The Influence of Timing of Form Instruction During Practice Using Task Repetition on Task Performance

Masaki DATE
University of Fukui

Abstract

This study focused on how timing of form instruction during task repetition influences development of proceduralization and accuracy of linguistic knowledge. Following a pretest using a narrative task, two treatment groups repeated narrative tasks in practice sessions. In each session, Group 1 was given time to find and correct errors in the instructional material as form instruction after the first task, and then repeated the first task. Conversely, Group 2 first received form instruction then performed the task two times. The control group was not given any instruction sessions. After the sessions, a posttest with the same pictures as the pretest was administered. The results showed that Groups 1 and 2 performed better than the contrast group, and although Group 1 performed more fluently than Group 2, only Group 2 improved accuracy. This study implies that: (1) practice using task repetition accompanied by form instruction is effective for facilitating proceduralization and accuracy; and (2) the timing of form instruction brings different effects for developing proceduralization and accuracy.

1. Background

According to VanPatten, Williams and Rott (2004), establishing Form-Meaning Connections (FMCs), connections between a language form and its meaning, is a fundamental aspect of language acquisition for both first (L1) and second languages (L2). Tasks can become a way to help learners notice FMCs in individual exemplars. One of the merits in output through tasks is to facilitate learners’ syntactic processing and their consciousness of forms, especially grammar (Swain, 1985). Such learners’ consciousness of forms can then help learners notice a language item, facilitate fixing its structure and accelerate the speed of language acquisition. However, it is questionable if learners are conscious of forms while speaking. First of all, because of the cognitive burden, they have limited capacity in what they can pay attention to and so tend to focus first on the content of their speech at conceptualization and then on the form to represent its meaning at formulation (Levelt, 1989). Learners’ tendency to pay more attention to meaning than form in speaking has been supported by several findings: recasts were adjudged not feedback on erroneous forms but negotiation of meaning (Ellis & Sheen, 2006), and most of the language-related episodes learners
produced were not syntactically oriented (Williams, 1999). It is thus possible that learners will not be particularly conscious of language forms simply by doing a speaking task.

Besides setting a task like a consciousness-raising task or a dictogloss, whereby learners naturally or essentially pay attention to forms, there are two ways to help learners be conscious of language forms during a speaking task. One is to manipulate how the task is done, with task repetition being the way in this case. Fukuta (2016) examined participants’ attention shifts during repeated task engagement. It was then clarified that the participants focused more on the conceptualizing process and less on the syntactic encoding process in the first task; however, it worked vice versa, with participants focusing more on the syntactic encoding process in the second, repeated task. The second way is to give learners some triggers for forms either before, while or after doing a task, such as form instruction, feedback (recasts/prompts), and reflection through transcription of the task performance.

1.1 Task repetition

Task repetition has been clearly identified for its effectiveness in improving language performance the second time around compared with the first time. A possible reason for this consistent improvement in the second performance may be that suggested by Bygate (1999): students are likely to focus initially on message content, then, subsequently, once the message content and the basic language needed to encode it have been established in the first performance, to switch their attention to the selection and monitoring of appropriate language in the second performance, as Fukuta (2016) proved. Alternatively, it may be the process proposed by Bygate (2001, p. 29): “part of the work of conceptualization, formulation and articulation carried out on the first occasion is kept in the learners’ memory store and can be reused on the second occasion.” However, these studies do not show precisely what cognitive processes are at work to improve the second performance; that is, if there is some change in the learner’s L2 knowledge representation.

Different from the previous studies which mainly focused on task performance right after task repetition, De Jong and Perfetti (2011) focused on the effectiveness of sessions using task repetition for proceduralization (i.e., a process of storing and developing procedural knowledge of the skill or cognitive act stored in the production system) leading to fluency development and examined their performances at the posttests. It was then concluded that task repetition in the sessions may cause changes in learners’ underlying cognitive mechanisms, and result in an increase in fluency and the long-term retention of the increased fluency. The mechanism of task repetition, keeping the work carried out on the first occasion in the learners’ memory store and reusing it on the second occasion, implies the importance of making them reuse what is kept in their memory after the previous performance for proceduralization and/or (implicit) learning. According to the Adaptive Control of Thought-Rational theory, when declarative knowledge stored in the declarative module is retrieved repeatedly, this process helps procedural knowledge in the production system to develop (Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004). It may be expected that task repetition can
facilitate learners’ consciousness of language forms at the second performance and that such repeated experience can lead to the development of proceduralization.

1.2 Timing of form instruction

Task repetition can work in facilitating learners’ consciousness of language forms at the second performance and, through repeated experience, lead to developing fluency. On the other hand, the combination of output (such as production-based instruction) and form instruction is effective for developing accuracy (Ellis & Shintani, 2014). Van de Guchte, Braaksma, Rijlaarsdam, and Bimmel (2016) examined whether the combination of task repetition and form instruction through prompts given in the first task was effective. They also found its effect for written accuracy as well as metalinguistic knowledge at the posttests compared with only form instruction in the first task (no repetition). Therefore, task repetition with noticing can be effective in facilitating not only fluency but also accuracy. In the study by Date (2015), where participants were given form instruction after the first task, they improved fluency and accuracy in the same task, and raised their consciousness of forms, leading to correct modifications while speaking. However, it is unclear if the ways in which learners are made to pay attention to forms have any influence. Timing of form instruction, i.e., when to give learners form instruction, is related to this issue.

There are two types of timing for form instruction. One is before the first task is conducted: PPP (Presentation-Practice-Production; Skehan, 1998), task-supported language teaching (Ellis, 2003), sequential focus on form (Doughty & Williams, 1998), and strategic planning are examples. The other is after the first task has been conducted. In Samuda (2001), participants first were given a task and created “a semantic ‘space’ which learners’ current IL [(interlanguage)] resources may be able to fill only partially” (p. 122), then target forms (TFs) were introduced and finally participants performed the task of preparing and presenting a poster. It was found that they could produce TFs at the second task. Sheppard (2006) studied the effectiveness of feedback provided after the first performance and found its effectiveness for improving accuracy, fluency and complexity in the repeat performance. Hawkes (2012) also investigated the effects of the combination of task repetition and form instruction given after the first task. The results showed the number of error corrections increased, implying greater attention to form was paid during the second task than the first task. Form instruction after the first task in task repetition thus has had its effectiveness proved in recent studies.

As seen above, either timing of form instruction seems to bring some benefit in task performance after the instruction. However, it is not clear whether different timings of form instruction with task repetition given in sessions influence development of proceduralization and accuracy and lead to different degrees of proceduralization and accuracy.
2. Study

2.1 Hypothesis
The overriding research questions were whether learners would facilitate proceduralization and/or increase in accuracy by practicing task repetition with form instruction, and if so, when form instruction should be given for more proceduralization and/or accuracy. The following two hypotheses were thus set:

1. Gains in proceduralization and accuracy from a pretest to a posttest would be larger for learners who had training sessions than for those who did not.
2. The gains would be larger for learners who received form instruction after the first task in the sessions than for those who received it before the first task in the sessions.

2.2 Procedure
The participants were university students taking English conversation classes once a week. Their TOEFL PBT scores ranged from 445 to 545. They were randomly assigned into three groups. Groups 1 (n = 15) and 2 (n = 15) were treatment groups given training sessions where they had task repetition with form instruction. Group 3 (n = 15) was a contrast group that was not given any sessions. Group 1 did the same task two times and had chances to notice errors after conducting the first task. Group 2 also repeated the same task, but received form instruction before the first task. Each group followed the procedure on different days. They had the same test two times as a pretest and a posttest, and four training sessions in four weeks (i.e., once a week).

A pretest was given one week before the first session, and a posttest given one week after the final session. In each test and training session, they performed a narrative task using a six-picture cartoon strip. In the study by Bygate (2001), after doing a task as a pretest, the experimental groups were given two tasks in each of the four sessions, the same task type but with different topics. In the final session, all groups performed the same task as in the first week and a new task different from the first week. Referring to Bygate’s procedure seemed valid in order to examine if task repetition with form instruction in each session would influence participants’ performance on a task as a posttest after the sessions. However, different from Bygate (2001) where two posttests, one with the same picture story as the pretest and one with a new picture story, were used, only one posttest, with the same picture story as the pretest, was administered. It is because there was the possibility that the difficulty of telling a story depends on the features of the pictures used. Tavakoli and Foster (2008) examined the effect of narrative type on learner output by using two types of narrative structures: tight (where the order of events was fixed), and loose (where the events could be reordered without loss of coherence). They found more accurate performance on tight tasks than loose tasks and pointed out the probability that the tight ordering of events releases attentional resources. For the purpose of deleting such influence of the pictures used, only the same picture...
story was used at the pretest and the posttest here.

Each session, as well as the posttest, was given during a class, without announcing to the participants beforehand that they would have four sessions and a posttest. In the sessions, each participant in Group 1 was first given an IC recorder and a sheet with a six-picture cartoon strip. Once they had had a quick look at the cartoon and made sure they understood the meaning of each picture, the researcher announced that they were to perform a narrative task by talking into the IC recorder for up to 90 seconds. They could stop recording by themselves if they finished talking before the 90 seconds had elapsed.

After telling the story the first time and turning over the sheet, each participant was given both a new sheet with a passage on it and a tablet PC. The passages given in each week’s sessions had different content, based on a topic, such as music, food or sports. The passages had been selected from those written by different students in different classes and had also been premodified by the researcher to contain similar total numbers of words, sentences, T-units, errors and TF errors. The average was 276.25 words, 22 sentences, 27.5 T-units and 18.25 errors including 11 erroneous TFs, either verbs or articles. The numbers of TFs included in each passage were as follows: five articles and six verbs (Session 1), eight articles and three verbs (Session 2), six articles and five verbs (Session 3), and six articles and five verbs (Session 4). All errors had been numbered in front and highlighted with an underline or a blank space by the researcher. The participants then corrected the errors underlined or filled in the blanks. When self-correction was completed, they compared their corrections to the same passage explicitly corrected on the PC. When they touched an error on the screen, a linked page came up and showed the correct form along with some metalinguistic explanation in Japanese. While they were checking, the researcher circulated and verified whether each participant now understood all the errors.

When all the participants had corrected the errors and checked them on the PC, in around 15 minutes, the original picture sheet was turned over. Then a new sheet with the same six-picture cartoon strip as for the first performance was given again, and the group told the story again into the IC recorder for up to 90 seconds. After telling the story for a second time, each learner’s three sheets, the IC recorder and tablet PC were collected.

As for Group 2, each participant was first given both a new sheet of pictures with a written passage on it and a tablet PC, so they corrected the errors and then checked them on the PC in around 15 minutes (i.e., the same as Group 1 but before doing the first task). After that sheet was turned over, they were given an IC recorder and a new sheet with a six-picture cartoon strip on it, and then started talking into the IC recorder for up to 90 seconds. Soon after telling the story and turning over the sheet, a new sheet with the same pictures as for the first performance was given again. The group simply told the same story again, immediately, into the IC recorder for up to 90 seconds. After the second task performance, they returned three sheets, the IC recorder and tablet PC.

A week after the last session (Session 4), a posttest was given with the same pictures as in the
pretest. Right after the posttest, two points had been announced orally: (1) the recorded data on each IC recorder would be used later in presentations at conferences or in papers, but their names would not be made public; and, (2) those who wanted their data to be omitted should declare so either now or later. Those who did not declare so were then adjudged to have approved of the use of their data.

2.3 Analysis

PRAAT 5.3.09 (Boersma & Weenink, 2012) was used to analyze all data on the IC recorders. The researcher transcribed all information and determined the beginning and end of each speech segment by using the PRAAT function ‘To textgrid (silences).’ The researcher then checked all pause boundaries and adjusted them as necessary, by listening to the recording and visually inspecting the spectrogram and wave-form. Nonverbal fillers, such as “uh,” and “mmm”, were not transcribed. Syllables were counted only from words uttered in their entirety, excluding false starts. Words repeated were also counted. Referring to De Jong and Perfetti (2011), the following three criteria were set: (1) any filler or silence, which was 0.20s or longer, was treated as a pause; (2) in each speech of each participant, the upper limit for pauses was set to 2.5 standard deviations above the mean, and a pause longer than that was replaced with the mean pause length plus 2.5 standard deviations; and, (3) for measuring proceduralization, two measures were used, namely, the mean pause length calculated by dividing the total length of pauses by the number of pauses and the mean length of fluent runs calculated by dividing the number of syllables by the number of pauses.

Verbs and articles were set as the TFs: verb errors included errors in tense, word choice, and subject-verb agreement, and article errors encompassed all types of article misuses. In the passages used for form instruction in the sessions, erroneous TFs were also intentionally embedded. All TFs produced by each participant were confirmed by the researcher with a native speaker of English. The numbers of erroneous TFs out of the TFs used (i.e., the combined number of erroneous forms and correct forms) were then tabulated. For measuring accuracy, the total error rates of TFs were calculated by dividing the number of erroneous TFs by the total number of the TFs used, and the error rates of the TFs in fluent runs were determined by dividing the total error rates of TFs by the mean length of fluent runs. The rates in fluent runs targeted how accurately they performed in fluent speech.

3. Results

3.1 Proceduralization

Table 1 shows the mean scores, standard deviations and median scores. No significant group difference was found at the pretest in either mean lengths of pauses, $F(2, 42) = 1.658, p = .203$, or mean lengths of fluent runs, $F(2, 42) = .157, p = .856$, indicating their parity in fluency at first.

Figure 1 presents the data distributions of the mean lengths of pauses produced by the three groups at each test. It can be seen that the median scores for pause lengths of the treatment
Table 1

Descriptive Statistics of Measures of Proceduralization

<table>
<thead>
<tr>
<th></th>
<th>Group 1 (n = 15)</th>
<th>Group 2 (n = 15)</th>
<th>Group 3 (n = 15)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( M )</td>
<td>( SD )</td>
<td>( Mdn )</td>
</tr>
<tr>
<td>mean lengths of pauses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pretest</td>
<td>1.269</td>
<td>.296</td>
<td>1.205</td>
</tr>
<tr>
<td>posttest</td>
<td>.990</td>
<td>.233</td>
<td>.980</td>
</tr>
<tr>
<td>mean lengths of fluent runs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pretest</td>
<td>2.371</td>
<td>.390</td>
<td>2.381</td>
</tr>
<tr>
<td>posttest</td>
<td>2.671</td>
<td>.552</td>
<td>2.538</td>
</tr>
</tbody>
</table>

groups became lower in the posttest than at the pretest whereas the scores of Group 3 looked similar between the two tests. Furthermore, Figure 1 illustrates that the median scores of Group 1 were the smallest at the posttest. A two-way repeated measures ANOVA revealed a medium sized effect in the interaction between test and group, \( F (2, 42) = 2.991, p = .061, \eta^2 = .125 \). Looking at simple main effects by test, significant differences were found in Group 1, \( t (14) = 4.504, p < .001, r = .649 \), and Group 2, \( t (14) = 3.167, p = .007, r = .514 \). Both groups produced significantly shorter pauses at the posttest than at the pretest with a large sized effect. Next, looking at simple main effects by group, a group difference was found at posttest, \( H (2) = 5.343, p = .069 \). Multiple comparisons by Mann-Whitney test with Bonferroni correction (\( \alpha = .016 \)) indicated that Group 1 made shorter pauses than Group 2 with a medium sized effect, \( z = -1.970, p = .049, r = -.360 \), and Group 3 also
with a medium sized effect, \( z = -2.012, p = .044, r = -.368 \). In short, although all groups first produced similar lengths of pauses, at the posttest the treatment groups made shorter pauses than at the pretest; especially, Group 1 made shorter pauses than Group 3 and even than Group 2. Group 3 did not show any improvement.

Figure 2 shows the data distributions of the mean lengths of fluent runs. The median score of Group 2 was larger in the posttest than at the pretest. The score of Group 2 at the posttest was also larger than the other two groups although it varied more than the others. A two-way repeated measures ANOVA found a medium sized effect in the interaction between group and test, \( F(2, 42) = 2.731, p = .077, \eta^2_p = .115 \). First, looking at simple main effects by test, significant differences were found in Group 1, \( t(14) = -2.611, p = .021, r = .443 \), and Group 2, \( t(14) = -5.636, p = .000, r = .730 \). Group 1 produced significantly longer runs in the posttest than at the pretest, and the effect size was medium. Group 2 also produced significantly longer runs in the posttest than at the pretest, and the effect size was large. Looking at simple main effects by group, no significant difference between the groups was found at the posttest, \( F(2, 42) = 1.493, p = .236, \eta^2_p = .067 \). In summary, all groups produced similar lengths of fluent runs at the pretest and the posttest; however, the treatment groups produced longer runs in the posttest than the pretest, but Group 3 did not.

### 3.2 Accuracy

Table 2 shows the mean scores, standard deviations and median scores at each test. There was no significant group difference found at the pretest in either error rates of TFs in total, \( F(2, 42) = .440, p = .647 \), or error rates of TFs in fluent runs, \( H(2) = .042, p = .979 \), indicating their parity in accuracy at first.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Descriptive Statistics of Measures of Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Group 1 (n = 15)</td>
</tr>
<tr>
<td></td>
<td>( M )</td>
</tr>
<tr>
<td>error rates of TFs in total</td>
<td></td>
</tr>
<tr>
<td>pretest</td>
<td>.428</td>
</tr>
<tr>
<td>posttest</td>
<td>.410</td>
</tr>
<tr>
<td>error rates of TFs in fluent runs</td>
<td></td>
</tr>
<tr>
<td>pretest</td>
<td>.187</td>
</tr>
<tr>
<td>posttest</td>
<td>.162</td>
</tr>
</tbody>
</table>

Figure 3 shows the data distributions of the error rates of TFs in total produced by the three groups at each test. Compared with the pretest, the median score in Group 2 was lower, and the score in Group 3 was higher at the posttest. Figure 3 also illustrates that the median score in Group 2 was the lowest while the score in Group 3 was the highest at the posttest. A two-way repeated
measures ANOVA on the error rates of the TFs revealed a medium sized effect in the interaction between group and test, $F(2, 42) = 2.168, p = .127, \eta_p^2 = .093$. Looking at simple main effects by test, a significant difference was found in Group 2 only with a medium sized effect, $t(14) = 2.629, p = .020, r = .445$, and Group 2 had a lower error rate in the posttest than at the pretest. Next, looking at simple main effects by group, a significant difference between the groups was found at the posttest, and the effect size was medium, $F(2, 42) = 3.295, p = .047, \eta_p^2 = .137$. Multiple comparisons by $t$ test with Bonferroni correction indicated that the ratio in Group 1 was lower than Group 3 with a medium sized effect, $t(28) = -1.844, p = .076, r = .330$, and the ratio in Group 2 was also lower than Group 3 with a medium sized effect, $t(28) = -2.517, p = .018, r = .430$. In other words, all groups first performed with equally high erroneous rates; however, at the posttest, the treatment groups produced erroneous TFs less than Group 3, with only Group 2 performing with a significantly lower rate than at the pretest.

Figure 4 illustrates the data distributions of error rates of TFs in fluent runs. It can be seen that, compared with the pretest, the median score in Group 2 was lower at the posttest. It also illustrates that the median score in Group 2 was lower than the other two groups at the posttest. A two-way repeated measures ANOVA on the error rates of the TFs showed a medium sized effect in the interaction between group and test, $F(2, 42) = 3.132, p = .054, \eta_p^2 = .130$. First, looking at simple main effects by test, a significant difference was found in Group 2 only with a large sized effect, $t(14) = 4.952, p < .001, r = .684$, and Group 2 produced significantly fewer erroneous TFs in the posttest than at the pretest. Looking at simple main effects by group, a significant difference between the groups was found at the posttest with a large sized effect, $F(2, 42) = 4.347, p = .019$. 

\[ \text{Figure 3. Distribution of error rates of TFs in total.} \]

\[ \text{Figure 4. Distribution of error rates of TFs in fluent runs.} \]
\(\eta^2 = .170\). Multiple comparisons by \(t\) test with Bonferroni correction revealed that the ratio in Group 2 was significantly lower than Group 3, and the effect size was medium, \(t (28) = 3.012, p = .005, r = .495\). In short, all groups performed equally at the pretest with higher erroneous rates in fluent runs; however, only Group 2 performed in the posttest with lower rates than it had at the pretest, and Group 2 also produced lower erroneous rates than Group 3 at the posttest.

### 4. Discussion

Table 3 is the summary of the data analysis of proceduralization and accuracy in the tests. Hypothesis 1 was that gains in proceduralization and accuracy from a pretest to a posttest would be larger for learners who had training sessions than for those who did not. Looking at proceduralization, it was found in within-group comparisons that the treatment groups not only produced longer fluent runs but also had shorter pauses in the posttest than at the pretest. However, the contrast group did not change in either the lengths of pauses or fluent runs from at the pretest. Such improved performance and gains in the treatment groups indicate the effectiveness of training sessions for facilitating proceduralization. On the other hand, in between-group comparisons, all groups were similar in the lengths of pauses and fluent runs at the pretest, and there was also no group difference in the lengths of fluent runs at the posttest. However, Group 1 needed shorter pauses than the contrast group to produce a similar length of fluent runs to the contrast group at the posttest. This means that Group 1 performed more fluently than the contrast group at the posttest, but Group 2 did not.

As for accuracy, it was found that, although the rates of erroneous TFs in total were similar between groups at first, the rates produced by the treatment groups were lower than those of the contrast group at the posttest. On the other hand, the rates of erroneous TFs in total produced by Group 2 became lower in the posttest than at the pretest, but the rates in Group 1 did not improve. Similarly, the error rates of TFs in fluent runs of Group 2 became lower in the posttest than at the pretest, but Group 1’s rates did not. In addition, although the error rates in fluent runs were similar between all groups at the beginning, only Group 2 produced a lower rate than the contrast group in
the posttest. In other words, when looking at error rates in total, whereas the treatment groups performed more accurately at the posttest than the contrast group, only Group 2 improved accuracy from the pretest. In addition, only Group 2 performed more accurately in fluent runs than the contrast group at the posttest and improved accuracy from the pretest.

Therefore, Hypothesis 1 is supported. It can then be said that training sessions were effective for improving proceduralization from before the sessions and for facilitating more accuracy in total after the sessions than no sessions.

Hypothesis 2 was that gains in proceduralization and accuracy from a pretest to a posttest would be larger for learners who received form instruction after the first task in the sessions than for those who did so before the first task in the sessions. As for proceduralization, both Group 1 and Group 2 had improved in the lengths of pauses and fluent runs in the posttest from at the pretest, and the two groups had been similar in the lengths of pauses and fluent runs at the pretest. However, in the posttest, a significant group difference was found in the lengths of pauses, with Group 1 needing shorter pauses than Group 2 in order to produce similar lengths of fluent runs to Group 2. Group 2 thus showed better performance at the posttest in proceduralization.

Looking at accuracy, it was found that there was no significant difference in the rates of erroneous TFs between Group 1 and Group 2 both in total and fluent runs. However, only Group 2 improved its error rates in the posttest from at the pretest both in total and fluent runs. Furthermore, only Group 2 performed more accurately than the contrast group in the posttest in total and fluent runs.

Therefore, Hypothesis 2 is partially supported. It can then be said that giving form instruction after the first task was effective for facilitating greater proceduralization than form instruction before the first task, whereas giving form instruction before the first task was effective for improving accuracy from before the sessions, but giving it after the first task was not effective. The study by Hawkes (2012) concluded that form instruction after the first task helped the participants pay more attention to form in the second task. However, referring to the results of Group 2 in this study, it is possible that such attention to forms before the second task may not be sufficient to develop their accuracy. This implies the effects of practice can be brought to proceduralization and accuracy differently, depending on when form instruction is given during the practice.

Both Groups 1 and 2 were given exactly the same form instruction in each session; the only difference was the timing of the instruction given. Both groups thus had opportunities to notice an erroneous TF and correctly connected the form to its meaning/function. As Ellis (2016) elucidated, “Pre-teaching [a TF] may help learners to attend to it while communicating and may also facilitate learning” (p. 413). Moreover, because they also have more opportunity to practice using the TF in the second task, “this may probably have raised or re-awakened their metalinguistic awareness [of it]” (Van de Guchte et al., 2016, p. 313). In other words, their consciousness of the connection might have been raised before the second task performance, and this raised consciousness might then have influenced their repeatedly retrieving declarative knowledge of the connection in the second
performance during the sessions. Consequently, it had been expected that when they undertook the posttest, not only the process of conceptualization but also that of formulation might have been boosted through the repeated practice in sessions, engendering more fluency and accuracy.

Then, why did Group 1 perform more fluently than Group 2, and why did only Group 2 improve accuracy in producing TFs from the pretest, at the posttest? Two factors could be related to the answer. One is the order of priority of learners’ attention to meaning and form in speaking. The other is the timing of form instruction. These two factors might have influenced the participants’ attention during the second task, and such repeated behavior through the sessions might have produced these impacts on their performance at the posttest. In the present study, Group 1 was asked to repeat the first task after feedback on erroneous TFs was given. Since previous knowledge of the TFs is available to them to build on in the second task, they had been expected to use the TFs more accurately than in the first task. However, because they might have needed to activate conceptualization again in the second task after the 15-minute instruction, it could be necessary for them to allocate their attentional capacity for remembering the content of what they had said in the first task. Naturally, they must have allocated less attention to the TFs which they were exposed to in the instruction, but they practiced accessing their linguistic resources (whatever they needed for task performance and had in IL system) quickly. Such repeated practice of quick access to their IL resources could result in developing their proceduralization.

On the other hand, Group 2, who were given form instruction at first, repeated the same task soon after the first task was completed. It could then be easier for them to remember the content of what they had just talked about in the first task. Consequently, they might have needed to activate their attentional capacity to formulation while doing the second task, not only easily accessing their linguistic resources but also remembering the TFs which they had noticed in the form instruction. Such repeated practice of accessing their linguistic resources, including the TFs which they had noticed in the instruction, could result in developing their accuracy in producing the TFs.

However, it is conceivable that paying such heavy attention to conceptualization might have had a negative corollary impact on Group 1’s accuracy, which relies more on formulization, resulting in greater development of fluency but no improvement in accuracy. Conversely, it is also conceivable that paying such attention to formulization might have had a negative corollary impact on Group 2’s fluency, which relies more on conceptualization. Moreover, because of easily remembering the content in the first task, Group 2 still paid attention to conceptualization to some extent although less than Group 1. Therefore, Group 2 could still improve proceduralization even though it was less than Group 1 did.

The aggregated results in this study present three implications. One is that form instruction through using not each learner’s own errors but rather others’ errors can be useful. Giving feedback on each learner’s errors during or after the first task is not practical. It is because error production by each learner must be waited for during the task, and feedback is then provided individually, so it takes time. In addition, there is no attention for the other learners while each learner is consulted.
with. However, it is feasible to give form instruction any time to all learners at the same time effectively and efficiently by using example errors which the teacher prepares beforehand. Another implication is that practice by task repetition with form instruction is necessary to facilitate proceduralization of linguistic knowledge and its accuracy to some extent. Van de Gucht et al. (2016) also pointed out the necessity of more repeated practice because the Repetition group, who repeated a task once, did not outperform the No-Repetition group. The last implication is that the timing of form instruction during task repetition influences the development of proceduralization and accuracy in different ways: greater fluency by giving form instruction after the first task, and great fluency and accuracy by providing the instruction before the first task. Therefore, when significant development of learners’ proceduralization only is focused on, form instruction should be given after the first task. Conversely, when the instruction is provided before the first task, both significant development of accuracy and some development of proceduralization can be expected.

5. Conclusion

The two major findings in this study were:

1. Fluency and accuracy of linguistic knowledge will improve through practice doing task repetition with form instruction.
2. The timing of form instruction brings different effects for developing proceduralization and accuracy.

Since “[p]roceduralization is considered a slow process that requires many encounters with the same items” (De Jong & Perfetti, 2011, p. 562), the many encounters with the TFs, articles and verbs, through task repetition and form instruction of the forms in this study, might have developed accurate procedural knowledge sufficiently for it to be used to improve fluency and accuracy. To address Ellis’s suggestion (2016, p. 414) that “[f]rom a psycholinguistic and pedagogical standpoint it is important to investigate … different ways of timing a focus on form,” it can be said that the timing might indeed have produced different effects on the development of fluency and accuracy.

In order to expand the potential of practice using task repetition accompanied by form instruction, it is necessary to conduct further research focusing on the following two points. One is the types of learners, especially those who have less linguistic knowledge. Because of interactions between feedback type, proficiency and TFs (Li, 2014), the effectiveness of the practice may depend on learners’ developmental stage, i.e., how much declarative knowledge of linguistic forms they have. Those who have less linguistic knowledge may heavily allocate their attention to formulation at the second task; however, they have little linguistic knowledge to access, so they will use little English in the task. Hawkes (2012) examined junior high school students aged 13 to 14 who were at “elementary proficiency” (p. 329); however, it is not clear how his treatment, task repetition with
form-focus, can influence their development of not only accuracy but also proceduralization. Otherwise, it is still not safe to conclude that the practice can work effectively for any learner.

The other point to be focused on is the types of TFs: new or known/learned forms. Williams and Evans (1998) focused on participial adjectives (which were actively used but often incorrectly) and passives (which were rarely found in the participants’ speaking). They found that form instruction was more effective for participial adjectives. Thus, it seems that the effectiveness of form instruction relates to how much learners already know the form when given the instruction. The TFs of articles and verbs in this study were items that the participants did not always produce correctly in speaking, and the form instruction after the first task was not effective for developing accuracy. Conversely, Samuda (2001), who used TFs new to the participants, found the effectiveness of form instruction after the first task for accurate use of the TFs. In short, it could be possible that the types of TFs may influence the effect of the timing of form instruction accompanied by task repetition. Therefore, it is necessary to examine if the practice used in this study can work effectively for any form.

Notes

1. The form instruction used here was of the partially unfocused type in the sense that TFs were embedded in every session although other forms were also included without overlapping between sessions. Whereas better effectiveness of focused instruction was found in Sheen, Wright, and Moldawa (2009), the effectiveness of unfocused instruction was demonstrated by Ellis, Sheen, Murakami and Takashima (2008). It seems more difficult for learners to pay attention to forms during unfocused form instruction than focused instruction. In order to verify if learners can derive some benefit through form instruction, the more difficult type of instruction (unfocused) was used.

2. The total error rates of TFs can be low in slow speech. Although the total rates show how accurately learners use TFs, the rates do not indicate how fluently as well as accurately learners perform. Therefore, the error rates of TFs in fluent runs should be calculated. For example, the total error rates of two learners (A, B) are equal (.45), but the mean lengths of fluent runs produced by A (3.00) are longer than by B (2.00). The error rates in fluent runs show that the error rates of A are .150, and the rates in B are .225, showing that A can produce longer runs than B with equal error rates. Another scenario could be that, whereas the total error rates of TFs in A (.45) are higher than of B (.30), the mean lengths of fluent runs produced by A (3.00) are longer than by B (1.50). The error rates of TFs in fluent runs of A (.15) then becomes lower than in B (.20). This shows that A can produce TFs more accurately in fluent runs than B.
Acknowledgement

This research was supported by JSPS KAKENHI, Grant Number 26370695.

References


What Spaced Learning is Effective for Long-Term L2 Vocabulary Retention?

Kohei KANAYAMA
Sapporo Otani High School
Kiwamu KASAHARA
Hokkaido University of Education

Abstract
The present study aims to compare the effects of expanding, equally-spaced, and contracting retrieval practice on L2 vocabulary learning. Group A (n = 34) learned 20 English and Japanese word pairs (e.g., ligament-jintai) under the expanding schedule (Day 1, 1, 8, and 22). In a similar way, Group B (n = 19) and Group C (n = 29) learned the same target words under either the equally-spaced (Day 1, 8, 15, and 22), or the contracting schedule (Day 1, 15, 22, and 22), respectively. Twenty-one days after the last learning session (Day 43), all the groups took a delayed post-test in which they were required to retrieve Japanese meanings for the target words (e.g., ligament:_____?). The results showed that: (a) Group A had a better score than did Groups B and C after the second learning session; (b) Group C outperformed Groups A and B after the fourth learning session. (c) However, no significant difference was observed between the three groups in the delayed test. These findings indicate that the length of the interval between each vocabulary session does not matter for long-term L2 vocabulary retention as long as learners are given three or more spaced learning sessions.

1. Introduction
1.1 The Importance of Linking Between the L2 Form and its L1 Translation Connection
A large number of studies have indicated that L2 learners should encounter new words in meaningful contexts or have a large amount of reading and listening input (Ellis, 2002) because this is almost the same process as when L1 learners acquire new words in their native languages. On the other hand, many scholars and language teachers have made the criticism that paired-associate learning, where learners intentionally focus on making connections between an L2 form and its L1 meaning, is not useful because it is not a natural process of language learning and does not contribute to normal language use (Oxford & Crookall, 1990). That is, making form-and-meaning connections does not include learning the functions of the target words, which is indispensable to real communication. Moreover, this criticism has been strongly supported by the acquisition-learning acquisition.