	2,005
France	1,847
New Zealand	1,025
Korea	989

Contrary to the downward trend in the number of students studying abroad, many Japanese colleges and universities place great value on overseas programs. Yokota (2006, p. 34) reports that MEXT announced a policy to rejuvenate a system to send Japanese students abroad as well as receive foreign students in Japan. Furthermore, more than 80 % of Japanese colleges and universities consider a short-term overseas language program important and

50％ actually implement such a program. However important such programs are deemed to be, university-sponsored overseas language programs have limitations in length due to the costs students have to pay in the present unfavorable economic climate. Mainstream programs nowadays are three- to five-week programs combined with language learning at the host institution and cultural experiences with homestay families or dormitory life. Although the lengths of such programs may not be sufficient to acquire proficiency in a foreign language, considering the great deal of money that students pay for such programs, many seem to perceive such programs as having high educational value. But pedagogically, what kind of educational effectiveness can we expect from such short-term study abroad programs?

2. Previous Studies

Adachi (2010) argues for the necessity to clarify the educational effectiveness of study abroad programs since universities integrate them into their curriculums. He categorizes the effectiveness and outcomes of study abroad programs into four types: academic effect, language proficiency, ability to adapt to a different culture, and personal progress. He argues that there is wide consensus among researchers that study abroad programs produce the first two outcomes – academic effect and language proficiency. The acquisition of language proficiency is closely related to the length of a study abroad program: the longer the program, the more students learn the target language. In the case of a short study abroad

program, what level of language proficiency can one expect to acquire?

While there are various studies that discuss the positive effects of studying abroad, there are also some studies that indicate weak associations between linguistic improvement and studying abroad. Yashima and Zenuk-Nishide (2008) are skeptical about the necessity of studying abroad to acquire language proficiency. From the results of their comparative study regarding linguistic gains between two groups of students – one in a "study abroad" environment and the other in an "at home" environment – they argue that students at home (in their native country) could develop just as many linguistic skills as the other group if they experience "acculturation," "motivation," and the "willingness to communicate" by imagining that they were in an international community. There seems to be considerable difficulties in creating all three of these aforementioned conditions in an actual classroom environment in Japan, but as Eguchi (2010) points out, beneficial changes such as improvements in the quality of Japanese English teachers and their teaching methods, expansions in the adoption of native English teachers, and above all, the effective use of computer technology for language education, have narrowed the differences between English learning environments in Japan and those in English-speaking countries. The advantages that study abroad programs have for the acquisition of English proficiency may no longer be limited to such programs, and may be readily available in Japan.

Nevertheless, many studies have found that studying abroad does have positive effects on English language acquisition. Sabet

(2007) investigated the effectiveness of study-abroad experiences on students' English proficiency by administering the Secondary Language English Proficiency (SLEP) exam to 53 college students. The results revealed positive effects and found that 60 hours of English classes in English-speaking countries and 140 hours of English classes in Japan are equivalent in terms of efficacy. Kuno (2011) tested the effectiveness of a three-week study abroad program with that of a ten-month e-learning program and found that the students in the three-week study program improved their listening skills to the same degree as the ones in the ten-month e-learning class. Matsumoto (2010) also investigated improvements in the listening skills of students who participated in a four-week study abroad program by using SLEP, and found that the program had positive effects on the students' listening skills. He also found that students with lower levels of proficiency improved more than those with higher levels, and from this, he deduced that listening skills do not improve homogeneously. Although some controversy may exist, much research has shown that short-term study abroad experiences positively affect a learner's listening skills.

　　　Improvements in speaking skills through short-term study abroad programs have not been adequately observed in extant studies as speaking ability is so complex, and combines many different skills that are difficult to improve over the short run. Furthermore, investigative methods for measuring improvements in speaking skills have not been adequately developed. Koizumi and Fujimori (2010) analyzed improvements in speaking performance across 67 "progress sensitive" measures. They carried out four months of instruction

on two groups of Japanese EFL learners – 10 high school students and 5 university students – and administered a pretest and posttest involving picture description. They computed the data on the 67 measures examined, and found that there were moderate to high differences between the two groups in 34 of the measures. On the basis of these results, they provide an interesting finding that "the students in the low proficiency group may generate more fluent, syntactically complex, and lexically complex speech, but show no improvement in accuracy" (Koizumi & Fujimori, 2010, p. 89). Kawaguchi and Kamimoto (2003) tried to identify the distinctive features of the fluent and non-fluent speech of low-intermediate Japanese EFL learners. They examined speech rate, hesitation factors, and facilitation factors, and concluded that the most distinctive features that differentiate fluent from non-fluent speakers were positioning and frequency of pauses, the use of connectives, and a prefabricated pattern of "when I" clauses. Sugita (2007) tried to verify the effectiveness of his speaking teaching method which was based on the use of the Dynamic Listening and Speaking Method. Utilizing picture descriptions, he administered a pretest and a posttest, and analyzed the data based on four evaluative dimensions – vocabulary, accuracy, fluency, and complexity – in order to observe changes in speaking ability. Even though improvements were limited, he found some improvements in speech production such as improvements in vocabulary and fluency.

The educational effectiveness of studying abroad has to be defined. Many studies on its effects on language proficiency – especially listening skills – have been conducted, finding positive

results. In contrast, its effectiveness on improvements in speaking skills is still controversial among researchers.

3. KIU English Study Program

Kyushu International University (KIU) has integrated study abroad programs into its curriculum. Several programs are carried out in different countries every year, and the English Study Program examined in this study was implemented from February 13 to March 6, 2012. The program is designed to help students acquire and develop their English proficiency and experience cross-cultural exchange. The 17 students in the program stayed in a university dormitory on campus for the first week and then spent two weeks with their homestay families. For the three-week duration of the program, they attended ESL classes at the English Language Institute (ELI) attached to Eastern Washington University (EWU) in the state of Washington in the United States. The students were exposed to English for 24 hours a day for three weeks in classes, at the dormitory, and at the homes of their homestay families. This gave them opportunities to mingle with American students of their own age as well as experience American family life. To further deepen their experience, the program included a San Francisco sightseeing tour on the return trip to Japan. The students stayed at a youth hostel for two nights and further broadened their cross-cultural experience.

At ELI, the students took four ESL classes per day from Monday to Friday: two fifty-minute reading and writing classes in

the morning and two fifty-minute listening and conversation classes (or note-taking classes) in the afternoon, totaling 60 hours of classes throughout the three weeks. There were ten to fifteen students in each class, and the classes were taught by experienced native English teachers who placed emphases on interactive communication. Since the students joined EWU's regular ten-week ESL classes, they had a golden opportunity to interact with other ESL students from non-English-speaking countries in their classes.

After school, International Peer Advisers (IPAs) organized out-of-class activities that provided the students with opportunities to improve their communicative skills by freely and actively interacting with the IPAs through various activities related to American culture. Even though most of the activities were optional, the students actively participated in most of the activities during the weekdays and spent time with their host families on the weekends. In sum, in line with the program's objectives, the students were intensively exposed to English for three weeks. At first, they appeared to be very nervous, but gradually they became less tense and even seemed to develop confidence in themselves. They spontaneously started to use English not only with the IPAs but also with their Japanese peers.

4. Purpose and Research Questions

This case study examines improvements in speaking performance among KIU students who participated in a three-week study abroad program in the United States. The study aims to

answer the following questions.

i. How does a three-week English study abroad program affect students' speaking performance?

ii. What aspects of the study abroad program correlate with improvements in speaking performance?

iii. What preparations should be made before the commencement of a short-term study abroad program to further enhance its effectiveness?

5. Method

5.1 Participants

To obtain permission to conduct this research and recruit participants, the purpose and content of the study were explained to the students beforehand, and eleven out of the seventeen students agreed to participate in the study. The research participants consisted of nine sophomores and two juniors majoring in international studies or law at KIU. One student had previously lived in the United States for one year, but all of the remaining students had never stayed in an English-speaking country for more than a month. Their English abilities ranged widely as Table 2 shows. Before the program, the students took a placement test for reading and writing and another test for listening and conversation. Based on their overall scores, they were placed in classes of the appropriate level. The class levels ranged from level 1 for the lowest to level 5 for the highest.

Table 2. Distribution of students across class levels

(based on performance on the ELI Placement Test)

English Levels		1	2	3	4	5
No. of students	R & W	2	2	7	3	3
	L & C	3	6	3	3	2

R: reading; W: writing; L: listening; C: conversation/speaking

5.2 Oral Assessment Test and Procedures

Oral assessment tests were conducted by utilizing the measurements of the "HOPE Chukousei no tame no Eigo Supiikingu Tesuto" (HOPE English Speaking Test for Junior and Senior High School Students). The test was given to each student twice – once before the ELI program started and once after the program ended. Each student's speaking performance was recorded with an IC recorder and transcribed at a later date. The following are the specific procedures that were taken:

Section 1: Warm up (30 seconds)

The purpose of this section was to create an environment where students could relax and speak English freely by exchanging greetings and engaging in a short conversation. Speaking performance in this step was not evaluated.

Section 2: Picture description task (60 seconds)

In this section, each student was shown a picture and asked to describe it in as much detail as possible for one minute. The picture was obtained from the "HOPE Chukousei no tame no Eigo Supiikingu Tesuto" and the same picture was used for both pretests and posttests.

Section 3: Follow-up questions (60 seconds)

The purpose of this section was to elicit maximum potential utterances from a student by carrying out a dialogue about a topic related to the picture used in Section 2.

Section 4: Role-play (90 seconds)

After evaluating a student's speaking performance in Sections 2 and 3, in this section, an appropriate role-play card was chosen by the tester and the student was asked to engage in a role-play conversation.　The purpose of this section was to see how much of an initiative a student could make in maintaining a conversation.

Section 5: Follow-up questions (90 seconds)

As was the case in Section 3, the purpose of this section was to elicit maximum potential utterances from a student.

Section 6: Wind down (30 seconds)

This section aimed to both make the student feel comfortable in having taken the test and motivate the student to keep improving his/her English speaking skills.　Speaking performance in this step was not evaluated.

5.3 Data Analysis

The students' recorded utterances from Sections 2 to 5 were transcribed, and the utterances from the picture descriptions in Section 2 were used as the main data for the analysis, whereas the utterances from the remaining sections were used as supplementary data.　The evaluative dimensions and objective indicators for assessing speaking performance were borrowed and translated from Sugita (2007, p. 57) and are presented in Figure 2 below.　Sugita

(2007) measures speaking competence along four dimensions – vocabulary, complexity, accuracy, and fluency – and operationalizes each dimension by establishing specific objective indicators.

Figure 2. Evaluative Dimensions & Objective Indicators

Evaluative dimensions	Objective indicators
vocabulary	1) the number of unpruned tokens (the number of all the words in an utterance)
	2) the number of pruned tokens (the number of unpruned tokens - the number of words repeated or corrected)
	3) the number of types (the number of first appeared words)
	4) the number of types ÷ the number of pruned tokens (the rate of not repeating the same words)
complexity	5) the number of pruned tokens ÷ the number of C-units (the number of pruned tokens per C-unit)
accuracy	6) the number of errors
	7) the number of C-units with errors
	8) the number of C-units with errors ÷ the number of C-units (the rate of C-units with errors among all C-units)
fluency	9) the number of unpruned tokens ÷ speaking time (seconds) x 60 (the number of unpruned tokens per minute)
	10) the number of pruned tokens ÷ speaking time (seconds) x 60 (the number of pruned tokens per minute)
	11) the number of pruned tokens ÷ the number of unpruned tokens (the rate of words that do not interrupt fluency)

(Sugita, 2007, p. 57; translated into English by the author)

6. Data

To examine the four dimensions of students' speaking performance according to the objective indicators above, the following three steps were taken to collect and analyze the data.

6.1 Data Collection

As mentioned above, data obtained from the picture description task were used as the main data, so following the metrics of the objective indicators presented in Figure 2, the numbers of words, errors, and clauses in each student's (students A to K) attempt at this exercise were counted and computed to produce numerical values for each objective indicator. This was done for both the pretest and posttest picture description tasks, and the results are presented in

Tables 3 and 4, respectively.

6.2 Statistical Analysis: Paired Samples T-tests

Paired Samples T-tests (2-tailed) were run to ascertain whether there are statistically significant differences between the pretest and posttest score averages. Tests were run on each the aforementioned evaluative dimensions.

6.3 Analysis of Individual Data

Although the number of subjects in this study is too small to generalize about the effects of short-term study abroad experiences on speaking performance, the performances of individual students were examined in greater detail to better understand the effects of studying abroad on speaking performance.

7. Results and Discussion

7.1 Descriptive Statistics

Tables 3 and 4 present the pretest and posttest speaking performance results of all the students. The letters A to K represent students and are arranged from low to high based on student placement. Noticeable changes between the pretest and posttest scores appear in certain evaluative dimensions and are discussed in further detail below.

(1) Vocabulary

Figures 3 to 6 respectively examine each of the four dimensions

Table 3. Pretest Results

evaluative dimensions	obj. ind.	A	B	C	D	E	F
vocabulary	1	20	31	32	31	43	40
	2	18	29	32	31	41	38
	3	17	19	21	23	25	30
	4	0.9444	0.6551	0.6562	0.7419	0.6097	0.7894
complexity	5	2	4.333	5.333	5.167	5.857	4.222
accuracy	6	8	4	3	6	7	8
	7	9	4	3	6	6	4
	8	0.889	0.667	0.5	1	0.857	0.444
fluency	9	14.286	18.416	48	33.818	36.857	22.857
	10	12.857	18.416	48	33.818	35.143	21.714
	11	0.9	0.935	1	1	0.9534	0.95

evaluative dimensions	obj. ind.	G	H	I	J	K
vocabulary	1	31	51	33	94	81
	2	31	50	31	94	79
	3	20	24	21	45	40
	4	0.6451	0.48	0.6774	0.4787	0.5063
complexity	5	6.2	7.143	6.2	8.545	6.077
accuracy	6	1	4	4	7	1
	7	1	4	3	6	1
	8	0.2	0.571	0.5	0.545	0.077
fluency	9	22.143	43.714	26.4	47.797	74.769
	10	22.143	42.857	24.8	47.797	72.923
	11	1	0.98	0.939	1	0.975

Table 4. Posttest Results

evaluative dimensions	obj. ind.	A	B	C	D	E	F
vocabulary	1	18	54	68	72	68	69
	2	18	53	68	72	67	68
	3	15	32	35	43	28	39
	4	0.8333	0.6037	0.5147	0.5972	0.4179	0.5735
complexity	5	2.25	5.889	5.667	5.538	6.091	6.181
accuracy	6	8	4	8	14	3	6
	7	7	3	6	13	3	4
	8	0.875	0.333	0.5	1	0.273	0.364
fluency	9	16.615	27.458	41.212	34.56	52.308	39.057
	10	16.615	26.95	41.212	34.56	51.538	38.491
	11	1	0.981	1	1	0.985	0.986

evaluative dimensions	obj. ind.	G	H	I	J	K
vocabulary	1	37	55	48	70	92
	2	37	55	47	70	92
	3	23	26	30	40	42
	4	0.6216	0.4727	0.6382	0.5714	0.4565
complexity	5	6.167	6.885	5.222	7.778	7.077
accuracy	6	3	4	6	3	3
	7	2	4	6	2	3
	8	0.333	0.5	0.667	0.222	0.23
fluency	9	38.276	53.226	37.894	46.667	96.842
	10	38.276	53.226	37.105	46.667	96.842
	11	1	1	0.979	1	1

of vocabulary presented in Figure 2, and show the differences between each student's pretest and posttest measures for the picture description exercise.　Figure 3 shows the number of unpruned tokens (all the words in an utterance), Figure 4 shows the number of pruned tokens (the number of unpruned tokens － the number of words repeated or corrected (i.e. the number of words after the omission of repeated or corrected words)), Figure 5 shows the number of types (the number of first appeared words), and Figure 6 shows the rate of not repeating the same words in an utterance (the number of types ÷ the number of pruned tokens).　For both tests, the students were asked to describe the picture for about one minute, and most of them finished within 80 seconds (except three students; Students B and C spoke for a longer time only on the posttest and Student J spoke for a shorter time only on the posttest).　The numbers in Figures 3 to 5 are affected by this inconsistency in the length of time.　However, more noteworthy is that it is clear that the students autonomously spoke more on the posttest.　This may indicative of the students' motivations and positive attitudes toward speaking English.　The number of pruned tokens per minute (evaluative dimension 10) was examined in order to assess fluency, and these results are introduced later in this section.

　　　Figure 6 shows the rate of new words in an utterance.　The posttest did not show improvement in vocabulary; rather, the rate of new words in an utterance dropped.　This suggests that the three-week study abroad program did not contribute to an increase in vocabulary because even though the students surely acquired the ability to speak more, they ended up using the same words and
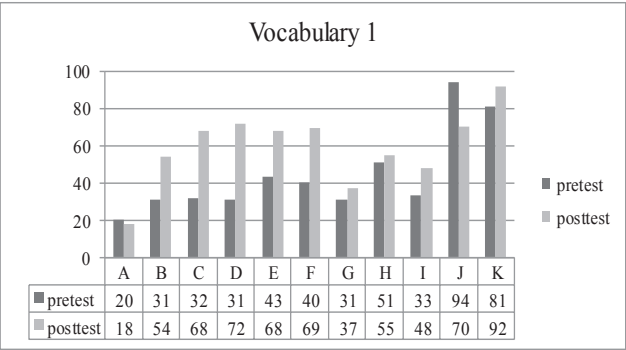
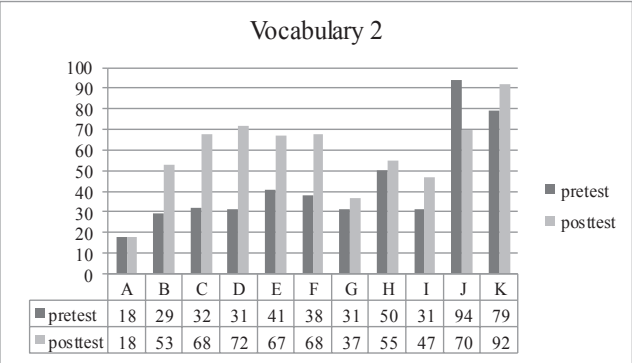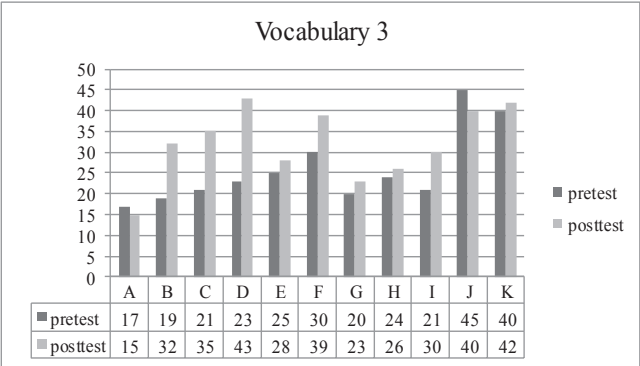Figure 3. Vocabulary 1 (number of unpruned tokens)

Vocabulary 1

	A	B	C	D	E	F	G	H	I	J	K
pretest	20	31	32	31	43	40	31	51	33	94	81
posttest	18	54	68	72	68	69	37	55	48	70	92

Figure 4. Vocabulary 2 (number of pruned tokens)

Vocabulary 2

	A	B	C	D	E	F	G	H	I	J	K
pretest	18	29	32	31	41	38	31	50	31	94	79
posttest	18	53	68	72	67	68	37	55	47	70	92

Figure 5. Vocabulary 3 (number of types)

Vocabulary 3

	A	B	C	D	E	F	G	H	I	J	K
pretest	17	19	21	23	25	30	20	24	21	45	40
posttest	15	32	35	43	28	39	23	26	30	40	42

phrases repeatedly to describe the picture.

Figure 6. Vocabulary 4 (rate of not repeating the same words)

	A	B	C	D	E	F	G	H	I	J	K
pretest	0.94	0.65	0.65	0.74	0.6	0.78	0.64	0.48	0.67	0.47	0.5
posttest	0.83	0.6	0.51	0.59	0.41	0.57	0.62	0.47	0.63	0.57	0.45
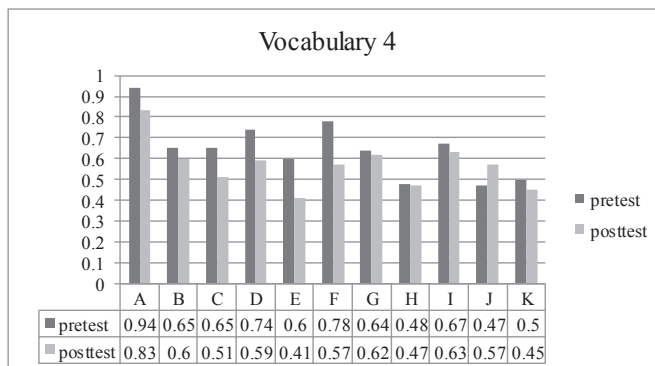
(2) Complexity

　　Complexity was calculated by taking the number of pruned tokens per Communication Unit (C-Unit). The C-Unit is one of the basic units of discourse analysis. Although it is similar to the T-Unit (a main clause with any subordinate clauses), the difference is that the C-Unit includes non-clausal structures which have communicative value. It is clear from Figure 7 that the students of lower English ability barely increased in the number of pruned tokens per C-Unit and the students of higher English ability slightly decreased. However, the changes are too small to infer the existence of effectiveness.

(3) Accuracy

　　Figures 8 to 10 examine the three dimensions of accuracy. Figure 8 shows the number of errors students made in their utterances and Figure 9 shows the number of C-units with errors.

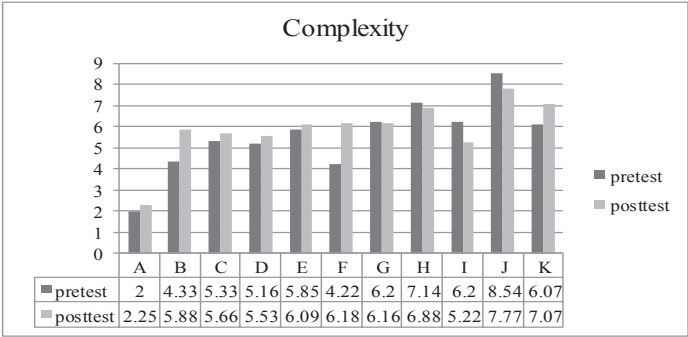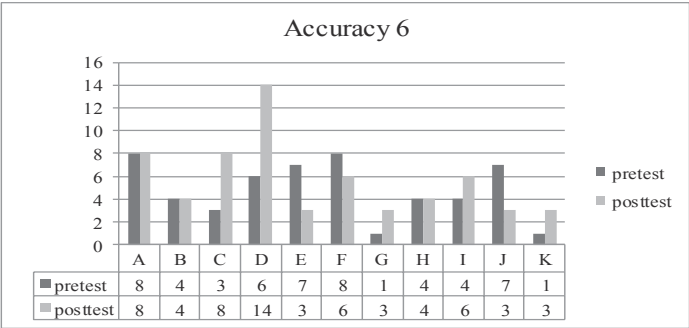Figure 7. Complexity (number of pruned tokens per C-unit)

Complexity

	A	B	C	D	E	F	G	H	I	J	K
pretest	2	4.33	5.33	5.16	5.85	4.22	6.2	7.14	6.2	8.54	6.07
posttest	2.25	5.88	5.66	5.53	6.09	6.18	6.16	6.88	5.22	7.77	7.07

Figure 8. Accuracy 6 (number of errors)

Accuracy 6

	A	B	C	D	E	F	G	H	I	J	K
pretest	8	4	3	6	7	8	1	4	4	7	1
posttest	8	4	8	14	3	6	3	4	6	3	3

Figure 9. Accuracy 7 (number of C-units with errors)

Accuracy 7

	A	B	C	D	E	F	G	H	I	J	K
pretest	9	4	3	6	6	4	1	4	3	6	1
posttest	7	3	6	13	3	4	2	4	6	2	3
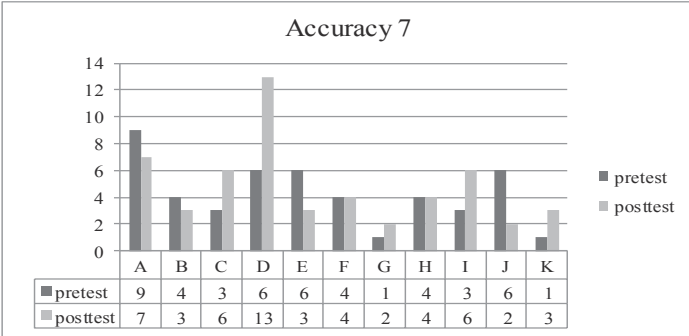
Figure 10. Accuracy 8 (rate of C-units with errors among all C-units)

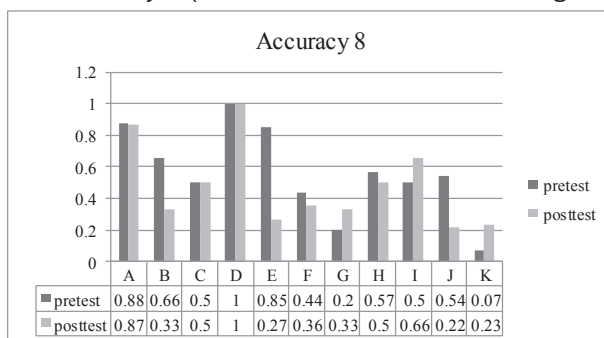|	A	B	C	D	E	F	G	H	I	J	K
pretest	0.88	0.66	0.5	1	0.85	0.44	0.2	0.57	0.5	0.54	0.07
posttest	0.87	0.33	0.5	1	0.27	0.36	0.33	0.5	0.66	0.22	0.23

Figure 10 shows the rate of accuracy calculated by dividing the number of C-Units with errors by the total number of C-Units per utterance. This rate was calculated for each student in both the pretest and posttest. Figures range from 0 to 1 with 0 being completely accurate and 1 being completely inaccurate. The results seem to be random; three students stayed the same, five increased, and three decreased in terms of accuracy. This inconsistency in progress across students occurred regardless of English abilities, indicating that the three-week program did have a systematic effect on accuracy.

(4) Fluency

Figures 11 and 12 show the results for fluency, or the number of unpruned or pruned tokens per minute. The results indicate noticeable changes between the two tests, indicating that the three-week program did have an effect on fluency levels. Most students improved along both dimensions of fluency. The average number of pruned tokens per minute also increased from 34.9 on the pretest to

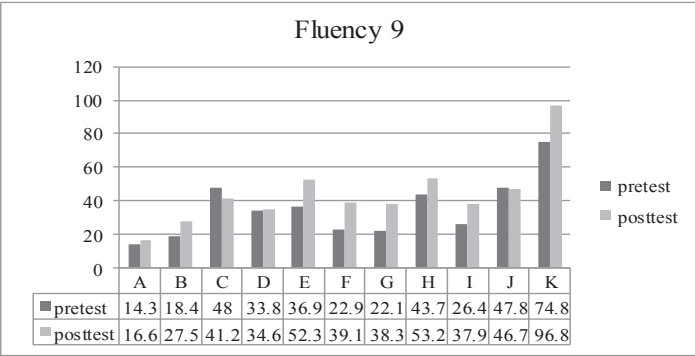Figure 11. Fluency 9 (number of unpruned tokens per minute)

Fluency 9

	A	B	C	D	E	F	G	H	I	J	K
pretest	14.3	18.4	48	33.8	36.9	22.9	22.1	43.7	26.4	47.8	74.8
posttest	16.6	27.5	41.2	34.6	52.3	39.1	38.3	53.2	37.9	46.7	96.8

Figure 12. Fluency 10 (number of pruned tokens per minute)

Fluency 10

	A	B	C	D	E	F	G	H	I	J	K
pretest	12.9	18.4	48	33.8	35.1	21.7	22.1	42.9	24.8	47.8	72.9
posttest	16.6	27	41.2	34.6	51.5	38.5	38.3	53.2	37.1	46.7	96.8
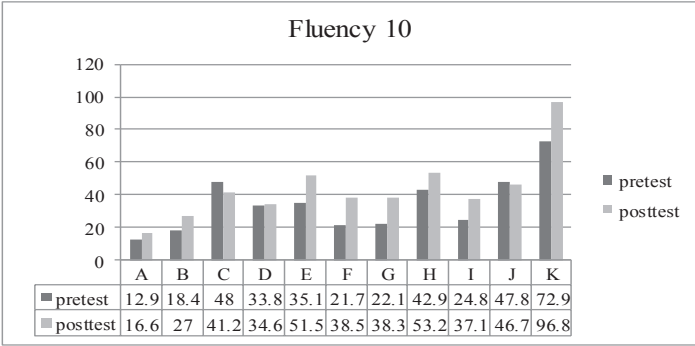
Figure 13. Fluency 11 (rate of the words that do not interrupt fluency)

Fluency 11

	A	B	C	D	E	F	G	H	I	J	K
pretest	0.9	0.94	1	1	0.95	0.95	1	0.98	0.94	1	0.98
posttest	1	0.98	1	1	0.99	0.99	1	1	0.98	1	1
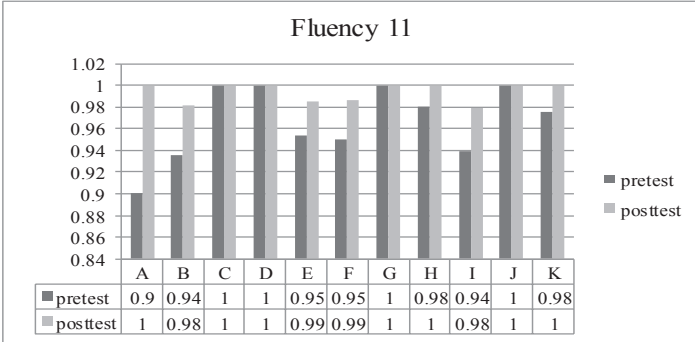
43.8 on the posttest (See Table 6 and 7). Furthermore, the degree of improvement tends to be higher for students with higher English abilities. Figure 13 shows the rate of words that do not interrupt fluency (calculated by dividing the number of pruned tokens by the number of unpruned tokens per utterance). Figures range from 0 to 1 with 0 being not fluent and 1 being fluent. These results indicate that the three-week program did have an effect on fluency.

7.2 Inferential Statistics: Paired Samples T-tests

　　　To see if the three-week study abroad program helped students make significant improvements in their spoken English performance, paired samples t-tests were used to ascertain whether the average scores of the posttest were significantly different from those of the pretest across all objective indices (1 to 11). The results are presented in Table 5.

　　　The results show that all objective indices (1 to 4) for vocabulary were significant at the $p<0.05$ level (2-tailed). As the differences between the pretest and posttest mean scores for indices 1 to 3 are all negative, posttest mean scores were higher than pretest mean scores for all three indices. This is also reflected in the negative t-values for these indices. These results suggest that the three-week study abroad program helped the students acquire the ability to produce quantitatively more words in an utterance with more variation in the words that were used. However, the average of the students' rate of new word usage in an utterance (objective index 4, or Vocabulary 4) dropped to a lower value in the posttest, and this difference was also statistically significant. This may at

first, seem to point to a contradictory decline in the students' level of vocabulary, but under closer inspection, this is not the case. Students increased the number of words they knew, but just as equally or to a larger extent also increased the number of pruned tokens in their utterances. Mean rates of new word usage may have declined between the pretest and posttest, but if we look at the magnitude of the difference in pretest and posttest means for Vocabulary 3 (number of types in an utterance) and Vocabulary 2 (number of pruned tokens in an utterance), we see that the Vocabulary 3 posttest mean only increased by 6.18 units compared to the 15.73 unit increase in the Vocabulary 2 posttest mean. As Vocabulary 4 is calculated by dividing Vocabulary 3 by Vocabulary 2, it makes sense that the rate obtained will decrease if the magnitude of growth for Vocabulary 2 is greater than that of Vocabulary 3. Substantively, this means that the students' vocabulary level did improve over the study abroad program, but this improvement was overshadowed by a greater improvement in an area that is relevant to level of fluency: the number of pruned tokens in an utterance.

With regard to complexity (objective index 5) and accuracy (objective indices 6 to 8) there were no significant differences between the pretest and posttest mean scores. These results indicate that the changes observed could have been due to chance. Therefore, we cannot infer that the study abroad program had an effect on these two evaluative dimensions.

Regarding fluency (objective indices 9 to 11), the pretest and posttest mean scores for the number of unpruned and pruned tokens per minute (objective indices 9 and 10) were significantly

different (p<0.01 level (2-tailed)).　Slightly weaker, but nevertheless significant differences were also observed between the mean pretest and posttest scores for the rate of words that do not interrupt fluency (p<0.05 level (2-tailed)).　These results indicate that there were significant changes across these indicators of fluency over the three-week study abroad program.

Table 5. Paired Samples T-test Results

Evaluative dimensions/Objective indicators		Paired Differences			
		Mean	Std. Deviation	t	Sig. (2-tailed)
Vocabulary	1)	-14.90909	18.72140	-2.641	.025*
	2)	-15.72727	18.62842	-2.800	.019*
	3)	-6.18182	7.49424	-2.736	.021*
	4)	.08032	.09033	2.949	.015*
Complexity	5)	-.33345	.89524	-1.235	.245
Accuracy	6)	-.81818	3.60051	-.754	.468
	7)	-.54545	3.14209	-.576	.578
	8)	.08664	.23558	1.220	.251
Fluency	9)	-8.64164	8.85937	-3.235	.009**
	10)	-9.18309	9.21315	-3.306	.008**
	11)	-.02715	.02988	-3.013	.013*

df = 10 for all indicators
*p<0.05; **p<0.01

7.3 Analysis of Individual Speaking Performance

　　In order to understand changes in the students' speaking performance in greater detail, the two highest level students, the two lowest level students, and two other students who showed notable

changes in their utterances were selected and examined in their performances along four select indicators: vocabulary 4 (the rate of not repeating the same words), complexity (the number of pruned tokens per C-unit), accuracy 8 (the rate of C-units with errors among all C-units), and fluency 10 (the number of pruned tokens per minute). In Tables 6 and 7, the average, highest, and lowest marks for the pretest and posttest are shown for all students across all four indicators. The students under detailed examination are highlighted in grey. Their speech manuscripts are presented in Appendix A.

Table 6. Pretest

	A	B	C	D	E	F	G	H	I	J	K	total	average	highest	lowest
vocabulary	0.9444	0.6551	0.6562	0.7419	0.6097	0.7894	0.6451	0.48	0.6774	0.4787	0.5063	7.1842	0.653109	0.9444	0.4787
complexity	2	4.333	5.333	5.167	5.857	4.222	6.2	7.143	6.2	8.545	6.077	61.077	5.552455	8.545	2
accuracy	0.889	0.667	0.5	1	0.857	0.444	0.2	0.571	0.5	0.545	0.077	6.25	0.568182	1	0.077
fluency	12.857	18.416	48	33.818	35.143	21.714	22.143	42.857	24.8	47.797	72.923	380.468	34.588	72.923	12.857

Table 7. Posttest

	A	B	C	D	E	F	G	H	I	J	K	total	average	highesst	lowest
vocabulary	0.8333	0.6037	0.5147	0.5972	0.4179	0.5735	0.6216	0.4727	0.6382	0.5714	0.4565	6.3007	0.572791	0.8333	0.4179
complexity	2.25	5.889	5.667	5.538	6.091	6.181	6.167	6.885	5.222	7.778	7.077	64.745	5.885909	7.778	2.25
accuracy	0.875	0.333	0.5	1	0.273	0.364	0.333	0.5	0.667	0.222	0.23	5.297	0.481545	1	0.222
fluency	16.615	26.95	41.212	34.56	51.538	38.491	38.276	53.226	37.105	46.667	96.842	481.482	43.77109	96.842	16.615

Students J and K – the two highest level students – were intermediate EFL learners with a good general foundation in English and were placed in level 4 or 5 at ELI. Student K had the experience of living in the United States for one year when she was in high school, and therefore, she was able to regain the English that she had learned and demonstrated fluency on the pretest. In her utterances, she used connectives quite frequently – a typical speech trait of students who have lived in English-speaking countries (Yamashita et al., 1995). On the other hand, J had never been to an English-speaking country and his fluency level was not as high as K's.

However, his complexity score was the highest in the group on both tests. His speech included relative pronouns such as "a boy who wears...," adjective phrases such as "the boy in the sea," and many connectives. Moreover, in addition to trying to describe the picture by using prefabricated patterns, he tried to express his own thoughts by saying "I think...." The use of these expressions contributed to his ability to create longer sentences and more words per C-unit. Figures 14 and 15 show that both J and K have good foundations in English. However, across the two tests, little or no improvements were observed across any of the speaking performance evaluative dimensions (with the exception of K's fluency which improved).
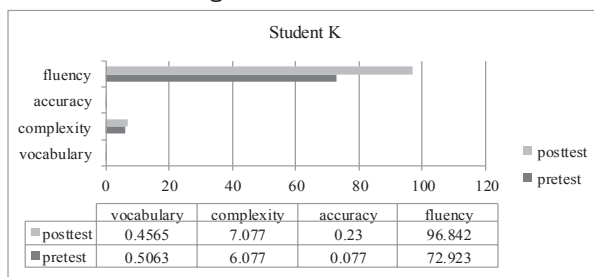
Figure 14. Student J

	vocabulary	complexity	accuracy	fluency
posttest	0.5714	7.778	0.222	46.667
pretest	0.4787	8.545	0.545	47.797

Figure 15. Student K

	vocabulary	complexity	accuracy	fluency
posttest	0.4565	7.077	0.23	96.842
pretest	0.5063	6.077	0.077	72.923
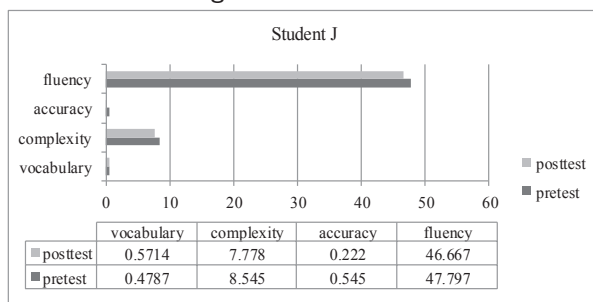
    Student A was a low-level EFL learner and placed in level
1 at ELI. She was not comfortable with talking in English at the
beginning. The picture descriptions consisted almost entirely of
one-word sentences on both tests, and there were no noticeable
improvements in the results. However, there was a slight increase
in fluency that could be interpreted as the development of self-
confidence in speaking English (Figure 16). Student B was placed
in level 2 at ELI. His utterances mainly consisted of one-word
sentences and noun phrases on both tests. He continued to make
the same mistakes such as "Boy eating a hot dog" instead of "A
boy is eating a hot dog" on both tests, indicating that he was not
cognizant of the grammatical mistakes he was making. Figure
17 shows that his vocabulary did not improve and his accuracy
level stayed the same. However, fluency and complexity slightly
improved, indicating that he may have also acquired a positive
attitude toward speaking.

    Student D was a low-level EFL learner. Like the other
students, improvements did not appear in vocabulary. The posttest
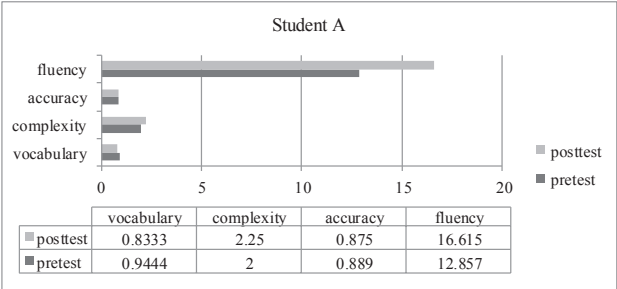showed slight differences in accuracy and complexity, but the
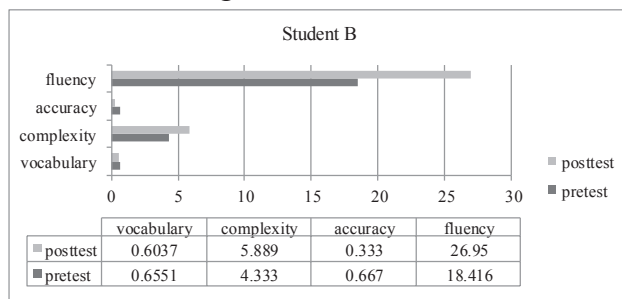
Figure 16. Student A

	vocabulary	complexity	accuracy	fluency
posttest	0.8333	2.25	0.875	16.615
pretest	0.9444	2	0.889	12.857

Figure 17. Student B

	vocabulary	complexity	accuracy	fluency
posttest	0.6037	5.889	0.333	26.95
pretest	0.6551	4.333	0.667	18.416

Figure 18. Student D

	vocabulary	complexity	accuracy	fluency
posttest	0.5972	5.538	1	34.56
pretest	0.7419	5.167	1	33.818

Figure 19. Student F

	vocabulary	complexity	accuracy	fluency
posttest	0.5735	6.181	0.364	38.491
pretest	0.7894	4.222	0.444	21.714
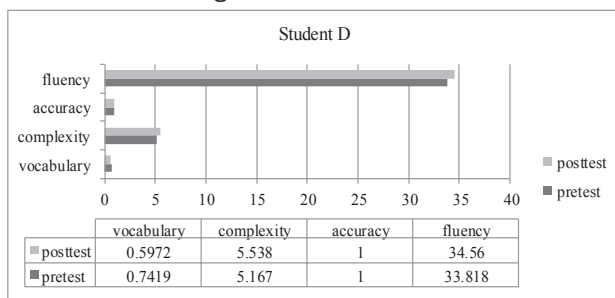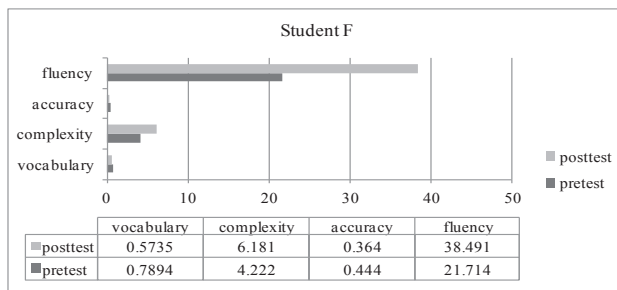
differences were negligible (Figure 18). He spoke with a higher level of fluency on the posttest, but did not learn the right order of English sentence structure and kept producing incorrectly ordered English sentences that mirrored Japanese sentence structure (e.g. "There is chair on the cat" instead of "There is a cat on the chair."). This is one of example of how fluency is related in some way to meaning and not to form.

Student F was a low-intermediate EFL learner. She had strengths in grammar and sentence structure, but was behind in listening and especially behind in oral production. She had a prudent attitude, and would not speak unless she was sure that what she was going to say was correct. She seemed to rehearse what she wanted to say before saying it. As Figure 19 shows, she made significant improvements in the areas of fluency and complexity.

7. Conclusion

This study examined how a three-week English study abroad program implemented at EWU in the United States in 2012 contributed to the improvement of KIU students' speaking performance. The statistical analysis focused on the four dimensions of speaking performance – vocabulary, complexity, accuracy, and fluency – and revealed that only vocabulary and fluency showed statistically significant improvements between the pretests and posttests. The changes in performance in the other two dimensions were inconsistent and did not reach statistically significant levels. However, distinct features were found in the students' attitudes

toward speaking.　In the picture description task of both tests, the students had one minute to talk, but they were allowed to speak as much as possible if they wanted.　Most of the students tried to produce longer utterances and actually spoke more voluntarily on the posttest than they did on the pretest.　It was apparent to the researcher that the students demonstrated higher levels of confidence in speaking English in the posttest.　Though these observations concerning confidence are difficult to quantify and demonstrate in numbers, this apparent attitudinal shift may point to one of the positive effects a study abroad experience may have on potential long-term improvements in speaking performance.

　　The second purpose of this study was to define the elements of a study abroad program that correlate with improvements in speaking performance.　Cultivating speaking ability is not a simple procedure, and it involves cultural, psychological, and linguistic factors.　As Yashima and Zenuk-Nishide (2008) pointed out, EFL learners may be able to develop their linguistic skills to some degree in an "at home" environment, though the development may be limited considering the complex nature of speaking.　Study abroad experiences in English-speaking countries could offer the unique opportunity for students to be exposed to numerous complex conversational situations that challenge them with their inputs, outputs, and interactive communicative strategies, thereby contributing to improvements in their speaking performance.

　　The third purpose of this study was to gain insights on the various preparations that should be made before a short-term study abroad program commences to further enhance its effectiveness.

Simply living and studying in an English-speaking country does not automatically result in improvements in linguistic ability.　It is considered impossible for college students aged 18 and over to acquire a language naturally and unconsciously during a brief period such as a short-term study abroad program.　As this study showed, the three-week program did not contribute to the development of speaking abilities other than vocabulary and fluency if the students did not have a good general foundation in English.　Before participating in a study abroad program, it is recommended that EFL learners should increase their vocabulary and learn grammar rules and structures in a conscious way.　A study abroad program should be considered as an opportunity to put one's knowledge into practice. Without a good foundation in English, the effectiveness of a study abroad program in improving speaking performance will be limited.

　　This research is a case study that focused on KIU students who participated in a three-week study abroad program in the United States, and though the study provided interesting findings that help us think about the factors that may help improve the effectiveness of short-term study abroad programs on speaking performance, we must also remember that the number of subjects in the study is too small to make generalizable claims about the general effectiveness of short-term study abroad programs on speaking performance.　Further studies are needed to further investigate the full effectiveness and potentials of short-term English study abroad programs.

References

Adachi, T. (2010). Purposes and benefits of undergraduate study abroad programs and their place within the curriculum. *Touyoueiwajyogakuinndaigaku Jinbun Shakaikagaku Ronshu, 28,* 77-91.

Eguchi, H. (2010). The effects of a short-term overseas English program for Hokusei Gakuen University English department students. *Hokusei Ronshuu(bun), 47,* 33-49.

Imai, H & Yoshida, T. (eds.). (2007). *HOPE Chuukousei notameno Eigo Supikingu tesuto [HOPE English Speaking Test for Junior and Senior High School Students]. Kyouiku Shuppan.*

Japan Student Services Organization (JSSO), 2011 Follow-up Research on Experienced Study Abroad 平成 23 年度「海外留学経験者追跡調査」報告書Ⅲ．留学実態調査／結果の詳細
Retrieved January 3, 2013, from
http://www.jasso.go.jp/study_a/documents/k_h23_04.pdf

Kawaguchi, C. & Kamimoto, T. (2003). Distinctive features of oral production by fluent and nonfluent EFL learners. *Language Laboratory, 37,* 21-36.

Koizumi, R. & Fujimori, C. (2010). An exploration of measures to detect changes in speaking performance: A case study based on picture description. *JACET Journal 50,* 81-91.

Kuno, H. (2011). Effect of a short-term (3-week) study abroad program on the English Proficiency of four-year college students. *Hokkaido Bunkyou Daigaku Ronshu, 12,* 127-145.

Matsumoto, M. (2010). English listening skill's improvement in the short-term overseas program. *Annual Review of English Learning and Teaching, 15,* 57-65.

Ministry of Education, Culture, Sports, Science, and Technology. (2012). Action Plan for University Reform 大学改革実行プラン－社会の変化のエンジンとなる大学

づくり－
Retrieved November 25, 2012, from
http://www.mext.go.jp/b_menu/houdou/24/06/_icsFiles/afieldfile/2012/06/05/1
312798_01_3.pdf#search='%E5%A4%A7%E5%AD%A6%E6%94%B9%E9%9D
%A9%E5%AE%9F%E8%A1%8C%E3%83%97%E3%83%A9%E3%83%B3'

Morita, M. (2010). How does a short term study abroad influence language
    learning strategies? – The case of the intercultural communication program at
    Yamagata University.　*Yamagatadaigaku Jinbungakubu kenkyunenpou, 7,* 23-36.

Sabet, M. (2007).　Are Study Abroad Programs Effective?　*Seigakuin daigaku
    Ronsou, 20 (2),* 177-184.

Sugita, Y. (2007). Effect of the dynamic listening and speaking method on
    the improvement of English speaking ability of nursing college students.
    Yamanashikenritsudaigaku Kangogakka Kiyou, 9, 55-65.

White Paper on Ministry of Education, Culture, Sports, Science, and Technology.
    (2011). Chapter 8: Improvement in International Exchange and Cooperation
    （国際交流・協力の充実）
    Retrieved November 23, 2012, from
    http://www.mext.go.jp/b_menu/hakusho/html/hpab201201/detail/1324676.htm

Yashima, T. & Zenuk-Nishide, L. (2006). The impact of learning contexts
    on proficiency, attitudes, and L2 communication: Creating an imagined
    international community.　System, 36(4), 566-585.

Yamashita, T., Yamamoto, S., & Viswat, L. (1994).　Acquisition of communicative
    competence through overseas study in America: A case study of eight English
    Japanese high school students, *Language Laboratory, 31,* 32-44.

Yokota, M. & Others. (2006). Kiro ni tatsu Nihon no Daigaku: Zenkoku
    4nenseidaigaku no Kokusaika to Ryuugakukouryuu ni kansuru Chousahoukoku.
    Hitotsubashi daigaku Kikan Ripojitori.　Hitotsubashi University Repository.

Appendix A: Students' Utterances during the Picture Description Task

	Pretest	Posttest
A	Chair three chairs, cat, sleeping dog, playing guitar (Who is playing the guitar?) girl enjoy in sea, boys eat hot dog, ship, three ship, bird ...table...umbrella. (20 words, 1'24'')	Cat on chair, playing guitar, eating hot dog, playing in sea, three ship, sleeping dog, three beach ...bird, three. (18 words, 1'5'')
B	Dog... the dog is sleeping under the tree. Boy eating a hot dog, and playing...she playing the guitar. Two boys eating hot dog. The boy ...(long pause)...the cat is standing on the chair. (32 words, 1'41'')	Two boys eats..eating the hot dog. The girl is playing the guitar, maybe singing. OK...two people is...playing in the sea...the white table, two cans on the table, white table, and umbrella, sun umbrella, three birds on the sun umbrella staying, and the dog is sleeping under the tree, that's all. (53 words, 1'58'')
D	There is chair on the cat...table on the drink, and sitting chair man is eating, and woman playing the guitar. Sea in the two people, three under the dog sleeping. (32 words, 55'')	First, there is...this picture of the left, tree under the dog. Dog is sleep, and there is table and parasol, over the parasol a few birds, and three chair, one chair is sitting boy, he's eating frankfurt, and second chair on the cat, and a boy, he's eating maybe sandwich, next people is woman. She playing the guitar. There is sea, three ships and girl and boy are swimming. (70 words, 2'05'')
F	Ah..this picture...sea side beach..picnic...two boys eating sandwich and hot dog, and girl is playing guitar, and cat, kitty on chair...three birds...umbrella, big umbrella, one girl and one boy is...are in the sea maybe calling three guys. (40 words, 1'45'')	...This picture is beach and there five kids, three of them are on the beach, and two kids are in sea, water, and three kids are enjoying maybe lunch and one girl playing guitar, and cat is standing on chair, dog is under the tree near the three kids, and there three boats on the sea, there are three birds on the parasol, and that's all. (66 words, 1'46'')

J	Ah...A boy who wears a orange shirt is eating hot dog. His face is so fun. And a cat is on the chair. The other boy who wear a white cap is also eat hot dog. A girl is playing guitar. I think the two boys is hearing her music...and I think their friends two friends is playing in the sea. A brown dog is sleeping under the tree. And a girl in the sea, a girl is playing in the sea and the boy in the sea is come...approaching to the, their friends. (101 words, 1'58")	There are five children in this picture. And a boy and a girl are playing in the sea, and the boy wearing yellow shirt is eating hot dog and other boys wearing orange shirt is also eating hot dog, and a girl is playing the guitar. A cat is on the chair. And the brown dog is under the tree...So I think they are school friends and they spend summer vacation. (72 words, 1'30")
K	It's a beach and there are three ships and there are one, two, three boys and one of them is swimming in the sea and there are two girls, one is playing the guitar and one...another is playing in the sea, and there is a cat and there is a dog sleeping under the tree. Two boys on a shore, and they are having hot dogs, maybe, and there are three sea seagulls on the...I should say parasol or...um... (81 words, 1'05")	One, two, three, four...five people at the beach, and three are boys and two are girls, and two boys are on the shore. They have table and parasol stuff and they are having hot dogs. There's a cat on the chair and one girl is playing the guitar and another girl is swimming with one boy. There are...ship, actually three ships are on the seashore, and there is a tree and under the tree, a dog is sleeping, and on the parasol there are three seagulls, that's it. (89 words,57")