

Developing Second Language Proficiency in a Study Abroad Context

Chieko Kawauchi
Sumiko Nagasawa

Abstract

This study attempts to investigate the development of English proficiency of four college students who took part in a one-year study abroad program in an American university. English proficiency before and after the study program was measured in terms of listening, reading, writing, speaking, and vocabulary. In addition, a questionnaire was also provided to examine students' perceptions of their own language development as well as their attitude changes in language use. The writing skill significantly improved in fluency, accuracy, and, in particular, complexity. The speaking skill in the interview test also indicated a significant increase in all the categories (accent, grammar, vocabulary, fluency, and comprehension). The listening skill changed favorably, but the gains were not statistically significant in spite of the students' high self-evaluations on listening development. The improvement in reading skill was not obvious. Development of English proficiency for the individual students varied. The most proficient student before the program demonstrated the least improvement, while the least proficient student showed the largest increase.

Introduction

Kurume University started its exchange program with Southern Oregon University in 1998, and for the first time four Kurume University students were selected and sent to take part in a one-year-study program in August of that year. This study attempts to investigate the development of their English proficiency as a result of the program.

A number of previous studies have looked into the effects of studying abroad on the development of language proficiency and on the affective changes of the students who took part in the study abroad programs. A summary of the major studies is displayed in Table 1.

Table 1

Summary of Studies on Second Language Development in the Target Countries

	Subject	Period	L2	Residents	Major findings
Yamane (1985)	43 English majors	14 week intensive course	English	Homestay	Listening comprehension was significantly higher than pre-study abroad and students at home
Kikuchi (1988)	108 (91 English & 17 non-English majors)	5 weeks	English	Dorm with homestay for 3 to 7 days	Listening significantly improved in the English majors but written test (grammar /reading) didn't; Learners have both integrative and instrumental motivation
Lennon (1989)	4 advanced English learners	6 months	English	3 in dorm & 1 in town	Students applied different production strategies under different circumstances; Seeking out feedback and practicing new linguistic items
Lennon (1990)	4 advanced English learners	6 months	English	3 in dorm & 1 in town	Reduction of filled pauses, repetitions and pause time, and increase of speech rate were associated with improvement of perceived fluency
Yamamoto (1992)	30 junior college students	4 weeks	English	Homestay	Dictation was significantly better than that of control group and functioned as a better indicator for total English proficiency; No improvement in grammar, cloze and listening

The findings of the previous research vary according to the research focus, the period of the study abroad programs, the student's proficiency levels, and the types of residences available to the students.

Iwakiri (1993)	205 college students	5 weeks	English	Homestay	Listening, structure and reading skills improved; Became more extrovert and gave more self-expressions after the study abroad
Yashima, Yamamoto & Viswat (1994)	8 high school students	1 year	English	Homestay	No progress in grammar and vocabulary levels; Used more discourse markers to make coherent text, which significantly correlated with NS rating
Freed (1995)	15 (Novice: 1, Intermediate :14)	1 semester	French	Unknown	Students spoke significantly faster and had a longer speech run than students at home; Fewer dysfluent silent pauses and fewer non-lexical pauses
Lafford (1995)	29 college students	1 semester	Spanish	Unknown	Used broader repertoire of communication strategies than students at home; a variety of fillers, connectors & back-channel signals; More self-repairs than repeats
Brecht, Davidson & Ginsberg (1995)	658 Russian majors	4 months	Russian	Unknown	Men were more likely to gain on listening and the OPI than women; Younger students gained on listening; Grammar & reading showed significant predictive value for speaking & listening
Towell, Hawkins & Bazergui (1996)	12 advanced French learners	6 months	French	Unknown	Increase in fluency is the result of length and complexity of linguistic units (speech run) which are uttered between pauses
Yashima (1997)	118 high school students	1 year	English	Homestay	No correlations between English proficiency and extroversion; Extroversion as an indicator of students' satisfaction in interpersonal relationships and communication effectiveness
Yashima & Viswat (1997)	16 high school students	1 year	English	Homestay	Fluency improved with higher speech rate and fewer silent pauses; False starts and repetition resulted in lower NS perceptions of fluency
Isoda (In press)	39 non English majors	1 month	English	Dorm	Listening significantly improved and personality changed into being more extroverted and confident; No improvement in reading and structure

Conflicting results are seen in terms of listening: Some studies showed significant progress (Yamane, 1985; Kikuchi, 1988; Iwakiri, 1993), while others didn't (Yamamoto, 1992; Isoda, in press). It appears that there isn't a great deal of improvement in grammatical accuracy even in long-term programs in a target country (Yashima, Yamamoto & Viswat, 1994, Yashima & Viswat, 1997), let alone short-term programs (Kikuchi, 1988; Yamamoto, 1992; Isoda, in press). The level of fluency, however, increases in all cases as a result of communication-oriented programs and of living in the target country where opportunities to use the language are abundant. To be more specific, learners increase their speech rate (Freed, 1995; Yashima & Viswat, 1997) or employ longer speech runs (Towell, Hawkins, & Bazergui, 1996) with fewer silent pauses (Lennon, 1990; Yashima & Viswat, 1997), employing discourse markers (Yashima, Yamamoto, & Viswat, 1994) or formulaic chunks (Towell et al., 1996), and a variety of production strategies (Lennon, 1990; Lafford, 1995). Moreover, the study abroad contributes not only to the development of various linguistic aspects of language but also to intercultural understandings and to changes of personality and attitude into extroversion (Kikuchi, 1988; Isoda, in press; Yashima, 1997).

Many of the previous studies, however, have focused on short-term study abroad programs, ranging from 4 weeks to one semester in length. Long-term study programs in which students take part in formal instruction are rarely seen in the literature. Five of the studies mentioned above investigated the effect of the long-term study abroad, ranging from six months to one year, but their subjects were, in fact, used repeatedly in the various research questions. Lennon's studies (1989, 1990) examined four German advanced learners of English, who didn't, however, receive any formal language instructions in the target country, but, rather, they were exposed to informal natural

learning for communicative purposes. On the other hand, Yashima, Yamamoto, & Viswat (1994), Yashima & Viswat (1997), and Yashima (1997) focused on the same group of Japanese high school students who resided with host families and attended local high schools for one year. To the best of our knowledge, there are no studies which examined the effect of long-term study programs for Japanese college students.

Moreover, the previous studies have limited the effects of study abroad programs to the communicative aspects of language, specifically listening and speaking. This might be quite reasonable because developing aural and oral skills in an authentic speech environment is one of the main goals for most programs. Consequently, the development of the learner's writing skill is rarely studied, even though writing is another important productive aspect of language performance which will reflect language growth and discriminate learners' proficiency levels (Hirano, 1991, 1993; Ishikawa, 1995; Larsen-Freeman & Strom, 1977).

The purpose of this study is to investigate the development of English proficiency for the students who were sent from Kurume University in a one-year-study abroad program. The study is concerned with not only the listening, speaking and reading skills on which many previous studies have focused, but also the writing skill, its development in fluency, accuracy, and complexity. Furthermore, we will look at the post-study-abroad vocabulary level. After they completed the program, the students were asked to fill out a questionnaire about the development of their language skills and their attitude changes in language use. Before going into the study, some theoretical and developmental issues on writing will be discussed in the following section.

Development of Writing for Japanese Learners of English

In EFL pedagogy such as in Japan, output-based teaching focusing on writing and speaking is still subordinate to input-based teaching with the emphasis on translation and grammar-oriented reading. Writing is limited largely to the sentence level as evidenced in many textbooks for high school and college students. Writing courses with the emphasis on essay and expository writing are not offered for the majority of students except for those who study English as their majors. According to Watanabe's (1990) survey, the skill Japanese college students want to develop most is speaking, but reading is still the primary concern of 80% of college teachers. Interestingly though, only 3% of the students in Watanabe's survey responded that writing is one of their primary goals. Because of the rapid advance of communication technology, however, people have realized more than ever before the need for the writing skill. For example, some recent studies have brought about a new genre which has been described as "Internet English" and have promoted new teaching syllabi for writing through the computer assisted language learning (CALL) system and the Internet (Asano and Saito, 1996). Some studies have begun to examine various features of writing styles observed in "Chat" and "E-mail" (Gonzalez-Bureno, 1998; Hirota, 1999).

Writing is, indeed, a latecomer to the EFL teaching in Japan, and more studies are needed to examine how Japanese learners develop writing skills both in the EFL and ESL contexts. First of all, we would like to review some influential studies on writing which have dealt with Japanese students as their subjects. By doing so, it will be easier to recognize some of the problems encountered in the development of the writing skill.

Using as many as 134 Japanese college students, Robb, Ross, & Shortreed (1986) evaluated the effects of different types of feedback

on error in EFL learners' writings. The four different methods of feedback are; (1) correction of all categories of lexical, syntactic and stylistic errors, (2) coded feedback of error types alone (3) uncoded but marked with a yellow highlighter and (4) marginal feedback only with the total number of errors written in the margins of the student's paper. One of the major findings was that there were no significant differences among these four methods in terms of helping students in improving accuracy. Corrective feedback was effective for improving fluency to some extent. This improvement was attributed to the weekly writing assignments, though. The study concluded that a less-time-consuming method of directing student attention to surface error may suffice, and that improvement of writing was independent of the types of feedback.

What, then, will facilitate the improvement of writing, if the types of feedback do not count? As mentioned in Robb et al. (1986), weekly writing brought about writing fluency which was measured by the number of words per composition. Briere (1966) also demonstrated that writing quantity facilitated writing quality, and Kawauchi (1999) supported this finding by analyzing the writing samples of low-intermediate Japanese students, particularly in the case of picture description task. In her study, a weekly writing activity in class and a weekly homework writing assignment significantly increased both the writing fluency and writing accuracy using only limited feedback from the instructor. It seems that the writing skill will be developed by writing more frequently and writing a larger quantity. In fact, the quantity of writing measured in the total number of words is found to have a discriminatory power, and is considered to be one of the indicators for the improvement, particularly in the case of picture description (Hirano, 1991).

When students are engaged in writing, what factors facilitate or

inhibit their writings? Many students say that when writing essays, they first write in Japanese and then translate it into English. In this type of writing, it is easy to imagine that transfer from L1 in syntactic, semantic, and stylistic structures will definitely influence their productions. Kobayashi & Rinnert (1992) provided some clues to support this supposition. They investigated Japanese college students' compositions written by using two differing writing processes; one was writing first in Japanese and then translating into English, and the other was composing directly in English. Forty-eight Japanese college students enrolled in English Composition wrote both of the two kinds of compositions. In terms of the quality of content, organization, and style, the lower-level learners benefited from translation, but the higher-level learners did not. Although overall syntactic complexity was greater in translation, the higher-level students tended to make more errors, because translation interfered with their intended meaning more frequently than direct writing did. The study showed that the most significant error type was lexical choice, and the next significant type was awkward structures.

Lexical choice errors were also one of the major factors in decreasing "communicability" in the writings of Japanese high school students (Harada, 1998). "Communicability" is defined here as the degree to which native speakers can comprehend when they read L2 compositions, even though there are some errors (Ando, 1991). Inappropriate lexical choice most strongly contributed to low communicability followed by a lack of explanation and coherency. These findings suggest that vocabulary use significantly affects the evaluation of compositions.

Another factor which contributes to the evaluation of writing is accuracy. Larsen-Freeman and Strom (1977) first proposed error-free T-unit analysis to measure the proficiency level of advanced ESL

learners. Hirano (1991) supported the error-free objective measures as a significant index for the writings of Japanese college students. By analyzing a total of 160 English compositions from 80 students, she drew the conclusion that the most appropriate indicator for Japanese EFL proficiency was the total number of words in an error-free T-unit.

Looking at another aspect, Ishikawa (1995) focused on the developmental changes in the writings of Japanese college freshmen (N=28) and found that the most vital and efficient measures for these learners are, first, the total number of words in error-free clauses, and second, the number of error-free clauses per composition.

In our review we have found that many studies agree that writing is one of the most effective indicators in assessing learners' proficiency levels and their development. As we will see in the comments on the questionnaire, the students said that they had a large number of writing assignments as well as reading in the study abroad program. This is a large contrast, when we compare a limited opportunities for writing in the current language curriculum in Kurume University. Writing is, therefore, the obvious skill to look at to examine the developmental changes for our students who studied English under a formal instruction at Southern Oregon University.

Method

Subjects

There were four students who participated in the one-year study abroad program at Southern Oregon University (SOU). These four students are called A, B, C, and D, in the study. Based on the initial TOEFL scores, the students were categorized into advanced (Subject A), intermediate (Subjects B and C), and low-intermediate (Subject D).

Subject A who satisfied the language requirement (TOEFL 520) entered the regular college program directly. Subjects B, C, and D attended an intensive English program at the American Language Academy (ALA) operated on SOU campus. After attending three 10-week terms in the ALA, they passed TOEFL with the required score of 520, enabling them to register for mainstream courses at SOU. Subjects B and C attended a one 10-week term while Subject D attended two terms including summer sessions. Therefore, we assume that the quality and quantity of language exposure between Subject A and the other three subjects might have been different during the study abroad, at least for the initial three terms at SOU.

Table 2 displays the personal information of each subject.

Table 2
Personal Data for TOEFL Scores and Study Situations at SOU

	Initial	Level	Latest	Courses attended	Language experiences; Majors at Kurume; Type of Residence in US
Subject A	547	Advanced	537	Started in the main stream majoring Political Science	1 year study abroad in high school; International Politics; Dorm
Subject B	460	Intermediate	527	Intensive English Courses (ALA); Main stream in the final term	No previous experiences of study abroad; English; Dorm & homestay
Subject C	460	Intermediate	540	Intensive English Courses (ALA); Main stream in the final term	No previous experiences of study abroad; English; Dorm & homestay
Subject D	433	low-intermediate	540	Intensive English Courses (ALA); Main stream in the final term	No previous experiences of study abroad; International Politics; Dorm & shared room with a Bulgarian friend

Tests

The students were tested on listening, reading, writing, and speaking in both the pre- and post-tests. The vocabulary section was included only in the post-test. The students were also asked to fill out the questionnaire about their language development and attitude changes.

Listening:

The test consisted of 15 short dialogues, one long dialogue, one monologue, and two news passages. There were a total of 40 multiple-choice questions: 15 for the short dialogues, 4 for the long dialogue, 6 for the long monologue, and 15 for the news passages. All the questions except for the news passages were taken from a TOEFL textbook (Kishigami, 1997). The questions about the news passages were made from TV news programs broadcast by ABC and CNN. Speech rates for these questions were rather high: on average, 172 words per minute in the TOEFL questions and 182 words per minute for the TV news.

Reading:

Five reading passages with the average of 230 words each were used. Each passage taken from a TOEFL exercise book (Tanaka, 1997) included five questions, making the total number of questions 25. Like the listening test, these passages were made into exercises following the TOEFL format for test-takers.

Vocabulary:

The vocabulary test was made following the frequency levels identified by Laufer and Nation in their 1999 investigation into the learners' productive vocabulary knowledge and vocabulary growth. We adopted three of these frequency levels: the 2000-level, the 3000-level,

and the University Words List level (UWL) for our use. The test for each level was made up of 18 items. In each item the students were asked to complete a statement by filling in the blanks with the target vocabulary word, where the first few letters of the target word were provided (e.g., I'm glad we had this opp_____ to talk).

Writing:

The students were assigned a specific topic either "A Memorable Childhood Experience" or "A Memorable Person from my Childhood" to write about. They were given 40 minutes to write in both tests. Direct L2 writing was employed, and the use of dictionaries was not allowed.

All the written passages were examined in terms of fluency, accuracy, and complexity. To determine fluency, we used the total number of words per writing. Accuracy was examined using the error-free clause analysis (Ishikawa, 1995), which consisted of the total number of words in each error-free clause and the number of error-free clauses per composition. Clauses include simple independent clauses, subordinate clauses, coordinate clauses, coordinate clauses with subject deletion, and embedded clauses. Three native speaker (NS) teachers evaluated all the clauses as either "correct/acceptable" or "incorrect." When there were cases of ambiguity, they were asked to give the most favorable interpretation. We used only those clauses which two of the three teachers agreed were error-free. Complexity was investigated by the ratio of clause per c-unit. Foster and Skehan (1996, p.310) state that "A c-unit is defined as each independent utterance providing referential or pragmatic meaning. Thus, a c-unit may be made up of one simple independent finite clause or else an independent finite clause plus one or more dependent finite or nonfinite clauses." For example, the following sentence has 3 clauses and 1 c-unit, and therefore, the

clause ratio accounts for 3:

"I have one memorable person who was a music teacher when I was a junior high school." (Subject B in the post-test)

Speaking:

Two forms of the speaking test were examined; narratives reflected by picture description and dialogues reflected by interview. In the narratives the students were asked to describe two drawings, each of which was made up of four sequential pictures. The students were shown each of the drawings and given one minute to prepare before talking about it for two minutes. We used the first picture as a warm-up and the second picture for the analysis.

The students were also interviewed individually as a part of the pre- and post-tests. These tests (approximately 15 minutes long each) were administered by a NS teacher of English. The questions to be asked were carefully controlled in both tests in an attempt to have the interactions with the interviewer be the same for all students.

All the oral productions were tape-recorded and evaluated by the same three NS teachers of English who rated the writing samples. The criteria for the oral production was the Foreign Service Institute (FSI) oral proficiency rating scale. The FSI numerical rating procedure evaluates accent, grammar, vocabulary, fluency and comprehension. Each category has a scale of 1 to 6 with detailed rating criteria (Vallete, 1977, pp.159-160). For example, "Grammar level 1" is described as "Grammar almost entirely inaccurate except in stock phrases" and "Grammar level 6" as "No more than two errors during the interview." The average scores of the three NS teachers were used for each category.

In order to examine linguistic growth in the speaking skill in greater detail, we transcribed the recorded narratives and examined them for

fluency, accuracy, and complexity using the same criteria that were applied to the writing. The narratives provide a clearer indication of linguistic growth than the interactive dialogues in the interview. The reason for this is that the dialogue usually involves feedback and sometimes linguistic supports from the interlocutor as pointed out by Lazaraton (1996).

Questionnaire:

The questionnaire consisting of nine questions was given to the students immediately before the post-test. The students were asked to evaluate themselves in the development of each language skill (listening, reading, writing, speaking). They were also asked to evaluate their attitude changes in terms of making mistakes, the degree of embarrassment, confidence, interests, and enjoyment when using English. We used a 7-interval Likert-type scale with anchors ranging from "strongly disagree" to "strongly agree."

Results and Discussion

1. Listening and Reading

First, the results of listening and reading tests will be shown and discussed. Table 3 displays the results for the pre- and post-tests as well as the results of t-tests (one-tailed) between the tests.

The listening scores for all the students increased in the post-test, although the difference between the tests failed to attain the conventional significance level ($p=.05$). The results of the questionnaire (See Table 9), by contrast, showed that all the students except for Subject B, thought their listening skills had developed to the highest degree (Scale 7). The lowest scores by Subject B (26 on the post-test) corresponded to her evaluation of her listening development in the

Table 3
Scores Changes in Listening, Reading, and Past-tense

	Listening (N= 40)		Reading (N= 25)	
	Pre	Post	Pre	Post
Subject A	34	36	16	14
Subject B	23	26	12	16
Subject C	19	36	11	6
Subject D	24	27	9	8
Average	25	31.25	12	11
(SD)	(6.38)	(5.5)	(2.94)	(4.76)
<i>t</i> -value	<i>t</i> =1.74 (<i>p</i> =.090)		<i>t</i> =0.53 (<i>p</i> =.315)	

questionnaire which she rated the lowest of all (Scale 4). The largest gain was found in Subject C whose score was almost twice on the post-test. This supported her comment in the questionnaire where she stated that “Intensive exposure to English everyday and participation in various activities helped a lot.” Subject A also commented that “It was very helpful to listen to teachers’ lectures and watch TV everyday, especially when paying attention to the programs I’m interested in.”

In spite of those favorable perceptions of their listening development, there appears to be a gap between the test scores and their perceptions. There could be several reasons for this gap. First, it might be due to the differences of topics. The listening test included diverse topics ranging from daily dialogues to social issues, while the topics they had in the study abroad program and among friends might have been rather limited. Second, unlike the real world communication, speakers in the test tape were not present, and consequently the lack of interaction made it harder for listeners to understand. This seems

to be supported by the fact that these students showed higher listening comprehension in the interview tests (See Table 6).

In the reading test, however, the level of only one student increased, while the levels of the rest of them decreased. It was an unexpected result when we consider the fact that they had a lot of reading assignments during the study abroad program. This raises the question that there might be a qualitative difference between reading for meaning required in reading assignments for class and reading for accuracy required in the TOEFL. There was a rather high correlation between the students' self-evaluations on reading development and the reading scores in the post-test ($r=.87$). This appears to imply that the students' self-evaluations on the reading development are reliable and are backed by the scores on the reading test.

2. Vocabulary

Vocabulary levels of the students after the study abroad program are shown in Table 4.

There is a tendency for the students to make lower scores as the vocabulary levels increase. Only Subject A succeeded in passing the 2000-level (15 correct out of 18), while all the rest of the students failed.

Table 4

Results of the Three Levels of Vocabulary Tests (% correct)

	2000-level	3000-level	UWL
Subject A	18 (100%)	10 (55.6%)	9 (50%)
Subject B	9 (50%)	8 (44.4%)	5 (27.8%)
Subject C	14 (77.8%)	8 (44.4%)	8 (44.4%)
Subject D	13 (72.2%)	5 (27.8%)	8 (44.4%)
Average	10.7 (76.4%)	7.5 (41.7%)	7.5 (41.7%)

In order to examine how the vocabulary relates to reading and listening, the total scores of the three levels were compared with the reading and listening scores of the post-test. There was a rather high correlation between vocabulary and listening ($r=.84$), but, interestingly enough, no correlation between vocabulary and reading ($r=-.22$).

We also examined the vocabulary level of the three who were chosen as 1999 study-abroad students. Their average scores were 10.7 (59.4%) in the 2000-level, 3.67 (20.4%) in the 3000 level and 5.33 (29.6%) in the UWL. In order to see the difference between the 1998 students and the 1999 students, unpaired t-testing was carried out. The results of the t-test revealed that there was a significant difference in the 3000-level ($t=2.670$, $p=.044$). No significance was found in the 2000-level ($t=1.28$, $p=.255$) or the UWL level ($t=1.36$, $p=.231$). Since we could not obtain the results of the pre-test for the 1998 subjects, it might be dangerous to attribute this significant difference directly to the study abroad, but considering the fact that the 1999 students' initial TOEFL scores ($\bar{x}=486.7$) were higher than the current students, it is fair to assume that the significant increase in the 3000 vocabulary level was due to the one-year study program.

3. Writing

The students were given a pre- and post-writing tests. We collected a total of eight writing samples, four samples each for the pre- and post-tests. Table 5 displays the results of pre- and post-tests in terms of fluency, accuracy, and complexity, as well as the results of t-test (one-tailed) between the two writing samples.

The results show fairly uniform improvement for all the categories, although, except for complexity, these increases failed to attain the statistical significance. However, considering the fact that the number of the subjects is very limited, we think these probability levels

Table 5
Results of Writing Tests on Fluency, Accuracy, and Complexity

	Fluency		Accuracy 1		Accuracy 2		Complexity	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Subject A	248	276	186	195	37	35	1.61	1.83
Subject B	257	279	149	196	25	30	1.32	1.83
Subject C	141	393	31	284	5	43	1.18	1.30
Subject D	169	356	60	222	9	37	1.11	1.77
Average	203.7	326.0	106.5	224.5	19.0	36.25	1.19	1.69
(SD)	(57.6)	(58.0)	(73.0)	(41.75)	(14.79)	(5.38)	(0.09)	(0.26)
<i>t</i> -value	<i>t</i> =2.18		<i>t</i> =2.12		<i>t</i> =1.83		<i>t</i> =3.76*	
	<i>(p</i> =.062)		<i>(p</i> =.062)		<i>(p</i> =.082)		<i>(p</i> =.017)	

Note Fluency: The total number of words per writing; Accuracy 1: The total number of words in error-free clauses; Accuracy 2: The number of error-free clauses; Complexity: Ratio of clauses per c-unit; SD: Standard deviation.

(*p*=.062) are significant enough to assume that the difference favored the post-test.

First, improvement in complexity was worth noting. Accuracy scores also increased almost twice in both measures. It is fair to say that in the post-tests the students wrote more fluently and more accurately, employing more complex syntactic structures.

The features of writing development are, however, not the same among the subjects. Subject B showed the greatest percentage increase in complexity (138%), Subject C in accuracy (916% in accuracy 1 and 860% in accuracy 2), and Subject D in fluency (211%). However, Subject A, who had the highest TOEFL score before the program, remained almost the same in all the categories with the average increase of 10%. The following examples illustrate the contrast between the two writing samples in terms of fluency, accuracy, and complexity. Subject B wrote the following introductory sentences in the pre-test:

My most memorable person is a music teacher. I first met her when I was a junior high school student. Her name is Ms. Takako Mori. She was very kind and taught me how nice singing was. Besides, She was familiar with not only music but also math or English... (48 words; 8 clauses; 5 c-units)

In the post-test, Subject B began the same paragraph with more complex structures as follows:

Each person have each memorable childhood experience even if it is good one or bad one. I have one memorable person who was a music teacher when I was a junior high school. She taught me precious thing in my life. (41 words; 6 clauses; 3 c-units)

Subject C started her story about the most impressive event in the pre-test as:

I like sports and exercises very much. I think it is influenced by the school before elementary school. I don't remember the details but in that school, I enjoyed very much. (32 words; 5 clauses; 4 c-units)

In the-post test, Subject C, whose gain in accuracy was largest among the four students, provided detailed descriptions with 84 words. Although hers was the least gain in complexity, she employed two embedded relative clauses which were not used in the pre-test. The following shows the first few sentences:

The kindergarden which I went was very very good. I have a lot of good memories there. In Spring, we touched rabbits which we had in the kindergarden. In summer, we went swimming in a river and we slept over in the kindergarden for one day... (46 words; 7 clauses; 5 c-units)

Only Subject D wrote totally different stories: a piano contest in the pre-test and the best friend in the post-test whom she met while studying in the SOU. Scores for all the categories were strikingly

increased, indicating that her total writing skill had improved greatly during the study abroad program. Introductory sentences in the pre-test were:

It is pianno contest that a memorable experience for me. When I was ten years old, my teacher said to me, "Let's practice for pianno contest together!" I had learned it since five years old and every year I attended to a contest in my town... (47 words; 7 clauses; 5 c-units)

The following is the first paragraph in the post-test.

On August 1, I just turned 22 and I've met several people who was influenced on my life so far. However, the only person who changed my character and the way to think is Anelia, who is from Bulgaria - I met her in America, and she is probably one of my best friends. (54 words; 8 clauses; 5 c-units)

Finally, Subject A was the most proficient learner before the study abroad program. The following are the first parts of the introductory paragraph for the pre- and post-tests, respectively:

The memorable person from my childhood is a elementary school's teacher, Mr. Kodama. When I was 3rd grade, I was in his class. He was very great man. The first day of his class, he played a few small magic for us... (42 words; 5 clauses; 4 c-units)

I had a great teacher when I was 3rd and 4th grade, who is Mr. Kodama. He was about 40 years old. I remember the first day of school in third grade. He introduced himself to us and showed us a trick... (41 words; 7 clauses; 5 c-units)

Such a small increase in Subject A's writing could be attributed to the ceiling effect, a plateau which she might have achieved before the program. Another possibility is that Subject A didn't need to take any of the intensive English course provided by the ALA, and unlike the other three students, she might not have had opportunities to improve her

writing systematically.

4. Speaking

Speaking was measured by two types of oral production, interview and picture description. Table 6 shows the results of interviews made in the pre- and post-tests.

Table 6
Results of Interview Test Based on FSI

	Accent		Grammar		Vocabulary		Fluency		Comprehen-sion	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Subject A	4.00	4.33	3.67	4.00	3.33	4.33	2.67	4.33	3.67	4.00
Subject B	3.33	3.67	3.00	3.67	3.00	3.67	2.67	4.33	3.67	4.67
Subject C	3.67	4.00	3.00	4.33	2.33	4.00	3.33	4.67	3.67	4.00
Subject D	3.33	4.67	3.33	4.00	3.00	3.67	3.00	4.33	3.67	4.67
Average	3.50	4.17	3.25	4.00	2.91	3.92	2.92	4.42	3.67	4.34
(SD)	(0.43)	(0.43)	(0.32)	(0.27)	(0.44)	(0.32)	(0.32)	(0.17)	(0)	(0.39)
<i>t</i> -value	<i>t</i> =2.80*		<i>t</i> =3.58*		<i>t</i> =4.16*		<i>t</i> =15.96**		<i>t</i> =3.44*	
	<i>(p</i> =.034)		<i>(p</i> =.018)		<i>(p</i> =.012)		<i>(p</i> =.000)		<i>(p</i> =.021)	

Significant differences were found in all the categories and confirmed our suppositions that the students greatly improved their speaking skills. The largest gains are found in fluency, followed by vocabulary and grammar. We found that the students were able to comprehend what was said to them, and they were able to respond fluently using a large vocabulary.

Development of speaking, however, varied among the subjects and among the categories. Subject A showed the greatest increase in fluency (162%) followed by vocabulary (130%). Subject B also improved most in her fluency (162%) followed by comprehension (127%). The

largest increase in vocabulary was found in Subject C (172%) whose grammar was also greatly improved (144%). Subject D indicated overall increases in all the categories, particularly in fluency (144%) and accent (139%).

The results of picture description are shown in Table 7. As in the interview analysis, all the categories except for comprehension were examined. Since picture description is narrative in nature, the comprehension category was taken out.

Table 7

Results of Picture Description based on FSI

	Accent		Grammar		Vocabulary		Fluency	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Subject A	4.00	4.33	4.00	4.33	4.33	4.00	4.33	4.00
Subject B	3.67	4.33	3.33	4.67	3.33	4.33	4.00	5.00
Subject C	3.67	4.33	3.00	4.33	3.00	3.33	3.33	3.67
Subject D	3.67	4.67	3.33	4.67	2.67	4.00	3.00	5.00
Average	3.75	4.42	3.42	4.50	3.33	3.92	3.67	4.42
(SD)	(0.17)	(0.17)	(0.42)	(0.20)	(0.72)	(0.42)	(0.61)	(0.69)
<i>t</i> -value	$t=4.84^{**}$ ($p=.008$)		$t=4.31^{*}$ ($p=.012$)		$t=1.58$ ($p=.106$)		$t=1.52$ ($p=.113$)	

All the raw scores increased in the post-test, indicating that narrative skills also improved as a result of the study abroad program. The results in the accent and grammar categories showed significant improvement, but vocabulary and fluency did not. In comparison with the interview, improvement in the picture description task was not so obvious. Detailed examination shows no statistical differences between the post-test interview and the post-test picture description in terms of accent ($t=1.57$, $p=.108$), vocabulary ($t=0.01$, $p=.497$), and fluency

($t=0.01$, $p=.498$). Grammar in the picture description was, however, significantly higher than that of the interview ($t=2.32$, $p=.05$). From these findings, we have confirmed that overall speaking skills are not so different and that these two types of speaking tests are successfully gauging the current speaking levels of the students.

The small increase in the narrative skill may be due to the task difficulty associated with the picture description. Closed tasks such as picture description require students to focus on the storyline and to employ appropriate vocabulary to depict it, and accordingly this might have sacrificed fluency (Kawauchi, 1998). This phenomenon is the kind of "tradeoff" effect explained by Foster & Skehan (1996) who focused on the tradeoff between complexity and accuracy. Two way communication, however, as seen in the interview, appears to facilitate speaking fluency because there is sufficient feedback from the interlocutor and back-channeling from the subjects.

Subject D proved to be superior to the other three subjects in terms of gains in all the categories. She scored lowest on the pre-test in fluency, but her score in that same category was the highest on the post-test. Subject A, on the other hand, showed decreases in the scores of the vocabulary and fluency in the post-test. Subject D's great development in writing as well as speaking would lead to the assumption that less proficient learners should benefit from a study abroad program more than highly proficient learners.

In order to examine the syntactic growth of oral production more clearly, the transcriptions of picture descriptions were re-analyzed quantitatively in terms of fluency, accuracy and complexity as seen in Table 8.

All the scores in the post-test were higher than those of the pre-test, but only fluency attained the significance level ($t=2.36$, $p=.049$). The results contradicted those of NS teachers' ratings shown in Table 7,

Table 8

Results of Picture Description in Fluency, Accuracy, and Complexity Based on Transcription

	Fluency		Accuracy 1		Accuracy 2		Complexity	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Subject A	59.0	74.5	63	51	10	6	1.50	1.15
Subject B	58.0	74.0	49	57	9	10	1.55	1.33
Subject C	41.0	48.0	3	6	1	2	1.29	1.25
Subject D	41.5	90.5	14	27	2	5	1.00	1.43
Average	49.88	71.75	32.25	35.25	5.50	5.75	1.34	1.29
(SD)	(9.97)	(17.57)	(28.37)	(23.41)	(4.65)	(3.30)	(0.25)	(0.12)
<i>t</i> -value	<i>t</i> =2.36*		<i>t</i> =0.56		<i>t</i> =1.07		<i>t</i> =0.26	
	<i>(p</i> =.049)		<i>(p</i> =.309)		<i>(p</i> =.439)		<i>(p</i> =.404)	

Note Fluency: The total number of words per writing; Accuracy 1: The total number of words in error-free clauses; Accuracy 2: The number of error-free clauses; Complexity: Ratio of clauses per c-unit; SD: Standard deviation.

where accent and grammar improved but vocabulary and fluency did not. Fluency measured by words per minute (Table 8) may not be equivalent to fluency measured by FSI criteria (Table 7). The correlation between the two ratings for fluency was, however, significantly high in the pre-test ($r=.95$) and rather favorable in the post-test ($r=.80$). Based on this evidence, it was considered that the number of words produced strongly influenced the evaluation by NS teachers.

However, the forgoing discussion may perhaps provide some limitations in relying solely on test scores to assess linguistic development. More detailed examinations will be necessary to explain the qualitative differences between these two rating measures.

5. Questionnaire

Table 9 displays the students' responses to the questions about the

Table 9
Students' Responses to Questionnaire

	Subject A	Subject B	Subject C	Subject D
1. Development of the listening skill	7	4	7	7
2. Development of the speaking skill	3	7	6	6
3. Development of the reading skill	6	7	5	4
4. Development of the writing skill	5	7	6	5
5. Not embarrassed when using English	6	7	5	7
6. Not afraid of making mistakes	5	7	6	7
7. Confidence in English	4	7	5	7
8. Enjoy speaking	5	6	7	7
9. Interest in English	7	7	7	7
Average	5.33	6.56	6.0	6.33

Note 1: Strongly disagree; 7: Strongly agree

development of their language skills (1-4) and their attitude changes in language use (5-9).

At the first glance, we could say that all the responses are highly positive, and that is typically true for the less proficient students like Subjects B and D. These students' attitudes toward using English were highly favorable, indicating that they are not afraid of making mistakes and have come to enjoy speaking it without feeling any embarrassment.

Subjects A, C, and D thought their listening skills had improved the most, while Subject B evaluated it as lowest, although she is completely satisfied with the development of the other three skills. The responses of Subject A and D are noteworthy. As shown in the previous section, the language development of Subject A was rather small, but that of Subject D was quite remarkable. The questionnaire revealed that Subject A was not very satisfied with the language use ($\bar{x}=5.3$) and showed the least positive attitude changes ($\bar{x}=5.4$). Subject D

responded the most positively in all the attitude categories ($\bar{x}=7$), although she still gave herself a low evaluation for language skills ($\bar{x}=5.5$). One possibility for the marked improvement in Subject D's scores could be attributed to individual personality.

Conclusion

The purpose of our study was to investigate the English development of four students who took part in the formal study abroad program at SOU. Since it is a small-scale case study with only four subjects, one must be cautious about drawing decisive conclusions. Bearing this in mind, we feel that the results of the pre- and post-tests provided some insights into the linguistic developmental processes which learners may have been undergoing during the study abroad program.

Linguistic improvement derived from extensive exposure to the target language did show up in scores on the tests. First, the speaking skill in the interview test dramatically improved for all the categories, particularly in fluency, indicating that the students developed greatly expanded communication skills during their one-year-stay in the U. S. Even in the narrative picture descriptions, which are the most difficult, the speaking skill was improved in the accent and grammar categories.

Second, the students showed the most dramatic improvement in the writing skill as a result of the study abroad program. In particular, students showed the greatest improvement in complexity between the pre-and post-tests. They wrote more fluently and accurately, employing more complex syntactic structures. The evidence of significant improvement in speaking and writing skills leads to the conclusion that the study abroad program contributed most effectively to the

development of productive skills.

Third, the listening skill was also improved just as the students responded most favorably on the questionnaire, but the increase was not statistically significant. The vocabulary level also appeared to have developed, especially in the 3000-level. The least improvement was found in the reading skill.

Last but not least is individual differences. The most proficient student increased least for almost all the tests, suggesting ceiling effect. Her responses on the questionnaire also indicated the lowest self-evaluation both in language skills and attitude changes. In contrast, the least proficient student developed the most significantly in writing and speaking. The fact that she responded to attitude changes with the highest scores on the questionnaire may partly explain the great development of her language skills.

There are some limitations in this study. In addition to the small number of subjects, we haven't dealt with the quantity and quality of English use during one-year-stay in SOU. To get clearer idea of the effects of study abroad, we will need to examine the content of the program provided at SOU and samples of the students' works while they were there. It will also be necessary to investigate the amount of English they were able to use outside of the classroom. In this study our main focus was on the comparison between the proficiency levels before and after the study abroad program. It would also be of interest if we could survey the development of English on a regular basis during the program. Further study will be necessary to bring about clearer understanding of the process for language learning and development through the study abroad program.

Acknowledgments

We are very grateful for the help of our colleagues, Brent Gaston, Simon Cole, and Donna Stripling who evaluated all the oral and written productions by our students. We would also like to thank Donna Stripling for her careful proofreading of the original manuscript and valuable comments.

References

- Ando, S. (1991). *Gendaikyouiku gendai key word jiten*. Tokyo: Zosindo.
- Asano, K. & Saito, N. (1996). Internet to eigo kyoiku. *Eigo kyoiku bessatsu*. Tokyo: Taishukan.
- 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 study abroad context* (pp.37-66). Amsterdam: John Benjamin.
- Briere, K. (1966). Quantity before quality in second language composition. *Language Learning*, 16, 141-151.
- Freed, B. (1995). What makes us think that students who study abroad become fluent? In B. Freed (Ed.) *Second language acquisition in study abroad context* (pp.123-148). Amsterdam: John Benjamin.
- Foster, P., & Skehan, P. (1996). The influence of planning and task type on second language performance. *Studies in Second Language Acquisition*, 18, 299-323.
- Gonzalez-Bureno, M. (1998). The effects of electronic mail on Spanish L2 discourse. *Language Learning & Technology*, 1, 50-65.

- Harada, K. (1998). Communicability no sokutei kekka. A paper read at the 37th JACET National Conference, Okayama.
- Hirano, K. (1991). The effect of mode of discourse on objective measures of EFL proficiency in Japanese university students. *JACET Bulletin*, 22, 15-34.
- Hirano, K. (1993). The effect of time limits on objective measures of EFL proficiency in Japanese university students. *JACET Bulletin*, 24, 21-39.
- Hirota, R. (1999). Chat and E-mail as discourse media. A paper read at the 31th Language Laboratory Association Kyushu Chapter Conference, Fukuoka.
- Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, 4, 51-69.
- Isoda, T. (In press). A study of personality change and English proficiency. *Annual Review of English Language and Teaching*, 3.
- Iwakiri, M. (1993). Effects of a study abroad program on the English development of Japanese college students. *JACET Bulletin*, 24, 41-60.
- Kawauchi, C. (1998). Cued/uncued oral narrative tasks in foreign language fluency. *JACET Bulletin*, 29, 83-96.
- Kawauchi, C. (1999). Developing speaking and writing by narrative tasks. A paper read at the 12th Word Congress of Applied Linguistics (AILA), Tokyo.
- Kikuchi, M. (1988). The summer language program and foreign language proficiency. *Shoin Literary Review*, 24, 51-73.
- Kishigami, H. (1997). *Kyo kara hajimeru TOEFL*. Tokyo: Senseishuppan.
- Kobayashi, H., & Rinnert, C. (1992). Effects of first language on second language writing: Translation versus direct composition. *Language Learning*, 42, 183-215.

- Lafford, B. (1995). Getting into, through and out of a survival situation: A comparison of communicative strategies used by students studying Spanish abroad and at home. In B. Freed (Ed.) *Second language acquisition in study abroad context* (pp.97-122). Amsterdam: John Benjamin.
- Larsen-Freeman, D., & Strom, R. (1977). The construction of a second language acquisition index of development. *Language Learning*, 27, 123-134.
- Laufer, B., & Nation, P. (1999). A vocabulary size test of controlled productive ability. *Language Testing*, 15, 33-51.
- Lazaraton, A. (1996). Interlocutor support in oral proficiency interview: The case of CASE. *Language Testing*, 13, 151-172.
- Lennon, P. (1989). Introspection and intentionality in advanced second language acquisition. *Language Learning*, 39, 375-396.
- Lennon, P. (1990). Investigating fluency in EFL: A quantitative approach. *Language Learning*, 40, 387-417.
- Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL Quarterly*, 20, 83-93.
- Tanaka, M. (1997). *TOEFL reading mondai 270*. Tokyo: Obunsha Editor.
- Towel, I., Hawkins, R., & Bazergui, N. (1996). The development of fluency in advanced learners of French. *Applied Linguistics*, 12, 164-178.
- Valette, R. M. (1977). *Modern language testing*. San Diego: Harcourt Brace Jovanovich.
- Watanabe, M. (1990). Gaikokugo kyoiku kaizen eno teigen. *Sony LL Tsushin*, 151, 6-8.
- Yamamoto, H. (1992). Effects of a study abroad program on the English development of Japanese college students. *Seinan Jogakuin*

- Tanki Daigaku Kenkyu Kiyo*, 39, 21-30.
- Yamane, S. (1985). Studying in Canada: Effects on listening comprehension. *Tetsukayama Tanki Daigaku Kiyo*, 22, 113-124.
- Yashima, T. (1997). Extroversion vs. introversion and foreign language learning. *Language Laboratory*, 31, 93-105.
- Yashima, T., & Viswat, L. (1997). Acquisition of "fluency in L2" through an overseas study program. *JACET Bulletin*, 28, 193-206.
- Yashima, T., Yamamoto, S., & Viswat, L. (1994). Acquisition of communicative competence through overseas study in America: A case study of eight English high school students. *Language Laboratory*, 32, 31-43.