L2 Pragmatic Development Through Study Abroad:
Change of Speed and Accuracy in Appropriateness Judgment

Seiji FUKAZAWA
Shusaku KIDA
Hiroshima University
Kyoko SHINOMURA
Akiko TATSUMI
Yuka YAMAUCHI
Graduate School of Education, Hiroshima University

Abstract

The present study investigates how study abroad in the United Kingdom affects the
development of second language (L2) pragmatic competence. Specifically, this study examined
the development of L2 learners’ speed and accuracy in judging the appropriateness of L2 requests.
The participants comprised 22 Japanese university-level English as a foreign language learners
who experienced a four-month study abroad in a homestay environment. They performed the
same pragmatic appropriateness judgment task before (as a pretest in February 2014) and at the
end (as a posttest in July 2014) of study abroad. The task consisted of six appropriate and 10
inappropriate scenarios. The participants were asked to judge whether or not an L2 request
sentence was appropriate in a given context as rapidly and accurately as possible and their reaction
times and accuracy of judgment were measured. Overall, the results showed that appropriateness
judgment became faster and more accurate over time. These results suggest positive effects of
study abroad on the development of L2 pragmatic competence.

1. Introduction

Previous studies in interlanguage pragmatics have investigated various factors that may be
responsible for second language (L2) pragmatic competence and its development. These factors
include learners’ first language (L1), explicit and implicit instruction, and learning environment, to
name a few examples (Fordyce, 2014). Of these factors, learning environment factors are amongst
the most featured in recent research (Taguchi, 2011). Learning environment in L2 pragmatic
research is further divided into sub-topics, such as the difference between English as a second or
foreign language (ESL or EFL) (Bardovi-Harlig & Dörnyei, 1998; Schauer, 2006), length of
residence (Xu, Case, & Wang, 2009), and study abroad (Barron, 2003; Schauer, 2006, 2009;
Shively, 2011; Taguchi, 2008, 2011). Of these, the present study focuses on the effects of a four-month (one semester long) study abroad program implemented by a university faculty. This kind of study abroad program is very popular in Japanese universities. Therefore, an attempt to provide a detailed investigation of the program would be beneficial to those who implement similar programs. In particular, the present study examined pragmatic competence, specifically a fluency (processing speed) dimension of pragmatic competence for L2 learners, which has rarely been dealt with in most previous L2 pragmatic research.

2. Literature Review

2.1 Research on L2 Pragmatic Awareness

2.1.1 Bardovi-Harlig and Dörnyei’s (1998) Seminal Study

In interlanguage pragmatics, one critical topic of investigation is the relationship between L2 learners’ grammatical and pragmatic awareness. A seminal paper about this topic was the study by Bardovi-Harlig and Dörnyei (1998). They developed a series of scenarios in which various grammatically incorrect or pragmatically inappropriate utterances were included, in addition to both grammatically correct and pragmatically appropriate utterances. Participants were asked to judge if the target sentence in each scenario had any problem in each context. If yes, they were further asked to indicate the severity of the problem on a six-point Likert scale from *not bad at all* to *very bad*. The task was given to EFL and ESL learners and their answers were compared. The results demonstrated that, generally, EFL learners tended to pay more attention to grammatical errors while ESL learners detected more pragmatic inappropriateness.

The original study has been (at least partially) replicated both abroad (Niezgoda & Röver, 2002) and in Japan (Tagashira, Yamato, & Isoda, 2011). The results of these studies have demonstrated that L2 learners’ pragmatic awareness is more complex than the original study had shown, and that not only the learning environment but also other factors (e.g., individual learner factors such as motivation) could affect the development of L2 learners’ pragmatic awareness. Therefore, more detailed investigations are necessary. Further, some unsolved limitations exist in the previous research. The first limitation is that previous studies did not distinguish between types of pragmatic error. One type of pragmatic error was *under-polite* utterances (e.g., a student addresses an unfamiliar student and says, *Tell me how to get to the library*). The other type was *over-polite* utterances (e.g., a man is ordering a snack at the bar and says, *Would you be so kind as to give me a sandwich and a yogurt please?*). Although these two types of pragmatic error can be theoretically distinguished (Culpeper, 1996, 2008; Watts, 2005), the original study by Bardovi-Harlig and Dörnyei (1998) and its replications treated them simply as pragmatic errors, and no clear distinction was made. The second limitation is that previous studies have usually measured L2 learners’ pragmatic awareness through off-line measurement such as a questionnaire. In natural communication however, L2 learners are required to rapidly process the L2 input,
necessitating the examination of L2 learners’ fluency by on-line measurement in pragmatic research.

2.1.2 Distinction Between Under- and Over-Polite Pragmatic Errors

In order to treat the first limitation discussed above, Sawai (2013) explicitly distinguished between under- and over-polite pragmatic errors and developed new questionnaire scenarios based on Bardovi-Harlig and Dörnyei’s (1998) original questionnaire. Sawai asked five native speakers of English to rate the politeness level of English request sentences from -3 to +3, as shown in Figure 1.

Based on the English native speakers’ ratings, Sawai (2013) developed six appropriate scenarios and 10 inappropriate scenarios. The 10 inappropriate scenarios consisted of five under-polite and five over-polite scenarios. The degree of under- and over-politeness was carefully manipulated, ranging from extremely under- or over-polite to just slightly under- or over-polite. The results are shown in Table 1.

Table 1
Results of Each Scenario’s Politeness Rating by Native Speakers of English in Sawai (2013)

<table>
<thead>
<tr>
<th></th>
<th>Appropriate (+P)</th>
<th>Appropriate (-P)</th>
<th>Under-polite</th>
<th>Over-polite</th>
</tr>
</thead>
<tbody>
<tr>
<td>M 0.6</td>
<td>0.4</td>
<td>0.4</td>
<td>-2.6</td>
<td>-2.4, -1.4, -1.4, -1.4, -0.8</td>
</tr>
<tr>
<td>SD 0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5, 0.8, 1.4, 0.5, 0.7</td>
<td>0.5, 0.6, 0.8, 0.7, 0.6</td>
</tr>
<tr>
<td>Minimum 0</td>
<td>0</td>
<td>0</td>
<td>-3, -3, -3, -2, -2</td>
<td>2, 2, 1, 0, 0</td>
</tr>
<tr>
<td>Maximum 1</td>
<td>1</td>
<td>1</td>
<td>-2, -2, -1, -1, 0</td>
<td>3, 3, 2, 2, 2</td>
</tr>
</tbody>
</table>

Note: +P and -P refer to the ±Power of the interlocutor’s social status; a and b refer to appropriate scenarios whereas c and d refer to inappropriate scenarios.
Based on the 16 scenarios, Sawai (2013) asked the Japanese university-level EFL learners to review the appropriateness of the target request sentence in each scenario in the given context. As in Bardovi-Harlig and Dörnyei (1998), when the participants judged that the target sentence was inappropriate, they were asked to indicate the degree of inappropriateness on a 6-point Likert scale. The results demonstrated that L2 learners’ pragmatic awareness varies depending on the types of pragmatic error; that is, it is more difficult to detect the inappropriateness of over-polite requests than under-polite requests for Japanese learners of English. Further, the degree of inappropriateness (i.e., extremely inappropriate or slightly inappropriate) also played a role in learners’ performance in the task. These results suggest the theoretical importance of the distinction between under- and over-polite pragmatic errors in L2 pragmatic awareness research.

2.1.3 Use of On-Line Measurement of Pragmatic Awareness

Sawai (2013) was a seminal paper that dealt with one important limitation of previous research in L2 learners’ pragmatic awareness, but the second limitation, namely, the lack of processing speed measurement was not resolved in her research. Therefore, Fukazawa, Maeda, Kida, Yamauchi, and Tatsumi (2015) examined L2 learners’ pragmatic awareness using an on-line pragmatic appropriateness judgment task in which the participants were asked to judge whether the target L2 request was appropriate or not in the context, using the 16 scenarios developed by Sawai. In their research, the reaction time (RT) of the judgment was measured from the time the target sentence appeared on the computer monitor to the time participants made the appropriateness judgment. The accuracy of judgment was also measured. The results of their study showed that (a) L2 learners processed pragmatically appropriate requests more rapidly and accurately than pragmatically inappropriate requests, (b) extremely under-polite requests were processed rapidly and accurately while slightly under-polite ones were processed more slowly and less accurately, and finally (c) both extremely and slightly over-polite requests were processed slowly and less accurately. These results confirmed Sawai’s conclusion that types of pragmatic error influence L2 learners’ pragmatic awareness. In particular, the results revealed that L2 learners’ fluency in processing L2 pragmatic input may vary according to the types and degrees of pragmatic error.

What remained unclear from these previous studies is whether L2 learners’ pragmatic awareness develops over time. Since previous research in this field has almost exclusively used a cross-sectional design, such as a comparison between ESL and EFL or lower and higher L2 proficiency, little is known about whether the same L2 learners develop their processing speed of L2 pragmatic input over time and what factors contribute to it.

2.2 Effects of Study Abroad on the Development of Pragmatic Competence

One of the most powerful factors that contribute to L2 acquisition is study abroad. Research investigating the effects of study abroad in general focuses on various facets of second language
acquisition, such as the development of speaking fluency (e.g., Freed, Segalowitz, & Dewey, 2004; Llanes & Muñoz, 2009), listening comprehension (e.g., Cubillos, Chieffo, & Fan, 2008; Llanes & Muñoz, 2009), and reading skills (Dewey, 2004). In general, these previous studies suggest that study abroad experience results in the positive development of various dimensions of L2 competence.

In interlanguage pragmatic research, many studies have investigated the effects of study abroad on L2 learners’ pragmatic development (e.g., Bataller, 2010; Cohen & Shively, 2007; Fukazawa & Fordyce, 2005; Matsumura, 2003; Shively, 2011; Schauer, 2006; Taguchi, 2008, 2011). The results of these studies have shown that study abroad is a powerful contributor to L2 pragmatic development, since it allows L2 learners to be exposed to a great deal of L2 input and to experience authentic L2 interaction that may be difficult to obtain in an EFL classroom. It is not clear, however, how L2 learners develop the fluency dimension of pragmatic competence. This is because most research thus far has measured L2 learners’ pragmatic ability using off-line measurement, such as questionnaires or discourse completion tests, and few studies have focused on learners’ processing speed through some sort of on-line measurement (Taguchi, 2005). Therefore, more research is needed to examine the development of L2 pragmatic competence by on-line measurement.

2.3 Summary and Research Questions

In sum, previous studies in interlanguage pragmatic research have shown that L2 learners’ pragmatic awareness may vary according to type of pragmatic error (i.e., under-/over-polite requests). Therefore, it is necessary to distinguish between them in pragmatic awareness research. The other point is that most previous studies have not measured L2 learners’ fluency in processing L2 pragmatic input, suggesting the importance of using some sort of on-line measurement. In terms of the development of L2 pragmatic competence, previous studies have shown that study abroad leads to various aspects of pragmatic development.

Based on the review of the previous research, the present study examines the effect of study abroad on the acquisition of L2 pragmatic competence. Specifically, this study investigates the development of L2 learners’ pragmatic awareness through study abroad using the pragmatic appropriateness judgment task developed by Sawai (2013) and measures L2 learners’ processing speed and accuracy. The present study addresses the following research questions:

(1) Does Japanese L2 English learners’ speed of pragmatic appropriateness judgment develop over time through study abroad?

(2) Does Japanese L2 English learners’ accuracy of pragmatic appropriateness judgment develop over time through study abroad?
3. Method

3.1 Participants

The participants consisted of 22 second-year university students (13 males and 9 females), who were all native speakers of Japanese. They were majoring in English language education at their university in Japan and studied at the University of Warwick or the University of Edinburgh in the UK for approximately four months (from April to August 2014) in a homestay environment. During their stay in the UK, they took academic courses specifically designed for international students (not the so-called ESL classes), such as phonetics, literature, teaching methodologies, cultural understanding, and so forth, which did not mainly aimed at improving their English proficiency. The students were required to speak English, even with their Japanese classmates during their stay. Further, they were encouraged to actively interact with their host family members. Their background information before going to study abroad is shown in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>19.13</td>
<td>0.46</td>
<td>19</td>
<td>21</td>
</tr>
<tr>
<td>Starting age of learning English</td>
<td>12.18</td>
<td>1.30</td>
<td>9</td>
<td>13</td>
</tr>
<tr>
<td>Years of formal instruction</td>
<td>6.68</td>
<td>0.97</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>TOEIC score</td>
<td>670.95</td>
<td>77.22</td>
<td>495</td>
<td>885</td>
</tr>
<tr>
<td>Self-rating: Listening</td>
<td>4.73</td>
<td>1.60</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>Speaking</td>
<td>4.32</td>
<td>1.69</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Reading</td>
<td>5.05</td>
<td>1.80</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Writing</td>
<td>5.00</td>
<td>1.78</td>
<td>1</td>
<td>8</td>
</tr>
</tbody>
</table>

Note: Self-rating ranged from 1 (I have minimum proficiency) to 10 (I have near-native proficiency).

3.2 Pragmatic Appropriateness Judgment Task

The same task materials as were used in Fukazawa et al. (2015) were employed in this study (see Fukazawa et al. for the complete set of materials), which were adopted from the questionnaire developed by Sawai (2013). There were six appropriate and 10 inappropriate L2 requests. Among the six appropriate requests, three were requests to those with higher power status in a social context (+Power; namely, in this case, teachers), while the other three requests were to those of equal power (-Power; peers). Among the 10 inappropriate requests, five were to those of higher status (categorized as under-polite) and the other five were to those of equal power (categorized as over-polite).
3. Method

3.1 Participants

The participants consisted of 22 second-year university students (13 males and 9 females), who were all native speakers of Japanese. They were majoring in English language education at their university in Japan and studied at the University of Warwick or the University of Edinburgh in the UK for approximately four months (from April to August 2014) in a homestay environment. During their stay in the UK, they took academic courses specifically designed for international students (not the so-called ESL classes), such as phonetics, literature, teaching methodologies, cultural understanding, and so forth, which did not mainly aim at improving their English proficiency. The students were required to speak English, even with their Japanese classmates during their stay. Further, they were encouraged to actively interact with their host family members. Their background information before going to study abroad is shown in Table 2.

Table 2

<table>
<thead>
<tr>
<th></th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>19.1</td>
<td>21.0</td>
</tr>
<tr>
<td>Starting age of learning English</td>
<td>12.0</td>
<td>18.0</td>
</tr>
<tr>
<td>Years of formal instruction</td>
<td>6.0</td>
<td>9.0</td>
</tr>
<tr>
<td>TOEIC score</td>
<td>670</td>
<td>885</td>
</tr>
<tr>
<td>Self-rating:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening</td>
<td>2.73</td>
<td>7.00</td>
</tr>
<tr>
<td>Speaking</td>
<td>4.32</td>
<td>7.00</td>
</tr>
<tr>
<td>Reading</td>
<td>5.05</td>
<td>8.00</td>
</tr>
<tr>
<td>Writing</td>
<td>5.00</td>
<td>8.00</td>
</tr>
</tbody>
</table>

Note: Self-rating ranged from 1 (I have minimum proficiency) to 10 (I have near-native proficiency).

3.2 Pragmatic Appropriateness Judgment Task

The same task materials as were used in Fukazawa et al. (2015) were employed in this study (see Fukazawa et al. for the complete set of materials), which were adopted from the questionnaire developed by Sawai (2013). There were six appropriate and 10 inappropriate L2 requests. Among the six appropriate requests, three were requests to those with higher power status in a social context (+Power; namely, in this case, teachers), while the other three requests were to those of equal power (-Power; peers). Among the 10 inappropriate requests, five were to those of higher status (categorized as under-polite) and the other five were to those of equal power (categorized as over-polite).

3.3 Apparatus and Procedure

In the experiment, Epson ST12E computers with Windows 7 Professional (32-bit, Core 2 Duo CPU, 2.00 GB RAM) were used. DMDX software (Forster & Forster, 2003) was employed for the presentation of the experimental materials and the measurement of participants’ RTs and accuracy data.

The participants were asked to judge whether or not the final sentence was appropriate in the provided context as quickly and accurately as possible. They pressed the yes button (the right control on the keyboard) if they thought the sentence was appropriate or the no button (the left control) if they thought the sentence was inappropriate. The four types of response obtained through the experiment were as follows:

(1) Yes response to an appropriate request (a correct judgment called Hit)
(2) Yes response to an inappropriate request (an incorrect judgment called False Alarm)
(3) No response to an appropriate request (an incorrect judgment called Miss)
(4) No response to an inappropriate request (a correct judgment called Correct Rejection)

Hits and correct rejections are correct responses in the task. Because appropriate responses were included as fillers, among the two types of correct response, the RTs and accuracy of correct rejections were mainly reported in this study.

Before the main session, the participants took a practice session in which two appropriate and two inappropriate L2 requests were displayed. First, the participants read a description of the situation in Japanese on the computer screen. They then proceeded to the next screen by pressing the space bar, where the dialogue appeared without the target sentence. After reading the dialogue, they once more pressed the space bar, and the underlined target sentence was displayed until the participant made a judgment. The RT was measured from the time the target sentence appeared on the screen to the time the participant made the yes/no judgment. The whole data collection process

![Figure 2](image-url)  
*Figure 2. Procedure of the appropriateness judgment task used in the pre- and posttests.*
lasted approximately 15 minutes without any intermission. The procedure is visualized in Figure 2. The participants took the same appropriateness judgment task before (as a pretest in February 2014) and at the end (as a posttest in July 2014) of study abroad.

4. Results and Discussion

4.1 Results of Reaction Time in Appropriateness Judgment

The results of RTs in the appropriateness judgment task in the pretest and posttest are shown in Table 3. Further, Figure 3 depicts the change of mean RTs for correct responses while Figure 4 depicts those for incorrect responses. Overall, the RTs of correct and incorrect responses in all scenarios in the posttest were faster than those in the pretest.

Examining the RTs of correct responses in each scenario specifically, in most cases participants’ RTs were faster in the posttest than in the pretest. However, an exception was found in Scenario c5. This item was processed more slowly in the posttest compared to the pretest (Table 4).

![Figure 3](image1.png)

*Figure 3.* Results of mean correct reaction times for each scenario in the pre- and posttests.

![Figure 4](image2.png)

*Figure 4.* Results of mean incorrect reaction times for each scenario in the pre- and posttests. Single plots (a3, b2, and b3) indicate that either the pre- or posttest has an omitted data because all the responses were correct.
Figure 4. Results of mean correct reaction times for each scenario in the pre- and posttests.

Figure 3. Results of mean incorrect reaction times for each scenario in the pre- and posttests.

4.1 Results of Reaction Time in Appropriateness Judgment

The participants took the same appropriateness judgment task before (as a pretest in February 2014) and at the end (as a posttest in July 2014) of study abroad. The procedure is visualized in Figure 2. In most cases, the RTs of correct responses in each scenario specifically, in most cases participants’ RTs were faster in the posttest than in the pretest. However, an exception was found in Scenario c5. This item was processed more slowly in the posttest compared to the pretest (Table 3). Further, Figure 3 depicts the change of mean RTs for scenarios in the posttest were faster than those in the pretest.

Table 3

<table>
<thead>
<tr>
<th></th>
<th>Correct responses</th>
<th>Incorrect responses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>Post</td>
</tr>
<tr>
<td><strong>Appropriate (+P)</strong></td>
<td>6,450 &amp; 3,575</td>
<td>4,875 &amp; 3,273</td>
</tr>
<tr>
<td>a1</td>
<td>7,823 &amp; 3,487</td>
<td>5,002 &amp; 1,960</td>
</tr>
<tr>
<td>a2</td>
<td>8,259 &amp; 3,517</td>
<td>5,746 &amp; 4,597</td>
</tr>
<tr>
<td>a3</td>
<td>4,291 &amp; 1,429</td>
<td>3,935 &amp; 2,252</td>
</tr>
<tr>
<td><strong>Appropriate (-P)</strong></td>
<td>5,096 &amp; 3,093</td>
<td>3,767 &amp; 2,315</td>
</tr>
<tr>
<td>b1</td>
<td>6,871 &amp; 3,295</td>
<td>5,177 &amp; 1,850</td>
</tr>
<tr>
<td>b2</td>
<td>4,635 &amp; 3,535</td>
<td>2,602 &amp; 913</td>
</tr>
<tr>
<td>b3</td>
<td>3,955 &amp; 1,575</td>
<td>3,928 &amp; 2,861</td>
</tr>
<tr>
<td><strong>Under-polite</strong></td>
<td>5,361 &amp; 2,948</td>
<td>3,324 &amp; 2,204</td>
</tr>
<tr>
<td>c1</td>
<td>4,019 &amp; 1,861</td>
<td>3,099 &amp; 1,939</td>
</tr>
<tr>
<td>c2</td>
<td>4,535 &amp; 1,094</td>
<td>2,242 &amp; 1,056</td>
</tr>
<tr>
<td>c3</td>
<td>6,342 &amp; 4,786</td>
<td>3,948 &amp; 2,416</td>
</tr>
<tr>
<td>c4</td>
<td>6,596 &amp; 2,942</td>
<td>3,366 &amp; 1,556</td>
</tr>
<tr>
<td>c5</td>
<td>6,288 &amp; 2,037</td>
<td>9,130 &amp; 3,301</td>
</tr>
<tr>
<td><strong>Over-polite</strong></td>
<td>8,851 &amp; 4,169</td>
<td>5,013 &amp; 2,559</td>
</tr>
<tr>
<td>d1</td>
<td>9,975 &amp; 3,763</td>
<td>5,176 &amp; 1,968</td>
</tr>
<tr>
<td>d2</td>
<td>12,298 &amp; 5,219</td>
<td>7,040 &amp; 2,904</td>
</tr>
<tr>
<td>d3</td>
<td>5,962 &amp; 2,910</td>
<td>3,561 &amp; 2,274</td>
</tr>
<tr>
<td>d4</td>
<td>7,284 &amp; 2,462</td>
<td>5,554 &amp; 1,410</td>
</tr>
<tr>
<td>d5</td>
<td>10,837 &amp; 2,875</td>
<td>4,278 &amp; 2,650</td>
</tr>
</tbody>
</table>

Note. Bold type refers to the average time of each category. n.a. in incorrect responses indicates that there was no data because all the responses were correct. The standard deviation for the incorrect a2 result is zero, because only one participant’s data has been used.

3 and Figure 3). The RTs for Scenario c5 were 6,288 ms in the pretest and 9,130 ms in the posttest. The delay in RTs in c5 was remarkable. On the other hand, the RT of incorrect responses for Scenario c5 was reduced in the posttest as well as other items (Table 3 and Figure 4).

4.2. Results of Accuracy in Appropriateness Judgment

Table 4 shows the results of appropriateness judgment accuracy for four situations (appropriate +P, appropriate -P, under-polite, and over-polite) in the pre- and posttests. Their
Figure 5. Change of accuracy rates between the pre- and posttest for each scenario.

Table 4
Results of Accuracy in the Pragmatic Appropriateness Judgment at the Pre- and Posttest Phases

<table>
<thead>
<tr>
<th></th>
<th>Pre</th>
<th>Post</th>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td>Appropriate (+P)</td>
<td>71%</td>
<td>94%</td>
<td>Appropriate (-P)</td>
<td>82%</td>
</tr>
<tr>
<td>a1</td>
<td>59%</td>
<td>86%</td>
<td>b1</td>
<td>73%</td>
</tr>
<tr>
<td>a2</td>
<td>64%</td>
<td>95%</td>
<td>b2</td>
<td>100%</td>
</tr>
<tr>
<td>a3</td>
<td>91%</td>
<td>100%</td>
<td>b3</td>
<td>73%</td>
</tr>
<tr>
<td>Under-polite</td>
<td>45%</td>
<td>65%</td>
<td>Over-polite</td>
<td>57%</td>
</tr>
<tr>
<td>c1</td>
<td>59%</td>
<td>82%</td>
<td>d1</td>
<td>50%</td>
</tr>
<tr>
<td>c2</td>
<td>50%</td>
<td>82%</td>
<td>d2</td>
<td>50%</td>
</tr>
<tr>
<td>c3</td>
<td>50%</td>
<td>82%</td>
<td>d3</td>
<td>86%</td>
</tr>
<tr>
<td>c4</td>
<td>41%</td>
<td>73%</td>
<td>d4</td>
<td>50%</td>
</tr>
<tr>
<td>c5</td>
<td>23%</td>
<td>9%</td>
<td>d5</td>
<td>50%</td>
</tr>
</tbody>
</table>

accuracy rates in the posttest tended to be higher than those in the pretest. However, there is a distinctive feature for Scenario b1, b2 (both are appropriate), and c5 (under-polite) unlike tendencies in other scenarios (Table 4 and Figure 5). As is the case with RT data, the participants’ judgments for Scenario c5 disimproved in the accuracy dimension after study abroad.

4.3 Discussion

In this section, we will discuss the results of the effects of study abroad on the development of participants’ speed and accuracy in the appropriateness judgment task in order to answer the research questions (1) and (2). Further, we will specifically discuss the results of Scenario c5.

The first research question addressed whether L2 learners’ speed is increased by the four-month study abroad. The answer was positive. The results of the pre- and posttests showed that the participants’ RTs for the appropriateness judgment increased in most cases. This suggests that L2 learners’ fluency at processing L2 pragmatic input increased as a result of their four-month
sojourn. The second research question addressed whether L2 learners’ accuracy develops over time. Again, the answer was positive since their accuracy rates in appropriateness judgment improved in most scenarios.

Although the overall results suggested the positive effects of studying abroad on the development of L2 pragmatic competence, a closer examination of the data revealed that there were some exceptions. The most noticeable item was c5. This is a scenario where a student wants to fill in a questionnaire at the teacher’s office and says, *Hello. My name is Kate Arista. If you don’t mind, I would like you to fill this in for me.* The sentence was almost acceptable, but the final expression *I would like you to ...* is slightly rude as a request to the +P interlocutor because it implies that the teacher is not allowed to refuse the student’s request. For this item, it is difficult to have any discussion with confidence about the RTs of the correct rejection response because the accuracy rates were too low. The accuracy was 23% (five participants among the 22 answered correctly) in the pretest and 9% (two participants answered correctly) in the posttest, and no participants correctly judged this item in both tests. Therefore, the difference in mean RTs for correct response between the pre- and posttests is not informative since it may simply demonstrate individual difference, not the change in within-participants’ fluency over time. Instead, mean RTs for incorrect responses are even more intriguing. The results demonstrated that the participants came to make an incorrect judgment more rapidly (7,840 ms in the pretest, 5,472 ms in the posttest). This seems to demonstrate that the participants made the incorrect judgment with more confidence in the posttest, which suggests that study abroad caused some pragmatic backsliding. Since the target request sentence in c5 was an inappropriate request (under-polite), an incorrect response to this item means that the participants judged the request to be acceptable or polite-enough in the given context.

How can this backsliding be explained? One possible explanation is that the participants changed their sociolinguistic mental range of acceptable appropriateness as a result of study abroad. That is, before they studied abroad, more participants thought that the c5 request was unacceptable, correctly judging that request to be under-polite (Pre in Figure 5). However, after experiencing a great deal of L2 authentic input, they grew accustomed to frank request sentences in the UK through daily interactions with peers and people with whom they had close relationships, such as their host family members. This, in turn, wrongly broadened the mental range of acceptable appropriateness and more participants came to accept the c5 request (Post in Figure 5). This change in the mental range of acceptable appropriateness resulted in the drop in accuracy rates of the c5 item from the pretest to the posttest, as shown in Figure 5. There are a number of studies that support this interpretation (e.g., Marriott, 1995; Regan, 1995). For example, Marriott (1995) observed that L2 Japanese learners preferred a more polite style before studying abroad, but after returning to their home country, they predominantly used a more plain style, arguing that their change of speech style was due to the fact that L2 learners staying in a home-stay environment had had greater exposure to plain style Japanese since this was employed
by host family members and close friends. The other sociolinguistic example by Regan (1995) demonstrated that L2 learners who experienced a one-year sojourn overgeneralized the non-prestige form of negation in French. These examples suggest that L2 learners develop their pragmatic awareness by changing the sociolinguistic mental range of acceptable appropriateness for L2 utterances. That is, at some point in their development, learners do not accept a slightly under-polite utterance because it is located outside of their acceptable range and therefore they judge the sentence to be under-polite. Later, the same learners may accept the same utterance in the same situation because they have modified their mentally acceptable range for appropriateness judgment as a result of a great deal of L2 exposure. In the case of c5 in the present study, the adjustment of mental range was unsuccessful and the participants incorrectly adapted their mental range as a result of study abroad (Figure 6). This interpretation suggests that L2 learners’ pragmatic development is not necessarily linear, but it could track complex developmental paths.

![Figure 6](image)

*Figure 6.* Image of the mentally acceptable range. White indicates acceptance as appropriate, while black indicates rejection as inappropriate.

The developmental shift can be visualized as in Figure 6. In this figure, the range of linguistic politeness of L2 utterances is continuously expressed from extremely under-polite (the most left) to extremely over-polite (the most right). The appropriate range is located between the two inappropriate ranges. Further, the white areas indicate learners’ acceptance of a statement as appropriate, while black areas indicate rejection as inappropriate. Finally, the gray indicates areas where learners have difficulty judging the utterance as either appropriate or inappropriate. Theoretically, the middle appropriate area should be colored white and the inappropriate areas on either side should be colored black. Before L2 learners experience study abroad, many areas are gray, demonstrating their difficulty in judging the appropriateness of a given L2 sentence. Only a limited appropriate range can be easily judged as appropriate. This tendency changes after study abroad; their judgments seem more confident with clearer contrasts between white and black. After they experience study abroad, however, both white and black areas become large, demonstrating that participants can judge more appropriate sentences as appropriate or
inappropriate sentences as inappropriate, reducing the areas colored in gray. However, their developmental level is inadequate since they mistakenly broaden the white area to include the slightly under-polite area, crossing the dotted border between the under-polite and appropriate ranges. Therefore, the development of L2 pragmatic awareness can be understood as a process in which learners reduce the gray area and fix the border between the appropriate and two inappropriate areas.

One thing to note here is that the hypothesized change of mental range of acceptable appropriateness was not observed in the case of over-polite requests. This may be because, generally, the frequency of over-polite requests is much lower than that of under-polite or frank requests for the participants in this study. Since the participants were university students, it is more likely that they were exposed to frank requests through daily interaction between peers, host family members, and teachers. It is not likely that they were exposed to over-polite requests to a similar degree as under-polite ones. Therefore, it seems more reasonable to suppose that the change of mental range of acceptable appropriateness is more likely to occur for under-polite requests, particularly in the case of just slightly under-polite utterances such as that in c5 in this study. The results of over-polite scenarios seem to support this hypothesis because speed and accuracy increased in almost all items and no item showed clear backsliding as shown in the c5 under-polite scenario. The hypothesis shown here is prominent, but it was not completely validated by the present study and therefore it should be tested more directly in future research.

In sum, the present study revealed that, during the four-month study abroad period, the participants were exposed to a great deal of L2 pragmatic input and experienced authentic interaction in L2. As a result, their speed and accuracy on appropriateness judgments were improved in most respects. Adding to the results of prior studies that demonstrated the effects of study abroad on learners’ pragmatic development in general, using off-line measurements such as a discourse completion test and a multiple choice test (Kondo, 1997a, 1997b; Code & Anderson, 2001; Fukazawa & Fordyce, 2005; Matsumura, 2001, 2003), this study confirms that the effects of study abroad on pragmatic development in L2 are robust.

There are some limitations to this study. The first is that the number of participants was small. Although there is a general agreement among researchers that it is difficult to have a large number of participants in study abroad research, more participants would be required in order to generalize the results of this study. The second limitation is that the present study cannot precisely specify what factors contributed to the development. The majority of the discussion assumed that a great deal of exposure to L2 resulted in the development, but there may have been some latent factors, such as explicit or implicit instruction, that were available during study abroad. Finally, we did not grasp in detail the amount and quality of participants’ language exposure during the period, particularly outside of class. The study abroad program in this study had general rules that the participants were required to use only English even among their Japanese peers and to actively interact with their host family members. However, it is difficult to completely examine the
participants’ language exposure during study abroad in this kind of longitudinal research. Future research is required to collect more qualitative data from participants’ own reports, such as interviews or participants’ daily journals. These types of data would help deepen our understanding of the effect of study abroad on the development of L2 pragmatic competence.

5. Conclusion

The present study investigated the effect of four-month study abroad in the UK on the development of L2 pragmatic competence. This study particularly focused on the change of speed and accuracy in the pragmatic appropriateness judgment task over time. The results showed that the participants improved both the speed and accuracy of their appropriateness judgment in most cases. In some situations, however, some pragmatic backsliding occurred, depending on the type or degree of pragmatic errors tested. In particular, the present study hypothesized that the cause of the backsliding is the L2 learners’ change of sociolinguistic mental range of acceptable appropriateness and it is more likely to occur in the case of slightly under-polite requests. More research is desirable to further investigate the process of L2 pragmatic development through study abroad.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 25370690, Grant-in-Aid for Scientific Research (C).

References


L2 listening comprehension skills. Foreign Language Annals, 41, 157-185. 

doi:10.1016/0378-2166(95)00014-3

A. Locher (Eds.), Impoliteness in language: Studies on its interplay with power in theory and 
practice (pp. 17–44). Berlin, Germany: Mouton De Gruyter.

domestic immersion and study abroad contexts. Studies in Second Language Acquisition, 26, 
303–327. doi:10.1017/S0272263104062072


Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond 
doi:10.3758/BF03195503

fluency in French: Comparing regular classroom, study abroad, and intensive domestic 
immersion programs. Studies in Second Language Acquisition, 26, 275–301. doi:10.1017/S02 
72263104062060

development of university-level Japanese EFL learners. Annual Review of English Language 
Education in Japan, 16, 31–40.

Fukazawa, S., Maeda, H., Kida, S., Yamauchi, Y., & Tatsumi, A. (2015). Speed and accuracy of 
appropriateness judgments for L2 requests by Japanese EFL learners. Annual Review of 
English Language Education in Japan, 26, 125–140.

Longitudinal study on interlanguage apologies. Sophia Linguistica, 41, 265–284.

Kondo, S. (1997b). Longitudinal study on the development of pragmatic competence in a natural 

353–365. doi:10.1016/j.system.2009.03.001

Freed (Ed.), Second language acquisition in a study abroad context (pp. 197–224). 
Amsterdam: Benjamin.


