How Does the Number of Incidental Recasts Affect Their Effectiveness on Addressees and Auditors?

Nobuhiro KAMIYA
(Graduate School, Michigan State University)

Abstract

Incidental recasts are a type of corrective feedback referring to recasts provided in response to errors that occur incidentally (e.g., Loewen, 2005). There are two variables that affect incidental recasts that have largely been ignored in previous studies: a) the number of recasts provided, and b) whether the learner is an addressee or an auditor of the recast. For the first variable, it is natural to assume that the greater the number of recasts provided for a target feature, the more effective the feedback. As for the second variable, we would anticipate that recasts are more effective for the addressee than for an auditor. The present study was conducted to examine these hypotheses. Twenty-six ESL learners participated in a treatment – immediate posttest – delayed posttest design of study.

Four grammatical features were chosen as target structures among all of the recasts provided during the implementation of a series of text-manipulation tasks that lasted two hours. These features were selected as target structures based on the total number of recasts provided on them. Two of the structures (major target structures) received a total of 39 recasts combined, whereas the other two (minor target structures) combined, received 11 recasts.

The results show that the effectiveness of incidental recasts for minor target structures was negligible regardless of whether the students were addressees or auditors. In contrast, the incidental recasts for major target structures were effective only for the addressees, but not for the auditors. This effectiveness, however, disappeared at the delayed posttest. This study suggests the necessity to take into account the number of recasts provided as well as to analyze data according to addresses and auditors when investigating the effectiveness of incidental recasts.

1. Literature review

1.1 Incidental Recasts

Focus on form (FonF; Long, 1991, 1996) is a topic that has caught the attention of many SLA researchers. FonF can be categorized into two types: proactive (or preemptive) FonF and reactive FonF (Long & Robinson, 1998). In the former case, either a teacher or a learner focuses on a linguistic element preemptively although no error has been made, whereas in the latter case, either a teacher or a learner gives corrective feedback (CF) in response to an error that another learner made, in the context of a second language (L2) classroom.

Ellis (2001) divided CF into two types: planned and incidental. Planned CF is
given when the teacher is focused on a pre-selected linguistic target feature, and therefore, tends to mainly give feedback on that particular target feature. In contrast, incidental CF is given when the teacher does not have any particular linguistic target feature in mind. In this case, the teacher gives CF on any kind of error that the learners make incidentally in the classroom. These two types of CF can coexist in L2 classrooms (Loewen & Chen, 2008), and both have been investigated in previous studies. Planned FonF has especially accumulated a large number of studies, and has yielded several meta-analyses that synthesize them (Lyster & Saito, 2010; Mackey & Goo, 2007; Norris & Ortega, 2000; Russell & Spada, 2006; Spada & Tomita, 2010). In contrast, the effectiveness of incidental CF has been investigated to a lesser extent than planned CF. This may be because (a) a pretest is not feasible (however, see Nassaji, 2009, for a solution to this problem) and (b) a posttest needs to be tailor-made for each participant, because it is impossible to predict ahead of time what linguistic features will be focused on.

One type of CF is recasts (Long, 1996), which is the reformulation of a learner’s utterance while maintaining its original meaning. According to Sheen (2004), recasts is the most common type of CF across various educational contexts. Whereas numerous studies have been conducted on recasts thus far, discussion about their effectiveness is still inconclusive. Some studies have argued for the effectiveness of recasts using test scores (Alcon, 2007; Loewen, 2005; Loewen & Philp, 2006; Nabei & Swain, 2002; Williams, 2001), the use of target features before and after treatment (Loewen, 2007; Williams, 2001), and learners’ error identification and correction of their own writing (Nassaji, 2009). Overall, these studies have shown that incidental CF is effective to some extent, approximately 50%. Meanwhile, other studies have shown that recasts are largely ineffective in terms of uptake, which is learners’ immediate response to CF (Lyster, 1998a; Sheen, 2006) and learner’s interpretation of CF (Carpenter, Jeon, MacGregor, & Mackey, 2006; Mackey, Gass, & McDonough, 2000). The conflicting results from previous studies suggest that the effectiveness of recasts remains a research topic that merits more empirical inquiries.

1.2 The Number of Recasts

A possible reason for such discrepancy in previous research results is that the number of incidental recasts provided differs from study to study. What further complicates the picture is that when incidental recasts are provided for two features in a single study, these two features may not have received the same number of incidental recasts. In other words, it is possible that one feature received more incidental recasts than the other in the study. Another problem lies in the way that the number of recasts have been reported in the literature. Thus far, the number of recasts has been reported in the form of the length of treatment. However, with this kind of data, it is difficult to determine the exact number of incidental recasts. Some of the more recent studies have reported the number of incidental recasts provided (e.g., Loewen & Philp, 2006; Lyster, 1998b). However, even these studies categorized the target features into rough groups, such as grammatical errors, lexical errors, and phonological errors. This type of categorization is not fine-grained enough to account for the exact number of incidental recasts provided for each target feature. It should be remembered that, according to the original definition of FonF (Long, 1991, 1996),

---

- 18 -
recasts should be effective even if they are provided only once. However, as stated earlier, the number of recasts may play a role in their effectiveness. It would be natural to assume that the larger the number of recasts provided for a feature, the more effective they are.

Ellis, Sheen, Murakami, and Takashima (2008) compared the effectiveness of two types of recasts in writing using three groups: a focused recasts (i.e., planned recasts) group, a unfocused recasts (i.e., incidental recasts) group, and a control group. The results showed that both types of recasts were equally effective compared to the control group. In addition, it was found that focused recasts tended to be more effective than unfocused recasts. They explained that this may be due to the sheer number of recasts provided. That is, their results may have stemmed from the greater number of recasts that were provided in response to errors in the target structure (English article) in the focused group than in the unfocused group. However, contrary to this explanation, Havranek (2002) found that the number of CF provided did not predict test scores; more interestingly, when CF was provided on a particular target feature more than three times, the test score decreased drastically. We must note, however, that in this study, the target features were grouped together according to the number of CF provided regardless of the type of CF and the type of errors.

To my knowledge, there has been no study thus far that has directly investigated the relationship between the number of incidental recasts provided for a certain target structure and their effectiveness using test scores. The present study was conducted to fill this gap.

1.3 Addresses and Auditors

Most of the previous studies on CF measured the effectiveness of CF only for its addressees; however, in a real language classroom, there are also auditors, who are ratified and known to be there by other members present in the classroom, but not directly addressed by the provider of CF (Bell, 1984).

Only a few studies have dealt with this issue to date. On the one hand, some researchers claim that learners do not benefit much from the interactions of other learners (Mackey, 1999) or CF provided by other learners especially when their L2 proficiency is low (Williams, 2001). Truscott (1999) claimed that CF for a particular learner may not be appropriate for other learners because each learner is at a different developmental stage. On the other hand, some researchers claim the opposite. Ohta (2000) found that auditors responded to the CF provided to other learners in their private speech, even when the addressee did not show any uptake. Havranek (2002: Havranek & Cesnik, 2001) found that auditors generally had higher test scores than addressees although the difference varied depending upon other variables as well, such as the target feature. Finally, using stimulated recall, Kim and Han (2007) found that both the addressees and the auditors were able to perceive recasts equally accurately. Nevertheless, due to the small number of studies that have directly addressed this issue, the results are still inconclusive.

The following two research questions were examined in the present study: 1) How does the number of incidental recasts affect their effectiveness, and 2) How does the effectiveness differ on the addressees and the auditors of the recasts?
2. Method

2.1 Participants

The participants were recruited from three ESL classes in the same low-intermediate level (Level 2) at an ESL program. The program consisted of four levels in total, divided on the basis of an in-house placement test at a large Midwestern university. The learners who agreed to participate in the present study from each class formed an extra Class outside their regular classes, yielding three Classes, in which two of them were assigned as experimental groups (henceforth, Classes E1 and E2: \( n = 6, 4 \), respectively, and Class E, combined) and the remaining one was assigned as a control group (henceforth, Class C, \( n = 16 \)). It was impossible to randomize the learners because the schedule of each of their regular classes was different. Only the data of the participants who came to all the sessions were included. Among the 26 valid participants, 16 were males and 10 were females. Their ages ranged from 17 to 40 (\( M = 21.8, SD = 4.9 \)), and their L1 was Chinese (16), Arabic (4), Korean (3), Japanese (1), Kurdish (1), and French and Bambara (1). The length of residence in an English-speaking country for these participants ranged from 0 to 24 months (\( M = 3.2, SD = 4.5 \)).

2.2 Procedure

2.2.1 Target Structure

The present study was conducted as a part of a larger study. Therefore, in addition to the target structures for incidental recasts, there was a pre-selected target structure, namely the English unreal conditional (e.g., \textit{If I were a bird, I would fly to you}). This structure is also known as the hypothetical.

Four structures were chosen for the target structures for incidental recasts: \textit{preposition}, \textit{third person singular}, \textit{modal verb}, and \textit{embedded question}. These structures were chosen among all the structures for which recasts were commonly provided in Classes E1 and E2, based on the number of incidental recasts provided.

2.2.2 Treatment

There were either two or three treatment sessions, lasting two hours in total. The researcher was the teacher in all cases. The treatment was implemented in a teacher-fronted whole class style so that the teacher was able to give feedback to each participant. The participants participated in four tasks in which they manipulated sentences of English unreal conditionals presented by the teacher, and then presented them orally. Their sentences were followed by some interactions between the participant and the teacher, who asked further questions related to the content of the sentence.

In Class E, the participants received both planned and incidental recasts. That is to say, they received recasts on errors in English unreal conditionals as well as any other linguistic feature. In contrast, the participants in Class C received recasts only on errors in English unreal conditionals.

2.3 Data Collection

The present study had a design of a treatment · immediate posttest · delayed posttest. Each test consisted of four tasks, each of which was conducted in the following order: An Elicited Oral Imitation Task, a Timed Grammaticality
Judgment Task, an Untimed Grammaticality Judgment Task, and a Metalinguistic Knowledge Task.

For logistical reasons, the interval between the end of the treatment and the immediate posttest (henceforth, Posttest 1) varied from two to four days. The delayed posttest (henceforth, Posttest 2) was conducted a week after the Posttest 1. The procedure and the test items were the same for the two tests.

All the participants from the same Class took the test together in a computer lab. Before each task, the researcher explained the procedure of the task, and then presented some examples and practice items. It took approximately one hour for the participants to finish the four tasks. Student background information was collected in advance.

2.3.1 Tests

2.3.1.1 Elicited oral imitation task.

This task consisted of 28 belief or opinion statements. Four belief/opinion statements (two grammatical, and two ungrammatical) were assigned to each of the four target structures, yielding 16 in total. The other 12 sentences were of the English unreal conditional (therefore, distractors in the present study). The order of the test items was counterbalanced within these two test times.

As the participants heard a sentence through a headset, they were required to say either “agree/yes,” “disagree/no,” or “not sure” to the content of the sentence in order for them to maintain their focus on the content. Then they were to repeat the sentence in correct English, which was recorded on a computer. Thus, if the sentence was grammatical, they could repeat it verbatim whereas in the case of an ungrammatical sentence, they needed to correct the error in the sentence as they repeated it. After 15 seconds, they heard the word “next,” and then heard the next sentence.

2.3.1.2 Timed grammaticality judgment task.

The number and the construction of the test items for this task was the same as the Elicited Oral Imitation Task; however, the sentences were different from those used in the Elicited Oral Imitation Task. The order of the test items was randomized for each test time.

An asterisk was shown in the middle of each participant’s computer screen. After 1.0 second, it was replaced by a sentence displayed in a single line. After the participants read the sentence, they judged whether it was grammatical or ungrammatical. If it was grammatical, they clicked the right shift key, and if it was ungrammatical, they clicked the left shift key as soon as possible. Then, another asterisk was shown and the next sentence was displayed. In order to keep the participants aware of the time limit, after 15 seconds from the onset of the display of the sentence, the sentence automatically disappeared, and the program proceeded to the next sentence.

2.3.1.3 Untimed grammaticality judgment task.

The test items for this task were identical to the Timed Grammaticality Judgment Task. The order of the test items was randomized for each test time. Each sentence was written on a separate sheet of paper. As the participants read a sentence, they judged whether it was grammatical or ungrammatical. If grammatical, they checked “correct,” and proceeded to the next sentence. If
ungrammatical, they checked “incorrect,” and corrected the sentence as well. There was no time limit and they were told not to go back to the previous sentences.

2.3.1.4 Metalinguistic knowledge task.

There were five sentences, all of which were ungrammatical. There was one sentence for each target structure, yielding four test items. The one remaining sentence included the unreal conditional structure (therefore, a distractor in the present study). The order of the test items was randomized for each test time.

After the participants corrected each sentence as they did in the Untimed Grammaticality Judgment Task, they wrote down the reason they thought the sentence was incorrect. There was no time limit and they were told not to go back to the previous sentences.

2.4 Data Analysis

After the Posttest 2, the researcher contacted all the teachers of the regular classes that the participants were attending while the present study was conducted. It was found that one of the teachers in Class E1 had taught modal verbs, and another teacher in Class E2 had taught the third person singular explicitly in class sometime during the present study. Therefore, these data were excluded from the analyses.

2.4.1 Coding of Addresses and Auditors

When a recast is provided, there is usually only one addressee, and all of the remaining students are auditors. However, these roles are not permanent, in that the role of each learner changes each time a recast is provided to a different learner. Thus, although it is possible that some (less vocal) students remain auditors throughout the class, it is nearly impossible that a single student continues to be an addressee throughout the entire class. In the present study, for the sake of analysis, all students who played the role of an addressee even once were coded as addressees and all of the remaining students were coded as auditors. This, admittedly, may be somewhat controversial, but was necessary in order to conduct the statistical analysis.

2.4.2 Coding of Test Answers

Obligatory parts were decided for each target structure as shown in Table 1. The following paragraphs explain how data regarding the third person singular was coded. The coding of other structures were the same except that these structures had different obligatory parts. All of the tests were coded dichotomously: 1 for correct and 0 for incorrect.

<table>
<thead>
<tr>
<th>Target structure</th>
<th>Obligatory part</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preposition</td>
<td>Preposition</td>
<td>To</td>
</tr>
<tr>
<td>Third person singular</td>
<td>Verb</td>
<td>Has</td>
</tr>
<tr>
<td>Modal verb</td>
<td>Modal + verb</td>
<td>Can cook</td>
</tr>
<tr>
<td>Embedded questions</td>
<td>Interrogative + subject + verb</td>
<td>What they are doing</td>
</tr>
</tbody>
</table>

2.4.2.1 Elicited oral imitation task.

Participant’s responses (i.e., agree/disagree/undecided) to the content of the statement they heard was not included in data analysis. The coding scheme of
participants' answers along with examples of its use is shown in Table 2. A second rater coded 20% of the data, and the Cohen’s Kappa was .884.

Table 2 Coding of the Elicited Oral Imitation Task

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Response</th>
<th>Example</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical: Every American drinks coffee every morning.</td>
<td>Repeat verbatim</td>
<td>Every American drinks coffee every morning</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Make an error in parts other than the obligatory part</td>
<td>Every American drinks the coffee every morning</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Make an error in the obligatory part</td>
<td>Every American drink coffee every morning.</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Unable to finish.</td>
<td>Every American drinks.</td>
<td>0</td>
</tr>
<tr>
<td>Ungrammatical: Everyone in the world like Mickey Mouse.</td>
<td>Correct in the obligatory part.</td>
<td>Everyone in the world likes Mickey Mouse.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Correct in the parts other than the obligatory part.</td>
<td>Everyone on the world like Mickey Mouse.</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Repeat verbatim</td>
<td>Everyone in the world like Mickey Mouse.</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Unable to finish.</td>
<td>Everyone in the world like.</td>
<td>0</td>
</tr>
</tbody>
</table>

2.4.2.2 Timed grammaticality judgment task.

In order to decide the maximum reaction time allowed for participants to respond, twenty native speakers of English completed the same task. After excluding the outliers (more than 1.5 interquartile from either the upper or the lower quartile), their reaction times were averaged, and the maximum reaction time for each test item was calculated by adding an additional 20 percent of the average, taking into account the slow processing speed of the participants (Ellis, 2005, 2006). Any response that took longer than the maximum reaction time was coded as 0 regardless of its accuracy. If the reaction time was within the maximum reaction time, and if the response was correct, it was coded 1. If the response was incorrect, it was coded 0. Cronbach’s alpha showed acceptable reliability for the Posttest 1 ($\alpha = .777$), but low reliability for the Posttest 2 ($\alpha = .615$). The results should therefore be interpreted cautiously.

2.4.2.3 Untimed grammaticality judgment task.

The coding scheme for this task and examples from it are shown in Table 3. A second rater coded 20% of the data, and the Cohen's Kappa was .891.

Table 3 Coding of the Untimed Grammaticality Judgment Task

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Response</th>
<th>Example</th>
<th>Coding</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grammatical: She talks with her friends even in class.</td>
<td>Grammatical</td>
<td>NA</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical: Correct in the parts other than the obligatory part.</td>
<td>She talks with her friends even in classes.</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Ungrammatical: Correct in the obligatory part.</td>
<td>She talk with her friends even in class.</td>
<td>0</td>
</tr>
<tr>
<td>Grammatical</td>
<td>Ungrammatical: Correct in the parts other than the obligatory part.</td>
<td>He like the present for her.</td>
<td>He likes the present from her.</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------------------------------------------</td>
<td>---------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Ungrammatical:</td>
<td>Correct in the obligatory part.</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

### 2.4.2.4 Metalinguistic knowledge task.

Because the purpose of this task was to measure the participants' knowledge about grammatical rules, the actual correction of the sentences was not analyzed. If the participants' reason for correcting the sentence accurately explained the grammar structure in the obligatory part, even without using any grammatical terminology, it was coded 1. Otherwise, it was coded 0. A second rater coded 20% of the data, and the Cohen's Kappa was .638. This low reliability was derived from the difficulty of coding participants' answers in this task. For example, there were several cases in which students seemed to partially understand the rule, but not perfectly. Because of the dichotomous nature of the coding scheme, the raters were not able to award partial points in these cases, which might have affected the inter-rater reliability. The results should therefore be interpreted cautiously.

### 2.4.3 Statistical Procedures

It should be noted that simply comparing data from the major target structures and the minor target structures within Class E would not be a valid method of statistical analysis because each target structure had a different level of difficulty, as did the difficulty of the test items for each target structure. Therefore, results from Classes E and C were compared. Because both classes were at the same level in the overall language program, it was assumed that neither Class should have an advantage over the other with regard to the majority of the tasks without the effectiveness of incidental recasts.

The Kolmogorov-Smirnov Test showed that some of the data were not normally distributed ($p < .001$). In addition, the Levene's Test showed that the variances were significantly different among the groups in some cases ($p = .006$). Because the data violated the assumptions of parametric data, non-parametric tests, specifically, the Kolmogorov-Smirnov Z test and the Wilcoxon signed-rank test were used in the following statistical analyses. The former was used in order to compare the scores of each group (addressees and auditors, major and minor, and Class C) and the latter was used in order to see how the scores changed between the two posttests. In addition, whether the scores of the addressees and the auditors fall within the range of the confidence interval of the scores of Class C was also examined in order to detect statistical differences that may not have been shown due to the small sample sizes. The small sample sizes also affected the reliability of the confidence intervals for the addressees and the auditors, and therefore, these two groups were simply compared by their mean scores. Admittedly, this may not be the optimal method of analysis; however, without data from a pretest, it seemed to be the most valid procedure. The alpha level was set at .05. SPSS 16 was used for all the statistics.
3 Results

Table 4 shows the total number of recasts provided in each Class. The data of the two target structures that received the most incidental recasts (preposition and third person singular; henceforth, major target structures) and the least incidental recasts (modal verb and embedded question; henceforth, minor target structures) were combined respectively for the following analyses, due to the small sample size.

<table>
<thead>
<tr>
<th>Class</th>
<th>Preposition</th>
<th>Third person singular</th>
<th>Modal verb</th>
<th>Embedded question</th>
<th>Other errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>13</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>64</td>
</tr>
<tr>
<td>E2</td>
<td>4</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>69</td>
</tr>
<tr>
<td>E (E1&amp;E2)</td>
<td>17</td>
<td>12</td>
<td>6</td>
<td>5</td>
<td>133</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>35</td>
</tr>
</tbody>
</table>

Tables 5-6 show the descriptive statistics.

Table 5 The Descriptive Statistics of the Major Target Structures

<table>
<thead>
<tr>
<th>Class</th>
<th>Addressee</th>
<th>Auditors</th>
<th>Difference 1</th>
<th>Class</th>
<th>Difference 2</th>
<th>Difference 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>(N = 5)</td>
<td>(N = 8)</td>
<td></td>
<td>C</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Posttest 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOIT</td>
<td>.7000</td>
<td>.2739</td>
<td>.4375</td>
<td>.1768</td>
<td>.2625</td>
<td>.5547</td>
</tr>
<tr>
<td>TGJT</td>
<td>.1250</td>
<td>.1768</td>
<td>.0000</td>
<td>.0000</td>
<td>.1250</td>
<td>.2578</td>
</tr>
<tr>
<td>UGJT</td>
<td>.9500</td>
<td>.1118</td>
<td>.6250</td>
<td>.1890</td>
<td>.3250</td>
<td>.8281</td>
</tr>
<tr>
<td>MKT</td>
<td>.4000</td>
<td>.5477</td>
<td>.1875</td>
<td>.3720</td>
<td>.2125</td>
<td>.3438</td>
</tr>
<tr>
<td>Posttest 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EOIT</td>
<td>.6500</td>
<td>.2850</td>
<td>.4844</td>
<td>.1695</td>
<td>.1656</td>
<td>.5859</td>
</tr>
<tr>
<td>TGJT</td>
<td>.1750</td>
<td>.2092</td>
<td>.1562</td>
<td>.1294</td>
<td>.0188</td>
<td>.3250</td>
</tr>
<tr>
<td>UGJT</td>
<td>.7750</td>
<td>.2236</td>
<td>.6857</td>
<td>.2216</td>
<td>.0893</td>
<td>.8571</td>
</tr>
<tr>
<td>MKT</td>
<td>.4000</td>
<td>.5477</td>
<td>.4375</td>
<td>.4955</td>
<td>-.0375</td>
<td>.3125</td>
</tr>
</tbody>
</table>

\[ ^a \text{Some participants were either addressees or auditors for both structures. Therefore, the number of the addressees and the auditors will not add up to the total number of participants}. \]

\[ ^b \text{Addressees} \quad ^c \text{Comparing addressees and Class C}. \]

\[ ^d \text{Comparing auditors and Class C}. \quad \text{The Elicited Oral Imitation Task}. \quad ^e \text{The Timed Grammaticality Judgment Task}. \quad ^f \text{The Untimed Grammaticality Judgment Task}. \]

\[ ^g \text{The Metalinguistic Knowledge Task}. \]
### Table 6 The Descriptive Statistics of the Minor Target Structures

<table>
<thead>
<tr>
<th>Tests</th>
<th>Tasks</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Mean</th>
<th>SD</th>
<th>Class C (N = 16)</th>
<th>Difference 1&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Difference 2&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Difference 3&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posttest 1</td>
<td>EOID&lt;sup&gt;d&lt;/sup&gt;</td>
<td>.2917</td>
<td>.2602</td>
<td>.2679</td>
<td>.2741</td>
<td>.0238</td>
<td>.2969</td>
<td>.1106</td>
<td>Ad=C</td>
<td>Au=C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGJT&lt;sup&gt;e&lt;/sup&gt;</td>
<td>.0000</td>
<td>.0000</td>
<td>.0357</td>
<td>.0945</td>
<td>.0357</td>
<td>.1875</td>
<td>.1768</td>
<td>Ad&lt;C</td>
<td>Au&lt;C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UGJT&lt;sup&gt;f&lt;/sup&gt;</td>
<td>.5000</td>
<td>.0000</td>
<td>.6071</td>
<td>.1967</td>
<td>-.1071</td>
<td>.7266</td>
<td>.2103</td>
<td>Ad&lt;C</td>
<td>Au&lt;C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MKT&lt;sup&gt;g&lt;/sup&gt;</td>
<td>.0000</td>
<td>.0000</td>
<td>.2143</td>
<td>.3934</td>
<td>-.2143</td>
<td>.2500</td>
<td>.3162</td>
<td>Ad&lt;C</td>
<td>Au=C</td>
<td></td>
</tr>
<tr>
<td>Posttest 2</td>
<td>EOID</td>
<td>.1250</td>
<td>.2165</td>
<td>.2321</td>
<td>.2925</td>
<td>-.1071</td>
<td>.3203</td>
<td>.1115</td>
<td>Ad&lt;C</td>
<td>Au&lt;C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TGJT</td>
<td>.0417</td>
<td>.0722</td>
<td>.0893</td>
<td>.1189</td>
<td>.0476</td>
<td>.3000</td>
<td>.1787</td>
<td>Ad&lt;C</td>
<td>Au&lt;C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>UGJT</td>
<td>.5417</td>
<td>.1909</td>
<td>.6250</td>
<td>.2500</td>
<td>-.0833</td>
<td>.7076</td>
<td>.2704</td>
<td>Ad&lt;C</td>
<td>Au=C</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MKT</td>
<td>.0000</td>
<td>.0000</td>
<td>.0714</td>
<td>.1890</td>
<td>.0714</td>
<td>.4062</td>
<td>.2720</td>
<td>Ad&lt;C</td>
<td>Au&lt;C</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>Addressees - Auditors. <sup>b</sup>Comparing addressees and Class C. <sup>c</sup>Comparing auditors and Class C. <sup>d</sup>The Elicited Oral Imitation Task. <sup>e</sup>The Timed Grammaticality Judgment Task. <sup>f</sup>The Untimed Grammaticality Judgment Task. <sup>g</sup>The Metalinguistic Knowledge Task.

#### 3.1 Comparing Addressees and Auditors

For the major target structures, the scores of the addressees were higher than those of the auditors in all tasks in the Posttest 1, and in three tasks in the Posttest 2. In contrast, for the minor target structures, the scores of the auditors were higher than those of the addressees in three tasks in the Posttest 1, and all tasks in the Posttest 2. No case showed a significant difference in mean scores ($p > .121$).

#### 3.2 Comparing Addressees and Class C

For the major target structures, there were two cases in which the scores of the addressees were above the confidence interval of Class C: the Elicited Oral Imitation Task and the Untimed Grammaticality Judgment Task in the Posttest 1. In other cases, the scores of the addressees were either within the range of (3 tasks) or below the confidence interval of Class C (3 tasks). In contrast, for the minor target structures, there was no case in which the score of the addressees was above the confidence interval of Class C; the score was below the confidence interval in most cases (7 tasks). No case showed a significant difference in mean scores ($p > .117$).

#### 3.3 Comparing Auditors and Class C

For the major target structures, there was no case in which the score of the auditors was above the confidence interval of Class C; the scores of the auditors were below the confidence intervals of Class C in most cases (7 tasks). Similarly, for the minor target structures, there was no case in which the score of the auditors was above the confidence interval of Class C; the scores were either within the range of (3 tasks) or below the confidence interval (5 tasks). Two cases showed significant differences in the mean scores at the Posttest 1 (the Timed Grammaticality Judgment Task, $p = .002$; the Untimed Grammaticality Judgment Task, $p = .031$);
however, the other 14 cases showed no significant differences ($p > .117$).

3.4 Score Change Between the Two Posttests

The Kolmogorov-Smirnov Z test showed that the scores of the Posttest 2 were significantly higher than those of the Posttest 1 in three out of the eight tasks in Class C: the Timed Grammaticality Judgment Task for the major target structures ($p = .037$) and for the minor target structures ($p = .021$), and the Metalinguistic Knowledge Task for the minor target structures ($p = .025$). Together, the data seems to suggest that there was a test effect in the Timed Grammaticality Judgment Task. In fact, all the treatment groups (addressees and auditors, major and minor) showed an increase in scores in the Timed Grammaticality Judgment Task between the two posttests, one of which showed a significant increase (auditors for the major target structures, $p = .025$); therefore, it is likely that this was due to a test effect. Thus, overall, excluding the test effect, the scores did not significantly change from the Posttest 1 to 2 in any tasks.

4. Discussion

First of all, the addressees in this study made errors during the tasks, and therefore, it can be said that they did not have a great amount of implicit knowledge of the target structures during the course of the present study. On the other hand, the auditors did not make any errors regarding the target structures. There are two speculative reasons for this: (a) they already possessed a high level of implicit knowledge, or (b) they learned from the interactions of other learners, including recasts. However, if their scores were low, it can be said that neither of these hypotheses are valid.

The results showed that, for the minor target structures, there was no case in which the score of the addressees or the auditors was above the confidence interval of the score of Class C. Therefore, it seems that a small number of incidental recasts was not effective regardless of whether the learners were addressees or auditors. However, in seven out of the eight tasks, the scores of the auditors were higher than those of the addressees. Thus, there is a possibility that recasts were effective to some extent only for the auditors, or simply, the auditors may already have possessed greater implicit and explicit knowledge of the minor target structures than did the addressees. However, without a pretest, such a question is not answerable in the present study.

The results of the major target structures showed a different pattern. The scores of the addressees were higher than those of the auditors in seven out of the eight tasks; in addition, the scores of the addressees were above the confidence intervals of the score of Class C in two tasks in the Posttest 1. In contrast, the scores of the auditors were below the confidence intervals of the scores of Class C in seven tasks. Thus, when the number of incidental recasts provided was relatively large, it seems that the recasts were effective for the addressees, but not for the auditors; however, the effectiveness for the addressees seemed to disappear in the Posttest 2.

The ineffectiveness of recasts for auditors found in this study is in line with the results of some previous studies (Mackey, 1999; Williams, 2001). Consideration of a few possible reasons are in order. First, as Truscott (1999) claimed, other
learners may not have been ready to intake recasts provided for an addressee. Second, because the present study was conducted outside their regular class with no effect on their grades, the participants may not have been very keen on learning during the treatment sessions, paying less attention to recasts provided to the other participants. Therefore, it is possible that, in a credit-bearing class, CF for auditors may prove effective.

5. Limitations

The first limitation of the present study is its small sample size, due to the large number of attritions. The second possible limitation is the lack of longitudinal data. For logistical reasons, the delayed posttest was conducted only a week after the immediate posttest, which was admittedly shorter than most previous studies. However, an inevitable problem that stems from a delayed posttest is that the later the delayed posttest is conducted, the greater the possibility that the participants will be exposed to the target structures. One possible solution is to choose a target structure that the participants will not likely learn in their regular classes; however, this was not a practical solution in the present study. The other limitation is that it was assumed that the proficiency level of the participants were equal between the two Classes because they belonged to the same level of classes in the ESL program. However, even among the participants in the same level, each participant had different strengths and weaknesses in grammatical knowledge and the four skills. Judging from the fact that the scores of Class E were mostly below the confidence intervals of those of Class C, it may be safe to say that the performance of the participants in Class C was better than that of Class E from the outset. This limitation could have been solved by the use of a pretest, the lack of which is the most crucial limitation of the present study. Although it was possible to predict which structures were more likely to be dealt with in treatment sessions from the task materials, this did not guarantee that CF would be provided for them. This was especially true because there were two Classes involved, and it was likely that some of them might receive more CF than the others.

6. Conclusion

The present study showed the importance of taking into account the number of recasts provided and the position of the participants (i.e., addressee or auditor) when investigating the effectiveness of incidental recasts. To my knowledge, this is the only study thus far that has investigated this issue while examining the effectiveness of recasts for each individual grammatical structure. While the present study certainly contributes to the current body of literature on the effectiveness of incidental recasts, this research topic still awaits further empirical inquiries.
Notes.
1 Originally, the former two tasks were used to measure implicit knowledge whereas the latter two tasks were used to measure explicit knowledge, following Ellis (2005, 2006), on the assumption that recasts would have different effects on these two types of knowledge, as shown in many previous studies (e.g., Ellis et al., 2009). However, the results did not show any differences between these two types of knowledge. Thus, these four tasks were used simply as four different tasks to measure learners’ performance.
2 The test items also had an equal number of old items and new items in each task. Old items were the structures that the participants used during one of the activities whereas new items were those that they did not. The former was to measure item learning whereas the latter was to measure rule learning. However, due to the small sample size, only the overall scores were reported.
3 Criteria to test the validity of this calculation has not been established to date, and therefore, the way that the time limit is set differs in various studies. In the present study, the calculation used in Ellis (2005, 2006) was adopted because the tasks were made based upon his study.

References


