Does Form Instruction During Task Repetition Facilitate Proceduralization and Accuracy of Linguistic Knowledge?

Masaki DATE
University of Fukui

Abstract

This study focused on how task repetition with form instruction influences proceduralization and accuracy of linguistic knowledge. Two experimental groups repeated narrative tasks in practice sessions. In each session, the Repetition + Noticing group was given time to find and correct errors in the instructional material distributed after the first task, and then they repeated the first task. Conversely, the Repetition group just repeated the first task. The control group was not given any repetition sessions. After the sessions, two posttests were given: posttest 1 with completely new pictures and posttest 2 with the same pictures as the pretest conducted in Session 1. It was found that: (1) at posttest 2, although all groups spoke with fewer errors than at the pretest, the Repetition group performed best; (2) whereas only the experimental groups spoke more fluently than at the pretest, the fluency of the Repetition group was better; and (3) at posttest 1, although all groups spoke with more errors than at the pretest, the control group spoke less fluently both than at the pretest and than did the experimental groups. This study implies (1) the effectiveness of practice using task repetition for facilitating proceduralization and accuracy, and (2) the possible effectiveness of form instruction during task repetition for producing more self-correction in performance as well as its negative corollary impact on them compared to task repetition only.

1. Background

According to ACT-R (Adaptive Control of Theory-Rational) theory (Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004), the learning of ACT-R develops through declarative, procedural and automatic stages. In the second stage, the skill or cognitive act is repeatedly used or performed with explicit knowledge (of what the skill or cognitive act is) acquired in the first stage so that implicit knowledge (of how to use the skill or act) is stored and develops. Procedural knowledge is necessary for language use; therefore, storing and developing procedural knowledge of the skill or cognitive act (implicit knowledge of how to use the skill or cognitive act) stored in the production system is thought to be an indispensable process for language learning. This process is called ‘proceduralization’. VanPatten, Williams and Rott (2004) stated that Form-Meaning Connections (FMCs) are identified as connections between a second language (L2) form
and its L2 meaning, and “[t]he establishment of FMCs is a fundamental aspect of both first and second language acquisition...[that] goes beyond lexical learning.” (p. 4). Tasks can become a way to help learners notice FMCs in individual exemplars and then experience those FMCs repeatedly through more exemplars. In other words, tasks have the potential to be effective in facilitating proceduralization of FMCs. Therefore, attention to tasks has been increasing in recent years, with many studies on the relationship between tasks and language learning having been conducted (e.g., Nunan, 2004; Willis & Willis, 2007).

However, the fact is that tasks are merely workplans for mental activity. By using tasks, teachers can ask, demand or invite learners to do meaningful operations with language and meanwhile pay attention to particular forms, but they cannot force the learners into anything. The gap between ‘what teachers expect learners to do’ and ‘what the learners actually do,’ can be wide. Furthermore, learners have limited capacity in what they can pay attention to and so tend to focus on the content of their speech during task performance. This argument has been supported by the findings of Swain (1985), that learners in a language immersion program had failed to acquire full grammatical and sociolinguistic competence, of Ellis and Sheen (2006), that learners who had received recasts adjudged them a part of negotiation of meaning rather than feedback on erroneous forms, and of Williams (1999), that from 65 hours of taped dialogue in tasks and exercises, around 80% of Language Related Episodes learners produced were lexically oriented. Therefore, it is possible that learners will not promote their language learning much simply by doing a task. The importance of attending to FMCs thus has been proposed by VanPatten et al. (2004). Batstone and Ellis (2009) also proposed ‘The Awareness Principle’, directed at making learners aware of how a particular meaning is encoded by a particular grammatical form. It is then questionable “how learners can be guided to attend to a specific form-meaning mapping ([FMCs]) in the context of communication that simulates real-operating conditions” (Batstone & Ellis, 2009, p. 199). The answer could be pedagogical interventions. Task planning is one such intervention.

The effect of task planning has been investigated in various studies (e.g., Bygate, 2001; Lynch & Maclean, 2001; Yuan & Ellis, 2003). In particular, task repetition has been clearly identified as being effective in improving the language produced in task performance (Skehan, Xiaoyue, Qian, & Wang, 2012). Other studies have also demonstrated the effectiveness of task repetition for producing better performance the second time than the first time, such as in accuracy (Lynch & Maclean, 2001) and fluency (Bygate, 2001). 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. Thus, these results may indicate that students switch attention to language from message content through task repetition. 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” (Ellis, 2005, p. 27). The next key question is whether task repetition is effective for language learning and/or language acquisition.

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 task repetition for proceduralization, i.e., changes in underlying cognitive mechanisms, for fluency development. They gave participants a 4/3/2 task (Nation, 1989) using monologue tasks and measured fluency development by the mean length of pauses, the phonation/time ratio and the mean length of fluent runs. During the training sessions, the Repetition group was given the same topic three times whereas the No-Repetition group was given a new topic three times. At the posttest one week, and the posttest four weeks, after the last session, all participants did a different task. Based on the finding that only the Repetition group maintained increased fluency during the training sessions in the two posttests, De Jong and Perfetti concluded that task repetition in the sessions may cause proceduralization, and result in an increase in fluency and the long-term retention of the increased fluency, as well as transfer of the increased fluency to a new task.

Nevertheless, there are two questions left unanswered by their study. First, as it is not clear if learners can notice target features in input, it is unclear to what extent learners notice anomalies in their production, as Philp and Iwashita (2013) pointed out. The participants might have devoted most of their attention to conceptualization, and spared too little attention to formulation, to notice their anomalies by themselves. Alternatively, even though learners will inevitably experience problems in producing output, it is unclear to what extent they can modify it by themselves, resulting in accuracy not improving without some specific instruction on what learners should attend to (Bygate, 2001). Therefore, it must be examined whether helping learners attend to and notice forms in repeating a task may be effective in facilitating proceduralization. The second question is whether or not proceduralization only facilitates an increase in fluency in a new task. According to ACT-R (Anderson & Lebiere, 1998), declarative knowledge takes the form of chunks in the declarative module, while procedural knowledge consists of production rules in the production system, with each production rule leading to the retrieval of one or, at most, a few declarative chunks. New production rules can subsequently gain strength so as to be able to compete with previously existing rules through repeated practice (Anderson et al., 2004). Therefore, it could be possible that an erroneous chunk is retrieved from the declarative module and then applied to a mistaken production rule, so that, even if the new production rule is erroneous, repeated practice of the rule may facilitate its proceduralization, and fluency may improve. In other words, the positive influence of task repetition on proceduralization must also be proved by fluency in the same task as well as accuracy in the same task and in a new task.

Focusing on these two questions found in De Jong and Perfetti (2011), Date (2013) investigated the effectiveness of task repetition with noticing for facilitating proceduralization and used the same measurements for fluency in De Jong and Perfetti (2011), plus an additional measurement for accuracy: the mean number of target forms used and the ratio of erroneous uses
of target forms. In a practice session, two groups were given a narrative task about a six-picture cartoon strip extracted from Heaton (1975), individually narrated it into IC recorders, and then transcribed what they had said on the back of the sheet with the cartoon on it and corrected their own errors first by themselves and later with teachers’ help using underlining and metalinguistic explanations in Japanese. After correcting the errors, one group repeated the same task, while the other group had a different task from the first task. After four sessions in four weeks, both groups were given two posttests using narrative tasks: the same task as at the pretest and a new task different from the pretest. The results showed that both groups improved fluency and accuracy in the same task, and that the group who repeated the same task in the sessions improved more than the group who did not. In a new task, only the group with task repetition improved accuracy.

However, there are three questions left unanswered by the study of Date (2013). Firstly, as mentioned as one of the limitations, “there was no group which was not given any opportunity for noticing erroneous [target forms]. Therefore, it was not possible to examine the effectiveness of the opportunities for noticing” (p. 122). Secondly, it is not clear if the improvement in fluency and/or accuracy was caused through repeating the same task two times in each session. When both groups transcribed what they had said in the first task on the sheet, they repeated the first task in written mode. In other words, the group who performed a different task after the first task was not necessarily the group without task repetition, and the group who performed the same task in the second task repeated the same task three times in the session. In the previous studies on task repetition, repetition groups orally repeated the same task two or three times with a 4/3/2 task in De Jong and Perfetti (2011). The influence of writing down the first task may be hard to overlook. Lastly, it is questionable whether the same procedure is practical with a large number of students. Time to transcribe and/or correct their own errors may vary depending on how much they talked and how many errors they made in the first task. Therefore, when more students join such practice sessions, not only does it take more time for teachers to check each student’s error correction but also there could be a time lag between students who talked more and/or made many errors and those who talked less and/or made fewer errors.

The present study thus focuses on fluency and accuracy in undertaking the same task and a new task and endeavors to examine the effectiveness of task repetition with noticing on proceduralization of linguistic forms, just as Date (2013) investigated. However, there were two different points from that study by Date: (1) a group with task repetition only and a control group without any practice sessions were set, and (2) the same instructional materials made by the researcher were given to participants to note erroneous forms after the first task in every session.

2. Study

2.1 Hypotheses

The overriding research question was whether learners would facilitate proceduralization
and/or increase in accuracy by task repetition with form instruction given before the second performance more than by task repetition only. The following two hypotheses were thus set:
1. Gains in proceduralization and accuracy from a pretest to a posttest on the same task and a new task would be larger for learners who had training sessions than for those who did not.
2. Gains in proceduralization and accuracy from a pretest to a posttest on the same task and a new task would be larger for learners who both had task repetition and received form instruction in the sessions than for those who had task repetition only in the sessions.

2.2 Procedure
The participants were university students taking English conversation classes once a week. Their TOEFL PBT scores ranged from 440 to 540. They were randomly assigned into three groups. Groups 1 \((n = 14)\) and 2 \((n = 14)\) were given training sessions. Group 3 \((n = 15)\) was a control group that was not given any sessions. Group 1, Repetition with form instruction, did the same task two times and had chances to notice errors before repeating the same task. Group 2, Repetition group, only repeated the same task. Each group followed the procedure shown in Table 1 on different days, respectively. They had three tests and four training sessions in four weeks. A pretest was given at the beginning of the training sessions, and two posttests were given one week after the sessions. In each test and training session, they performed a narrative task with a six-picture cartoon strip extracted from Heaton (1975). Posttest 1 was a new task, different from the pretest (a different picture story), and posttest 2 was the same task as the pretest (the same picture story). This procedure was modified from Bygate (2001). In his study, after doing a task as a pretest in the first week, the experimental groups were given two tasks four times in nine weeks, the same task type but with different topics. In the 10th and final week, all groups performed two tasks: the same task as in the first week plus a new task different from the first week. Referring to Bygate’s procedure seemed to be valid in order to examine if task repetition with/without form instruction in each session will influence participants’ performance in two tasks as posttests after sessions: one with the same picture story as the pretest and one with a new picture story.

Each session and posttest was given during a class, without announcing to the participants

| Table 1 |
| Schedule of Tests and Training Sessions |
| training sessions | 1 | 2 | 3 | 4 | posttests |
| pretest | a/[instruction]/a | b/[instruction]/b | c/[instruction]/c | d/[instruction]/d | e/a |
| Group 1 | a/a | b/b | c/c | d/d | e/a |
| Group 2 | a/a | b/b | c/c | d/d | e/a |
| Group 3 | a | b/b | c/c | d/d | e/a |

Note. The letters refer to the picture stories given.
beforehand that they would have four sessions and two posttests. In the sessions, each participant 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. Soon after telling the story, a new sheet with the same pictures as for the first performance was given again to Group 2. The group simply told one more story again into the IC recorder for up to 90 seconds (i.e., no opportunity for form instruction). On the other hand, after telling the story the first time, each participant in Group 1 was given both a new sheet with a passage on it and a tablet PC. The passages given in each week’s sessions were different but contained 276.25 words, 22 sentences, 27.5 T-units and 18.25 errors on average, 11 of which were errors in either verbs or articles. All errors had been 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, shown below, came up and showed the correct form along with metalinguistic explanation in Japanese. While they were checking, the researcher walked around the room and verified whether each participant now understood all the errors. When all the participants had corrected the errors and checked them on the PC, the 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 a story again into the IC recorder for up to 90 seconds (i.e., after form instruction). After telling the second story, each learner’s two sheets, recorder and tablet PC were collected.

One week after Session 4, two posttests were given: posttest 1, with completely new pictures, and posttest 2, with the same pictures as in the pretest. Right after the posttests, 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.

When I study or do ① __ homework, I always listen to jazz music, because jazz music makes me ...

① · 定冠詞 the を使わない（不特定な homework であるため）
· a は不要（homework は数えられない不可算名詞であるため）

2.3 Analysis
All data in IC recorders were analyzed with PRAAT 5.3.09 (Boersma & Weenink, 2012). After the researcher transcribed all information, the beginning and end of each speech segment was then determined by using the PRAAT function ‘To textgrid (silences).’ All pause boundaries
were checked and adjusted by the researcher as necessary, by listening to the recording and visually inspecting the spectrogram and wave-form. Nonverbal fillers, such as “uh,” and “mmmm”, were not transcribed. Any filler or silence, which was 0.20s or longer, was treated as a pause. This cutoff point was the same as that used by De Jong and Perfetti (2011). In each speech of each participant, the upper limit for pauses was set to 2.5 standard deviations above the mean. A pause longer than the upper limit was replaced with the mean pause length plus 2.5 standard deviations, as in De Jong and Perfetti. Syllables were counted only from words uttered in their entirety, excluding false starts. Words repeated were also counted. For measuring proceduralization, based on De Jong and Perfetti (2011) and Date (2013), two measures for fluency were used: 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.

Referring to Date (2013), verbs and articles were set as the target forms: verb errors included errors in tense, word choice, and subject-verb agreement, and article errors encompassed all types of article uses. In the passages used for form instruction in the sessions, erroneous target forms were also intentionally embedded. All erroneous target forms produced by each participant were confirmed by the researcher with a native speaker of English. Either when an erroneous form was replaced by a correct form soon or when a correct form was replaced by an erroneous form soon, the form was judged as an error. The number of erroneous target forms, and the total number of the target forms used (i.e., the combined number of correct forms and erroneous forms) were then counted. For measuring accuracy, different from Date (2013) who focused on the total error rates (dividing the number of errors of target forms by the number of forms used and multiplying by 100), the extent to which participants performed correctly while speaking fluently was targeted. Thus, the error rates of the target forms in fluent runs were calculated through dividing the total error rates of target forms by the mean length of fluent runs.

3. Results

Table 2 shows the mean scores and standard deviations at each test. There was no significant group difference found at the pretest (mean length of pauses: $F(2, 42) = .326, p = .724, \eta^2 = .02$; mean length of fluent runs: $F(2, 42) = 1.225, p = .304, \eta^2 = .06$; error rate of target forms in fluent runs: $H(2) = 1.420, p = .492$), indicating their parity in fluency and accuracy at first.

3.1 Proceduralization

A two-way repeated measures ANOVA on the mean length of pauses revealed a medium sized effect in the interaction between test and group ($F(4, 80) = 1.539, p = .199, \eta_p^2 = .07$). Looking at simple main effects by test, significant differences were found in Groups 1 ($\chi^2(2) = 18.429, p = .000$), 2 ($\chi^2(2) = 9.927, p = .007$) and 3 ($F(2, 28) = 10.797, p = .000, \eta^2 = .44$). Multiple comparisons by $t$ test and Wilcoxon signed-rank test with Bonferroni correction ($\alpha = \ldots$)
Table 2

Means and Standard Deviations of the Measures of Proceduralization and Accuracy in the Tests

<table>
<thead>
<tr>
<th></th>
<th>pretest</th>
<th>posttest 1 (new task)</th>
<th>posttest 2 (same task)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proceduralization</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mean length of pauses</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>1.28 (0.29)</td>
<td>1.23 (0.23)</td>
<td>1.00 (0.22)</td>
</tr>
<tr>
<td>Group 2</td>
<td>1.17 (0.49)</td>
<td>1.30 (0.93)</td>
<td>.94 (0.33)</td>
</tr>
<tr>
<td>Group 3</td>
<td>1.24 (0.20)</td>
<td>1.51 (0.41)</td>
<td>1.16 (0.24)</td>
</tr>
<tr>
<td>mean length of fluent runs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>2.36 (0.40)</td>
<td>2.05 (0.31)</td>
<td>2.65 (0.57)</td>
</tr>
<tr>
<td>Group 2</td>
<td>2.62 (0.54)</td>
<td>2.25 (0.20)</td>
<td>3.18 (0.54)</td>
</tr>
<tr>
<td>Group 3</td>
<td>2.44 (0.39)</td>
<td>2.15 (0.40)</td>
<td>2.63 (0.47)</td>
</tr>
<tr>
<td><strong>Accuracy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>error rate of target forms in fluent runs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Group 1</td>
<td>18.64 (6.08)</td>
<td>24.86 (5.68)</td>
<td>16.50 (5.27)</td>
</tr>
<tr>
<td>Group 2</td>
<td>16.64 (4.97)</td>
<td>22.36 (8.90)</td>
<td>18.60 (5.93)</td>
</tr>
<tr>
<td>Group 3</td>
<td>19.67 (7.36)</td>
<td>25.00 (7.90)</td>
<td>18.20 (6.01)</td>
</tr>
</tbody>
</table>

*Note.* Values enclosed in parentheses represent standard deviations.  
\(^a n = 14. ^b n = 14. ^c n = 15.*

.016) showed that: Group 1 produced significantly shorter pauses at posttest 2 than at the pretest ($z = -3.045$, $p = .002$, $r = -.58$) and similar lengths of pauses at posttest 1 to at the pretest; Group 2 produced significantly shorter pauses at posttest 2 than the pretest ($z = -2.972$, $p = .003$, $r = -.56$) and similar lengths of pauses at posttest 1 to the pretest; and different from Groups 1 and 2 was that Group 3 produced significantly longer pauses at posttest 1 than the pretest ($t(14) = -2.998$, $p = .010$, $\Delta = 1.30$) and the pauses at posttest 2 which were not shorter than at the pretest. Next, looking at simple main effects by group, the group difference was found only at posttest 2 with a medium sized effect ($F(2, 42) = 2.511$, $p = .094$, $\eta^2 = .11$). Multiple comparisons by t test with Bonferroni correction indicated that Group 3 made longer pauses than Groups 1 ($t(27) = 1.804$, $p = .082$, $r = .33$) and 2 ($t(27) = -2.018$, $p = .054$, $r = .36$). There was no significant difference between Groups 1 and 2 with a small sized effect ($t(26) = .561$, $p = .579$, $r = .11$). In short, at posttest 2, Groups 1 and 2 made shorter pauses than Group 3, and they made shorter pauses than at the pretest whereas Group 3 did not. At posttest 1, all groups made similar lengths of pauses, and Group 3 made longer pauses than at the pretest whereas Groups 1 and 2 did not.

A two-way repeated measures ANOVA on the length of fluent runs documented a medium sized effect in the interaction between group and test ($F(4, 80) = 2.338$, $p = .062$, $\eta^2 = .11$). Looking at simple main effects by test, significant differences were found in Groups 1 ($F(2, 26) = 13.464$, $p = .000$, $\eta^2 = .51$), 2 ($\chi^2(2) = 21.000$, $p = .000$) and 3 ($F(2, 28) = 12.598$, $p = .000$, $\eta^2 = .47$). Multiple comparisons by t test and Wilcoxon signed-rank test with Bonferroni correction
then revealed that: Group 1 produced significantly shorter runs at posttest 1 than at the pretest ($t(13) = 4.125, p = .001, \Delta = -0.76$) and longer runs at posttest 2 than at the pretest ($t(13) = 2.354, p = .035, \Delta = 0.72$); Group 2 produced shorter runs at posttest 1 than the pretest ($z = 2.354, p = .019, r = -.45$) and significantly longer runs at posttest 2 than the pretest ($z = -3.108, p = .002, r = -.59$); and Group 3 produced significantly shorter runs at posttest 1 than the pretest ($t(14) = 3.612, p = .003, \Delta = -0.76$) whereas no significant difference was found between the pretest and posttest 2 with a small sized effect ($t(14) = -1.492, p = .158, \Delta = 0.47$). Looking at simple main effects by group, a significant difference between the groups was found at the posttest 2 ($F(2, 42) = 4.915, p = .012, \eta^2 = .20$). Multiple comparisons by $t$ test with Bonferroni correction showed that, whereas Group 2 produced longer runs than Groups 1 ($t(26) = 2.522, p = .018, r = .44$) and 3 ($t(27) = 2.899, p = .007, r = .49$), there was no significant difference between Groups 1 and 3 with no effect ($t(27) = -0.96, p = .924, r = .02$). In summary, all groups similarly produced shorter fluent runs at posttest 1 than the pretest. Groups 1 and 2 produced longer runs at posttest 2 than the pretest, but Group 3 did not. Moreover, Group 2 produced longer runs than Group 1.

### 3.2 Accuracy

A two-way repeated measures ANOVA on the error rates of the target forms in fluent runs showed significant main effect in the test ($F(2, 80) = 33.545, p = .000, \eta_p^2 = .46$). Multiple comparisons by Wilcoxon signed-rank test with Bonferroni correction ($\alpha = .016$) indicated that the error rate in fluent runs was significantly higher at posttest 1 than the pretest ($z = -4.209, p = .000, r = -.27$). Furthermore, the rate at posttest 2 was significantly lower than the pretest ($z = -2.504, p = .012, r = -.27$). Looking at the error rate at posttest 2 then revealed a significant group difference ($F(2, 42) = 5.349, p = .009, \eta^2 = .21$). Multiple comparisons by $t$ test with Bonferroni correction indicated that the ratios in Group 2 were lower than Groups 1 ($t(26) = -2.324, p = .028, r = .42$) and 3 ($t(27) = -3.293, p = .003, r = .54$). In other words, although all groups performed equally at posttest 1 with higher erroneous rates in fluent runs than at the pretest, they similarly performed at posttest 2 with lower rates than at the pretest. In particular, Group 2 produced with lower erroneous ratios than Groups 1 and 3.

### 3.3 Self-modification

Table 3 shows the numbers of participants who correctly modified erroneous target forms by themselves during performance at each test, and of those who correctly modified at both posttests although they had not modified anything at the pretest. The following two points can be seen: (1) whereas the number of participants who modified erroneous target forms at each posttest decreased in Group 3, the number increased in Group 1 only; and (2) the number of participants in Group 1 who correctly modified at both posttests, although they had not modified anything at the pretest, was the largest; however, the number was zero in Group 3.
Table 3

<table>
<thead>
<tr>
<th></th>
<th>pretest</th>
<th>posttest 1</th>
<th>posttest 2</th>
<th>both posttests</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1</td>
<td>3</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Group 2</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Group 3</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>0</td>
</tr>
</tbody>
</table>

4. Discussion

Table 4 is the summary of the data analysis of proceduralization and accuracy in the tests. Here the two hypotheses are examined. Hypothesis 1, that gains in proceduralization and accuracy from a pretest to a posttest on the same task and a new task would be larger for participants who had training sessions than for those who did not, was partially supported. At the pretest, all groups were similar in fluency and accuracy. In posttest 2, the experimental groups needed shorter pauses than the control group in order to produce similar lengths of fluent runs to, or longer runs than, the control group. Furthermore, the experimental groups produced longer fluent runs than at the pretest as well as having shorter pauses than at the pretest. However, the control group did not change in either the lengths of pauses or fluent runs from at the pretest. Such improved performance of, and a gain in, the experimental groups may indicate the effectiveness of practice with task repetition for facilitating proceduralization compared to no practice. As for accuracy, the rates of erroneous target forms in fluent runs produced by all groups were lower than at the pretest. Group 2 produced a lower rate than the control group. In other words, whereas all groups similarly performed more accurately than at the pretest, only one of the experimental groups performed more accurately than the control group. Therefore, it can be said that, when doing a posttest with the same task, (1) gains in proceduralization from a pretest were seen only in the participants who had training sessions, and (2) although all groups gained in accuracy from the pretest, larger gains

Table 4

Summary of the Data Analysis of Proceduralization and Accuracy in the Tests

<table>
<thead>
<tr>
<th>Measure</th>
<th>Between comparison</th>
<th>Within comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proceduralization</td>
<td></td>
<td></td>
</tr>
<tr>
<td>posttest 1 length of pauses</td>
<td>Group 1 = 2 = 3</td>
<td>(Group 1, 2) = pretest; (3) &gt; pretest</td>
</tr>
<tr>
<td></td>
<td>length of fluent runs</td>
<td>Group 1 = 2 = 3</td>
</tr>
<tr>
<td>posttest 2 length of pauses</td>
<td>Group 1 = 2 &lt; 3</td>
<td>(Group 1, 2) &lt; pretest; (3) = pretest</td>
</tr>
<tr>
<td></td>
<td>length of fluent runs</td>
<td>Group 2 &gt; 1 = 3</td>
</tr>
<tr>
<td>Accuracy</td>
<td></td>
<td></td>
</tr>
<tr>
<td>posttest 1 error rate of target forms</td>
<td>Group 1 = 2 = 3</td>
<td>(Group 1, 2, 3) &gt; pretest</td>
</tr>
<tr>
<td>in fluent runs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>posttest 2 error rate of target forms</td>
<td>Group 2 &lt; 1 = 3</td>
<td>(Group 1, 2, 3) &lt; pretest</td>
</tr>
</tbody>
</table>
were seen in the participants who had the sessions without form instruction. In posttest 1, on the other hand, all groups produced similar lengths of fluent runs to each other, and the lengths produced by each group were all lower than at the pretest. Furthermore, all groups made similar lengths of pauses to each other. However, the experimental groups did not change from the pause lengths at the pretest whereas the control group made longer pauses. A gain in fluency from the pretest was not seen in either the experimental groups or the control group; however, the less fluent performance by the control group compared to at the pretest may indicate the effectiveness of practice with task repetition for facilitating proceduralization compared to no practice. As for accuracy, on the other hand, the error rates of target forms in fluent runs produced by each group were all higher than at the pretest. In addition, all groups produced a similar rate of erroneous target forms to each other. In other words, all groups similarly performed less accurately than at the pretest. It can then be said that, when doing a posttest with a new task, (1) although all groups lost in proceduralization from the pretest, more loss was seen in the participants who did not have any training sessions, and (2) all groups similarly lost in accuracy from the pretest.

In Bygate (2001), a narrative group had three treatment sessions in total and two different narrative tasks in each session. At the posttest nine weeks after the pretest, when the narrative group and the control group performed the same narrative task as in the pretest, neither of them improved their fluency and accuracy from the pretest. The control group in the present study did not improve their fluency either, just as in Bygate’s study. However, Groups 1 and 2 in this study did improve their fluency, different from the narrative group in Bygate’s study, although both studies gave the experimental groups three sessions and two narrative tasks in each session. Furthermore, Group 2 gained more accuracy than the control group did. A possible reason for such differences between the two studies is that Groups 1 and 2 in this study had task repetition in the session while the narrative group in Bygate did not. Therefore, the effectiveness of task repetition for facilitating proceduralization proposed by De Jong and Perfetti (2011) and Date (2013) was confirmed here. Repeating the same task may assist subsequent language performance if “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” (Bygate, 2001, p. 29). Such a theory about the mechanism of task repetition may constitute the reason for better task performance the second time. It should not be overlooked that the mechanism implies the importance of making learners reuse what is kept in their memory after the previous performance for proceduralization and/or (implicit) learning. ACT-R also insists that, when declarative knowledge stored in the declarative module is retrieved repeatedly, it helps procedural knowledge in the production system to develop (Anderson et al., 2004).

Next, Hypothesis 2, that gains in proceduralization and accuracy from a pretest to a posttest on the same task and a new task would be larger for participants who both had task repetition and received form instruction in the sessions than for those who had task repetition only in the sessions, was not supported. Analyzing posttest 1, no significant difference between Groups 1 and 2 was
found in the lengths of pauses and fluent runs, or in the error rates of target forms in fluent runs. In other words, both groups performed with similar proceduralization and accuracy to each other. On the other hand, in doing posttest 2, Group 2 produced longer fluent runs than Group 1, despite having similar lengths of pauses to Group 1’s. Furthermore, Group 2 produced a lower rate of erroneous target forms in fluent runs than Group 1. This means that Group 2 showed more gains in proceduralization and accuracy than Group 1. Therefore, it can be said that form instruction during task repetition did not facilitate either greater proceduralization or more accuracy than task repetition only; rather, form instruction had a negative impact on proceduralization and accuracy and disturbed fluent and accurate speaking.

Group 1 was given form instruction in each session. In the instruction, they first looked through a passage on a sheet which highlighted errors through underlining or a blank space, prepared by the researcher. After the participants corrected the errors on the sheet by themselves, they compared their corrections to the same passage explicitly corrected using metalinguistic explanation in Japanese on the PC. Group 1 thus had opportunities to notice an erroneous target form and correctly connected the form with its meaning/function. Therefore, their consciousness of the connection might have been raised before the second performance. This raised consciousness might then have influenced their repeatedly retrieving declarative knowledge of the connection in the second performance. Consequently, it had been expected that when they undertook the same task at the posttest, not only the process of conceptualization (message generation) but also that of formulation (grammatical encoding) might have been boosted, engendering more fluency and accuracy than Group 2, who would have been required to pay more attention to, and be burdened by, formulation. Such a possible benefit of giving form instruction through written and metalinguistic feedback when repeating the same task had been also expected from the positive results found in some previous studies. Sheen (2010) argued that written corrective feedback more effectively prompts learners’ noticing of errors and uptake than oral feedback inasmuch as the former feedback places a lesser burden on their cognitive abilities, especially the short-term memory. Li (2014) focused on oral feedback using metalinguistic information and found that metalinguistic correction (provision of the correct form and explicit rule explanation in the learners’ first language) produced better effects than recasts for low-level learners. Shintani, Ellis, and Suzuki (2014) proved that written corrective feedback (WCF) plus revision was more effective than WCF alone for producing greater accuracy in new writing.

Then why did Group 1 not perform as well as Group 2? The possible answer is that they did not focus on their own errors in the form instruction. According to Shintani et al. (2014), the reason for the better performance by WCF plus revision than WCF alone is that, “WCF in conjunction with the opportunity to revise results in pushed output …. The need to revise involves explicit attention to the initial error and its correction, which may promote storage of the target features in memory” (p. 108). In other words, it is important to make learners pay explicit attention to the initial error and its correction in the second performance so that this attention to the
error and its correction may promote storage of the target features in their memories. This idea of ‘feedback plus revision’ is also applicable to ‘feedback plus task repetition’. When learners receive feedback about erroneous forms during task repetition, the feedback should be given on the initial errors made in the first task performance. Consequently, when they repeat the same task at the second performance, they will pay explicit attention to the initial error and its correction, and the second performance becomes a revision of the first performance. In the present study, Group 1 was asked to repeat the first task after feedback on erroneous forms was given, i.e., direct corrective feedback with metalinguistic clues in Japanese. However, these instructional materials presented errors which were not actually produced by individual participants in Group 1. Two unexpected scenarios might then have conceivably played out. First, when learners try to notice how a form in the material is erroneous, they might not know how the form is used in that particular context. Thus, they might have some confusion about, or even difficulty in, making the Form-Meaning Connection (FMC) of the form and noticing the erroneous feature through metalinguistic explanation in Japanese. Consequently, they might not process the feedback deeply enough to make their declarative knowledge of the FMC be stored in the declarative module, and they then cannot use the knowledge to facilitate proceduralization or accuracy. Second, the repeated encounters with the same types of error forms of other people in the materials might make the participants too focused on those forms rather than their own FMCs. Consequently, they might pay more attention to formulation and less attention to conceptualization in the second performance and/or in the posttests. From Table 3, it can be seen that form instruction during task repetition might have raised learners’ consciousness of language forms and led to correct modifications while speaking after the instruction. However, it is also conceivable that paying such heavy attention to formulation might have had a negative corollary impact on Group 1’s fluency, which relies more on conceptualization. The negative influence of using erroneous forms made by others may help explain the different, more positive results of both the Repetition + Noticing group in Date (2013), who correctly modified their own errors after the first task before repeating the same task again, and the Direct Corrective Feedback + Revision group in Shintani et al. (2014), who received direct corrections to their own errors and then revised the first writing.

There are two other possible reasons that might have caused the poor performance of Group 1. One is that erroneous forms presented in the instructional materials could have been somewhat obscure, and thus insufficiently focused on, by the participants. This study intentionally embedded distracters in the materials. Consequently, the participants might not have focused on the target forms so much even though they were given in every session. However, in the study of Mochizuki and Ortega (2008) which proved the effectiveness of strategic planning for accuracy, the participants received teacher’s guidance to pay attention to relative clauses only. Furthermore, Sheen, Wright, and Moldawo (2009) reported that focused WCF was more effective in grammatical accuracy gain in L2 writing than unfocused WCF. The other possible reason for the poor performance of Group 1 is individual learner factors. For example, Li (2014) pointed out the
interactions between feedback type, proficiency and target structures. Therefore, it is still not safe
to conclude that it was the form instruction itself that was solely responsible for undermining the
effectiveness of task repetition in Group 1.

The aggregated results of these two hypotheses present two implications. One is that
practice by task repetition is necessary to facilitate proceduralization of linguistic knowledge
greatly, and accuracy of it in fluent runs to some extent. According to Ellis (2005), evidence for
some change in the learner’s linguistic knowledge representation can be found in: (1) the learner’s
use of some previously unused linguistic forms; (2) an increase in the accuracy of some linguistic
forms that the learner can already use; (3) the use of some previously used linguistic forms to
perform some new linguistic functions or in new linguistic contexts; and, (4) an increase in
fluency. Each item shows a change in underlying cognitive mechanisms. Such changes in
underlying cognitive mechanisms represent proceduralization of linguistic knowledge (De Jong &
Perfetti, 2011). The gains in proceduralization in the same context and the new context, which is
related to item (4), were found in the experimental groups only in the present study. A greater gain
in accuracy in the same context, which is related to items (1) and (2), was found in the group by
task repetition only than in the control group. The other implication is that form instruction during
task repetition promotes self-correction in speaking but does not necessarily facilitate
proceduralization and accuracy in fluent runs, when the instruction presents erroneous forms
which are not learners’ own errors for the purpose of practicality in the classroom.

5. Conclusion

The three major findings in this study were:
1. Fluency and accuracy in the same task will improve from the present level through practice
doing task repetition either with or without form instruction. However, practice doing task
repetition only will engender greater fluency and accuracy than providing form instruction
during task repetition.
2. Fluency in a new task will be positively influenced by practice doing task repetition either with
or without form instruction. However, accuracy in a new task will not improve from the present
level through task repetition practice either with or without form instruction.
3. Form instruction during task repetition will raise learner’s consciousness of language forms and
lead to correct modifications while speaking. However, the consciousness of forms will also
have a negative impact, leading to shorter fluent runs and less attention to target forms among
other forms.

The many encounters with the target forms, articles and verbs, through task repetition in this study,
might have developed accurate procedural knowledge enough for it to be used in the same task to
improve fluency and accuracy and in the new task to positively influence fluency. In order that
learners develop procedural knowledge little by little and enact the slow process of
proceduralization, it is useful to have learners repeat tasks for frequent encounters with the same forms. De Jong and Perfetti (2011) also mentioned that, “Proceduralization is considered a slow process that requires many encounters with the same items” (p. 562).

However, there are at least two limitations in this study. The first limitation is that, as pointed out in the Discussion section, it is not clear if the form instruction used here was appropriate to measure its effectiveness for proceduralization and accuracy in fluent runs. The second limitation is that it is not clear if the posttests were accurate in to measure the improvement of fluency and accuracy. 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. Through a within-group comparison between two narrative tasks, they found that the tasks with a clear order of narrative events were performed significantly more accurately than the tasks with loose structures probably because the tight ordering of events releases attentional resources that would otherwise be spent on finding connections between pictures. It is possible that, in this study, the pictures used at posttest 1 were of a loose type while those at posttest 2 were of a tight type; consequently, Groups 1 and 2 performed less accurately at posttest 1 than at the pretest, and the control group performed more accurately at posttest 2 than at the pretest.

Considering such limitations, further research is necessary to expand the potential of task repetition with form instruction to facilitate the proceduralization and accuracy of linguistic knowledge. Thus, it is necessary to examine the effects of various types of form instruction, including what the instructional materials cover: for example, learner’s own/others’ errors and focused/unfocused errors. It is also necessary to use various types of pictures: for example, tighter pictures at posttest 1 and less tight pictures at posttest 2 than those used in the present study.

References


