The Effect of Task Repetition With Noticing on Proceduralization of Linguistic Knowledge

Masaki DATE
University of Fukui

Abstract

This study focused on how task repetition influences proceduralization of linguistic knowledge. The participants performed two narrative tasks once a week. All participants were given the same picture stories for the first task and had time to notice errors after the first task. At the second task, the Repetition group \((N = 14)\) was given the same picture story as the first task whereas the No Repetition group \((N = 14)\) was given a different picture story from the first task. One week after Session 4, two posttests were given: posttest 1 with completely new pictures and posttest 2 with the same pictures as the pretest conducted in Session 1. At the posttest 1, the fluency of both groups was similar; however, the No Repetition group spoke less fluently than at the pretest whereas the Repetition group retained their fluency. Furthermore, the Repetition group spoke more accurately than at the pretest whereas the No Repetition group showed no difference. The Repetition group also spoke more accurately than the No Repetition group. At the posttest 2, both groups spoke more fluently and accurately than at the pretest; however, the Repetition group spoke more fluently and accurately than the No Repetition group. This study implies the possible greater effectiveness of task repetition with noticing for facilitating proceduralization rather than no repetition with noticing.

1. Background

The use of task has been paid attention to in the area of teaching English, as seen from teaching approaches such as Task-based Language Teaching (Ellis, 2003) and Task-based Teaching (Willis & Willis, 2007). Such attention to tasks has been increasing in recent years. Many studies on the relationship between tasks and language learning have been conducted (e.g., Bygate, Skehan, & Swain, 2001; Ellis, 2003; Willis & Willis, 2007).

There are two different types of linguistic knowledge: declarative knowledge and procedural knowledge. Procedural knowledge is the type of performative knowledge which is the direct source for application in the use of language, especially oral spontaneous use. ACT-R (Adaptive Control of Theory-Rational) theory (Anderson & Lebiere, 1998) claims that proceduralization is the process for storing and developing procedural knowledge of the skill or
cognitive act stored in the production module, while procedural knowledge is composed of production rules and formed through the process of production compilation. For creating a production rule, the following process is required: first learners find and experience the connection between a linguistic form and its meaning/function in an exemplar as an instance of the connection; then they repeatedly experience the connection through more exemplars; and, finally, they generalize the connection as an abstract rule and learn it as such. Tasks can become the way to help learners experience the connections between linguistic forms and their meanings/functions in individual exemplars and then experience those connections repeatedly through more exemplars. Therefore, tasks have the potential to be effective in facilitating proceduralization of L2 knowledge (Johnson, 1996).

However, learners have limited capacity in what they pay attention to and tend to focus on the content of their speech during task performance. Even when given a specific instruction to pay attention to forms before speaking, they tend to attend more to meaning (Crookes, 1989; Foster & Skehan, 1996). Learners simply doing a task will probably not promote their language learning much. Some approach then must be given to help them attend to the connection between forms and their meanings/functions (Skehan, Xiaoyue, Qian, & Wang, 2012). Focus-on-form (Doughty & Williams, 1998) is one approach for facilitating learners’ attention to the connection between a form and its meaning/function. Several techniques have been tried, and their effectiveness examined: for example, recasts (Leeman, 2003) and dictogloss (Qin, 2008).

Task planning is another technique for focus-on-form, as Ellis (2005) elucidated: “Providing learners with the opportunity to plan a task performance constitutes a means of achieving a focus-on-form pedagogically” (p. 10). The effect of task planning has been investigated in various studies (e.g., Ahmadian & Tavakoli, 2011; Bygate, 2001; Crookes, 1989; Foster & Skehan, 1996; Lynch & Maclean, 2001; Mehnert, 1998; Yuan & Ellis, 2003). Ellis (2005) summarized the previous studies on three types of planning: strategic planning, giving learners time for “preparing to perform the task by considering the content they will need to encode and how to express this content” (p. 3); task repetition, “repetition with the first performance of the task viewed as a preparation for a subsequent performance” (p. 3); and, on-line planning, giving “the time made available to the learners for the on-line planning of what to say/write in a task performance” (p. 4). Furthermore, Skehan et al. (2012) added two further variables, namely, familiarity with the content domain involved, “engaging with material that has been encountered before, and that may be known well” (p. 177), and post-task influence, “anticipation of a post-task activity to follow a task” (p. 179), to the three types of planning. They examined which variable is more influential on the language produced in task performance, in terms of accuracy, fluency, and complexity. Through these studies on task planning, it has been clarified that task repetition is effective in improving the language produced in task performance. Skehan et al. (2012) also pointed out its specific effectiveness: (1) task repetition can be stronger in its effects than strategic planning or on-line planning; and, (2) there are huge effect sizes in conceptualization, developing the ideas to
be expressed, and formulation, clothing the ideas in language elements. However, what the results in the previous studies suggested is only either that language in task repetition becomes better than language in the first task performance or that language in task repetition becomes better than language in other planning conditions or other variables. In other words, it is not clear if task repetition can facilitate proceduralization of linguistic knowledge and, consequently, improve fluency and accuracy in a new task as well as the same task.

De Jong and Perfetti (2011) focused on the effectiveness of task repetition on proceduralization. They measured fluency development by using three measures of fluency: the mean length of pauses, the phonation/time ratio (the percentage of time spent speaking as a proportion of the total time taken to produce the speech sample), and the mean length of fluent runs. These measures are good predictors of fluency, and, when used in combination, they can also be indicators of proceduralization (Towell, Hawkins, & Bazergui, 1996). De Jong and Perfetti (2011) used a 4/3/2 task (Nation, 1989), in which students did a speaking task for four minutes and then retold it twice, as close to verbatim as possible, in three and two minutes. Twenty-four students enrolled in speaking courses at high-intermediate level in an institute for ESL at a university performed three 4/3/2 tasks and were given three speaking tests. Monologue tasks were used in the sessions and tests, and the students were given a topic (e.g., How do you feel about pets?) and spoke about it. For each 4/3/2 task, the Repetition group was given the same topic three times whereas the No Repetition group was given a new topic three times. At the pretest before the training sessions, the immediate posttest one week after the last session, and the delayed posttest four weeks after the last session, all students did a different task for two minutes. It was then found that although both groups increased fluency during a training session, only the Repetition group maintained this increase in the two posttests. De Jong and Perfetti concluded that task repetition 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.

However, there are two questions left unanswered by the study of De Jong and Perfetti (2011). First, it is unclear if merely repeating a task facilitates proceduralization. ACT-R claims that, for the facilitation of proceduralization, the skill or cognitive act must be repeatedly used or performed with declarative knowledge of the skill or act stored in the declarative module. When learners repeat a task, they are likely to switch their attention to the selection and monitoring of appropriate language (Bygate, 1999). They also have the benefit of having used certain grammatical constructions, which can facilitate retrieval of the constructions through syntactic priming (Kim & McDonough, 2008). Therefore, task repetition seems to facilitate the skill or cognitive act to be repeatedly used or performed with declarative knowledge. On the other hand, it was found that accuracy did not improve without specific instruction on what learners should attend to (Bygate, 2001); however, with specific, discrete instructions about what forms learners should attend to before undertaking a task, they could be more mindful of those forms during the task performance (Mehnert, 1998; Yuan & Ellis, 2003). In other words, for the repeated use of the
skill or repeated performance of the cognitive act with declarative knowledge in order to facilitate proceduralization, learners must pay explicit attention to declarative knowledge stored in their declarative module before task performance and then attend to the knowledge again and use it during the ensuing speaking task. Consequently, there would be a greater likelihood of proceduralization occurring. Therefore, it must be examined whether helping learners attend to and notice forms in repeating a task may be effective on facilitating proceduralization.

The second question arises as it is not clear if proceduralization only facilitates an increase in fluency of a new task. ACT-R claims that declarative knowledge takes the form of chunks in the declarative module, while procedural knowledge consists of production rules in the production module, with each production rule leading to the retrieval of one or, at most, a few declarative chunks. However, it could be possible that an erroneous chunk is retrieved from the declarative module and then used in a mistaken production rule. New production rules can subsequently gain strength so as to be able to compete with previously existing rules through repeated practice (Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004). Therefore, it could be possible that, even if a new production rule is erroneous, repeated practice of the rule may facilitate its proceduralization, and fluency may improve. The process of acquisition can affect some changes in the learner’s L2 knowledge representation regarding fluency in the same task and/or in a new task, and/or accuracy in the same task and/or in a new task (Ellis, 2005). Proceduralization also shows changes in underlying cognitive mechanisms (De Jong & Perfetti, 2011). In other words, the positive influence of task repetition on proceduralization must be proved by not only fluency in a new task but also fluency in the same task and accuracy in the same task and in a new task.

The present study 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.

2. Study

2.1 Hypotheses

The overall research question was whether learners would increase in fluency and accuracy through facilitating proceduralization by task repetition with noticing their errors before the second performance more than no repetition. The two hypotheses were thus set:

1. Gains in fluency from a pretest to a posttest on the same task and a new task would be larger for participants who repeated than for those who did not.
2. Gains in accuracy from a pretest to a posttest on the same task and a new task would be larger for participants who repeated than for those who did not.

2.2 Procedure

The participants were university students taking English conversation classes once a week.
Their TOEFL PBT scores ranged from 450 to 540. They were randomly assigned into two groups (Repetition group: \( N = 14 \), No Repetition group: \( N = 14 \)).\(^1\) Table 1 shows the procedure of this study. Each session and posttests were given during a class, without announcing to the participants beforehand that they would have four sessions in total and posttests after that. In a session, each participant was first given an IC recorder and a sheet with a six-strip cartoon extracted from Heaton (1975). Once they had a quick look at the cartoon and made sure they understood the meaning of each strip, the researcher announced that they were to perform a narrative task, which required them to make a story using the pictures, in order, by talking into the IC recorder for 90 seconds. They could stop recording by themselves if they finished talking before the 90 seconds had elapsed. After recording, the participants listened back to their story and transcribed it on the back of the sheet with the cartoon on it. They were told to write down everything without modifying the content. This transcribing activity was set in order to give the participants three opportunities for noticing. While or after transcribing, they corrected their own errors by using dictionaries and color pens. This was the first opportunity for noticing. When self-correction was completed, the transcriptions were submitted. Immediately, a native speaker checked them with the researcher. While they were being checked, the participants were allowed to do anything except discuss the story and the errors. All errors in the transcriptions (except spelling) were underlined though not explicitly corrected and were returned to each participant. Participants were then given time to correct the underlined errors by themselves. This was the second opportunity for noticing. After correcting, each participant submitted the transcription. When correction was still needed, explicit and direct feedback using metalinguistic explanation in Japanese was given by the researcher as the third opportunity for noticing. When all the participants had corrected and understood the errors, the sheet was collected. Then a new sheet with a six-strip cartoon was given again: the same pictures as the first performance to the Repetition group, and different pictures than the first performance to the No Repetition group. After both groups made sure they understood each strip, they talked into the IC recorder, transcribed it on the back of the sheet, and made self-corrections. The sheet and the recorder were then collected. One week after Session 4, posttests comprising a narrative task were given: posttest 1 with completely new pictures and posttest 2 with the same pictures as the pretest. This procedure was implemented from Bygate’s study (2001). In his study, after doing a task as a pretest, the participants were given two tasks, the

### Table 1

**Schedule of tests and training sessions**

<table>
<thead>
<tr>
<th>Training Sessions</th>
<th>Pretest/1</th>
<th>1 Week</th>
<th>2</th>
<th>1 Week</th>
<th>3</th>
<th>1 Week</th>
<th>4</th>
<th>1 Week</th>
<th>Posttests 1/2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repetition</td>
<td>a(noticing)/g</td>
<td>b(noticing)/h</td>
<td>c(noticing)/i</td>
<td>d(noticing)/j</td>
<td>e/a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Repetition</td>
<td>a(noticing)/f</td>
<td>b(noticing)/g</td>
<td>c(noticing)/h</td>
<td>d(noticing)/i</td>
<td>e/a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. The letters refer to the picture stories given.*
same task type with different topics, every two weeks, and at the final week, they performed two tasks: one with the same topic as the pretest and one different topic from the pretest.

2.3 Analysis

All data in IC recorders were analyzed with PRAAT 5.3.09 (Boersma & Weenink, 2012). First, the researcher confirmed that all information was transcribed correctly. Pauses were then found as follows. The beginning and end of each speech segment was 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,” “ah,” “um,” and “mmm”, 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 plus 2.5 standard deviations, as in De Jong and Perfetti. Syllables were counted by targeting only words uttered in their entirety, excluding false starts. Words repeated were also counted. To obtain a reliability measure, the number of pauses, the length of each pause, and the number of syllables were all reexamined by two research assistants.

For measuring fluency, based on De Jong and Perfetti (2011), the mean pause length, the phonation/time ratio, and the mean length of fluent runs were calculated as follows:

- (the mean pause length) dividing the total length of pauses by the number of pauses;
- (the phonation/time ratio) dividing the total time filled with speech (excluding the total pause length) by the total time spent speaking (including the total pause length) and then multiplying by 100;
- (the mean length of fluent runs) dividing the number of syllables by the number of pauses.

For measuring accuracy, the ratio of erroneous uses of target forms was calculated through dividing the number of errors of the target forms by the number of the forms used and multiplying by 100. The frequency of a target form has a big impact on the ratio of erroneous uses of the form. For avoiding cases whereby a target form is used less frequently or inconsistently in each task, verbs and articles were set as the target forms in this study because they were expected always to be used more frequently and consistently than other forms. Verb errors included errors in tense, word choice, and subject-verb agreement. Article errors covered all types of article uses.

For analyzing variables for accuracy and fluency, a two-way repeated measures ANOVA and $\eta^2$ were used. For analyzing main effect or simple main effect, $\eta^2$ was used. In multiple comparisons, $r$ for independent $t$-test and $\Delta$ for dependent $t$-test were used.

3. Results

Table 2 shows the mean scores and standard deviations of each group’s fluency and
accuracy at the pretest, and the posttests 1 and 2. No difference was found between the groups at the pretest (the mean length of pauses: $t = .154$, $r = .28$; the phonation/time ratio: $t = .229$, $r = .24$; the mean length of fluent runs: $t = .505$, $r = .13$; the mean number of target forms, verbs and articles: $t = .932$, $r = .02$; the ratio of erroneous uses of verbs and articles: $t = .833$, $r = .04$).

### Table 2

<table>
<thead>
<tr>
<th>Measures of proceduralization</th>
<th>pretest</th>
<th>posttest 1 (new task)</th>
<th>posttest 2 (same task)</th>
</tr>
</thead>
<tbody>
<tr>
<td>the mean length of pauses (in seconds)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Repetition</td>
<td>1.18 (0.28)</td>
<td>1.18 (0.27)</td>
<td>1.01 (0.24)</td>
</tr>
<tr>
<td>Repetition</td>
<td>1.06 (0.15)</td>
<td>1.17 (0.25)</td>
<td>0.93 (0.18)</td>
</tr>
<tr>
<td>the phonation/time ratio</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Repetition</td>
<td>44.00 (6.23)</td>
<td>41.50 (9.94)</td>
<td>48.86 (7.97)</td>
</tr>
<tr>
<td>Repetition</td>
<td>46.57 (4.68)</td>
<td>44.50 (4.68)</td>
<td>53.64 (6.02)</td>
</tr>
<tr>
<td>the mean length of fluent runs (in syllables)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Repetition</td>
<td>2.64 (0.64)</td>
<td>2.18 (0.52)</td>
<td>2.84 (0.74)</td>
</tr>
<tr>
<td>Repetition</td>
<td>2.51 (0.36)</td>
<td>2.39 (0.48)</td>
<td>3.34 (0.78)</td>
</tr>
<tr>
<td>the mean number of target forms used</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Repetition</td>
<td>24.43 (7.55)</td>
<td>23.36 (6.04)</td>
<td>28.29 (8.05)</td>
</tr>
<tr>
<td>Repetition</td>
<td>24.64 (5.37)</td>
<td>24.93 (6.08)</td>
<td>25.29 (4.79)</td>
</tr>
<tr>
<td>the ratio of erroneous uses of target forms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Repetition</td>
<td>37.79 (14.78)</td>
<td>40.71 (13.87)</td>
<td>28.07 (8.30)</td>
</tr>
<tr>
<td>Repetition</td>
<td>38.86 (11.71)</td>
<td>29.36 (14.76)</td>
<td>19.86 (12.12)</td>
</tr>
</tbody>
</table>

Notes. Values enclosed in parentheses represent standard deviations. $^a n = 14$, $^b n = 14$.

#### 3.1 Fluency

##### 3.1.1 The length of pauses

A two-way repeated measures ANOVA on the length of pauses showed significant main effect in the test with a medium sized effect ($F (2, 52) = 10.744$, $p = .000$, $\eta_p^2 = .13$), not in the group ($F (1, 26) = .961$, $p = .336$, $\eta_p^2 = .02$) and interaction between test and group ($F (2, 52) = .887$, $p = .418$, $\eta_p^2 = .01$). Multiple comparisons by $t$ test with Bonferroni correction ($\alpha = .016$) indicated that the participants made shorter pauses at the posttest 2 than the pretest with a medium sized effect ($t (27) = 3.420$, $p = .002$, $\Delta = -.66$) and the posttest 1 with a large sized effect ($t (27) = -5.995$, $p = .000$, $\Delta = -.95$). The difference was not significant between the pretest and the posttest 1 ($t (27) = -.994$, $p = .329$, $\Delta = .25$). In short, both groups made similar lengths of pauses at each test. On the other hand, when given the same task as the posttest, they made shorter pauses than the pretest. However, when given a new task, their pauses were not shorter than the pretest.
3.1.2 The phonation/time ratio

A two-way repeated measures ANOVA on the phonation/time ratio showed significant main effect in the test with a medium sized effect ($F(2, 52) = 19.529, p = .000, \eta^2_p = .22$), not in the group ($F(1, 26) = 2.821, p = .105, \eta^2_p = .06$) and interaction between test and group ($F(2, 52) = .371, p = .692, \eta^2_p = .01$). Multiple comparisons by $t$ test with Bonferroni correction ($\alpha = .016$) indicated that the participants made a higher phonation/time ratio at the posttest 2 than the pretest with a large sized effect ($t(27) = -5.044, p = .000, \Delta = 1.07$) and the posttest 1 with a large sized effect ($t(27) = -6.367, p = .000, \Delta = 1.12$). The difference was not significant between the pretest and the posttest 1 ($t(27) = 1.486, p = .149, \Delta = -.41$). In short, both groups had a similar phonation/time ratio at each test. On the other hand, when given the same task as the posttest, they produced a higher phonation/time ratio than the pretest. However, when given a new task, their ratio was not higher than the pretest.

3.1.3 The length of fluent runs

A two-way repeated measures ANOVA on the length of fluent runs showed significant main effect in the test with a medium sized effect ($F(2, 52) = 32.054, p = .000, \eta^2_p = .25$) and significant interaction between test and group with a small sized effect ($F(2, 52) = 4.866, p = .012, \eta^2_p = .05$), not in the group ($F(1, 26) = .970, p = .334, \eta^2_p = .03$). Looking at simple main effects by test, a significant difference was found in the No Repetition group with a medium sized effect ($F(2, 26) = 13.446, p = .000, \eta^2_p = .17$). Multiple comparisons by $t$ test with Bonferroni correction ($\alpha = .016$) indicated that the No Repetition group produced shorter runs at the posttest 1 than the pretest with a medium sized effect ($t(13) = 4.417, p = .001, \Delta = -.72$) and the posttest 2 with a large sized effect ($t(13) = -4.563, p = .001, \Delta = .89$). The difference between the pretest and the posttest 2 was not significant ($t(13) = -1.426, p = .178, \Delta = -.31$). A significant difference was also found in the Repetition group with a large sized effect ($F(2, 26) = 21.976, p = .000, \eta^2_p = .37$). On the other hand, the Repetition group produced longer runs at the posttest 2 than the pretest with a large sized effect ($t(13) = -4.642, p = .001, \Delta = 2.28$) and the posttest 1 with a large sized effect ($t(13) = -6.409, p = .000, \Delta = 1.22$). The difference between the pretest and the posttest 1 was not significant ($t(13) = .906, p = .381, \Delta = -.32$). Next, looking at simple main effects by group, the difference between the groups was significant at the posttest 2 with a medium sized effect ($t(26) = 1.745, p = .093, r = .32$), not at the pretest ($t(26) = -.676, p = .505, r = .13$) and the posttest 1 ($t(26) = 1.117, p = .274, r = .21$). In short, both groups first produced similar lengths of fluent runs. However, four weeks later, the Repetition group produced longer runs than the No Repetition group at the same task, but not at the new task. On the other hand, both groups produced longer runs at the same task than the new task. However, compared with the pretest, the Repetition group produced longer runs at the same task, but the No Repetition group’s runs were not longer. Furthermore, compared to the pretest, the length of runs by the Repetition group was similar at the new task, but the No Repetition group produced shorter runs at the new task.
3.2 Accuracy

A two-way repeated measures ANOVA on the number of target forms used, articles and verbs, showed significant main effect in the test with a small sized effect \( (F(2, 52) = 3.861, p = .027, \eta^2 = .04) \) and significant tendency in the interaction between group and test with a small sized effect \( (F(2, 52) = 2.616, p = .083, \eta^2 = .02) \), not in the group \( (F(1, 26) = .037, p = .850, \eta^2 = .00) \). Looking at simple main effects by test, a significant difference was found in the No Repetition group with a small sized effect \( (F(2, 26) = 5.213, p = .012, \eta^2 = .08) \). Multiple comparisons by \( t \) test with Bonferroni correction \( (a = .016) \) indicated that the No Repetition group produced more articles and verbs at the posttest 2 than the pretest with a medium sized effect \( t(13) = -2.874, p = .013, \Delta = .51 \) and the posttest 1 with a medium sized effect \( t(13) = -2.807, p = .015, \Delta = .61 \). The difference between the pretest and the posttest 1 was not significant \( t(13) = .635, p = .537, \Delta = -.14 \). Next, looking at simple main effects by group, the difference between the groups was not significant at all the tests: the pretest \( t(26) = .087, p = .932, r = .02 \), the posttest 1 \( t(26) = .686, p = .499, r = .13 \), and the posttest 2 \( t(26) = 1.198, p = .242, r = .23 \).

In short, both groups always produced similar amounts of articles and verbs. The Repetition group always produced similar amounts of articles and verbs. On the other hand, the No Repetition group produced more articles and verbs when given the same task.

A two-way repeated measures ANOVA on the ratio of erroneous uses of articles and verbs showed significant main effect in the test with a medium sized effect \( (F(2, 52) = 15.334, p = .000, \eta^2 = .20) \) and significant tendency in the interaction between group and test with a small sized effect \( (F(2, 52) = 2.829, p = .068, \eta^2 = .04) \). Looking at simple main effects by test, significant differences were found in the No Repetition group with a medium sized effect \( (F(2, 26) = 5.908, p = .008, \eta^2 = .17) \) and in the Repetition group with a large sized effect \( (F(2, 26) = 12.284, p = .000, \eta^2 = .28) \). Multiple comparisons by Wilcoxon signed-ranks test and \( t \) test with Bonferroni correction indicated that the ratio of erroneous uses of articles and verbs by the No Repetition group was lower at the posttest 2 than the pretest with a large sized effect \( Z = -2.028, p = .043, r = -.54 \) and the posttest 1 with a large sized effect \( t(13) = 4.317, p = .001, \Delta = .91 \). The difference was not significant between the pretest and the posttest 1 \( Z = -.630, p = .529, r = -.17 \).

On the other hand, the ratio of erroneous uses of articles and verbs by the Repetition group was lower at the posttest 2 than the pretest with a large sized effect \( t(13) = 5.191, p = .000, \Delta = -1.62 \) and the posttest 1 with a medium sized effect \( t(13) = 2.702, p = .018, \Delta = -.78 \). The ratio at the posttest 1 was also lower than the pretest with a large sized effect \( t(13) = 2.219, p = .045, \Delta = -.81 \). Next, looking at simple main effects by group, significant differences were found at the posttest 1 with a medium sized effect \( t(26) = -2.097, p = .046, r = .38 \) and at the posttest 2 with a medium sized effect \( t(26) = -2.091, p = .046, r = .38 \), not at the pretest \( t(26) = 2.130, p = .833, r = .04 \). In short, both groups first produced a similar ratio of erroneous articles and verbs. However, at the posttest, the ratio of the Repetition group was lower than the No Repetition group at both tasks. Furthermore, compared to the pretest, the Repetition group produced a lower ratio of
erroneous articles and verbs at both tasks whereas the ratio of the No Repetition group was lower at the same task only.

4. Discussion

Table 3
Summary of the data analysis of fluency and accuracy in the tests

<table>
<thead>
<tr>
<th>Fluency</th>
<th>Measure</th>
<th>Between comparison</th>
<th>Within comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>posttest 1</td>
<td>length of pauses</td>
<td>Repetition = No Repetition</td>
<td>= pretest</td>
</tr>
<tr>
<td></td>
<td>phonation/time ratio</td>
<td>Repetition = No Repetition</td>
<td>= pretest</td>
</tr>
<tr>
<td></td>
<td>length of fluent runs</td>
<td>Repetition = No Repetition</td>
<td>(Repetition) = pretest</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(No Repetition) &lt; pretest</td>
<td></td>
</tr>
<tr>
<td>posttest 2</td>
<td>length of pauses</td>
<td>Repetition = No Repetition</td>
<td>&lt; pretest; &lt; posttest 1</td>
</tr>
<tr>
<td></td>
<td>phonation/time ratio</td>
<td>Repetition = No Repetition</td>
<td>&gt; pretest; &gt; posttest 1</td>
</tr>
<tr>
<td></td>
<td>length of fluent runs</td>
<td>Repetition &gt; No Repetition</td>
<td>&gt; pretest; &gt; posttest 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Repetition) = pretest; &gt; posttest 1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Accuracy</th>
<th>Measure</th>
<th>Between comparison</th>
<th>Within comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>posttest 1</td>
<td>number of target forms used</td>
<td>Repetition = No Repetition</td>
<td>= pretest</td>
</tr>
<tr>
<td></td>
<td>ratio of erroneous target forms</td>
<td>Repetition &lt; No Repetition</td>
<td>&lt; pretest</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Repetition) = pretest</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(No Repetition) = pretest</td>
<td></td>
</tr>
<tr>
<td>posttest 2</td>
<td>number of target forms used</td>
<td>Repetition = No Repetition</td>
<td>= pretest; = posttest 1</td>
</tr>
<tr>
<td></td>
<td>ratio of erroneous target forms</td>
<td>Repetition &lt; No Repetition</td>
<td>&lt; pretest; &lt; posttest 1</td>
</tr>
</tbody>
</table>

Table 3 is the summary of the data analysis of fluency and accuracy in the tests. Here the two hypotheses are examined. Hypothesis 1, that gains in fluency from a pretest to a posttest would be larger for participants who repeated than for those who did not, was partially supported. At the pretest, both groups were similar in the length of fluent runs, the length of pauses and the phonation/time ratio. In the same task at the posttest, the Repetition group produced longer fluent runs than the No Repetition group whereas no difference was found in the length of pauses and the phonation/time ratio. In other words, in the same task, the Repetition group produced longer stretches of speech than the No Repetition group without additional time for pausing. Next, looking at within comparisons to the pretest, both groups needed shorter times for pausing than the pretest. However, the Repetition group produced longer fluent runs and the No Repetition produced similar lengths of fluent runs. In other words, in the same task, the Repetition group produced longer stretches of speech with shorter pauses than the pretest whereas the No Repetition group produced similar stretches of speech to the pretest with shorter pauses than the pretest. Therefore, it can be said that a gain in fluency from pretest to posttest on the same task
was seen in the participants who did not have task repetition but the gain was larger for the participants who had task repetition. This may indicate the greater effectiveness of task repetition in facilitating proceduralization compared to no task repetition.

In the study by Bygate (2001), participants had two different narrative tasks in each of three sessions and two posttests: one using the same task as a pretest and the other the same task type but different from the pretest. The results then showed that there was no difference in fluency between the same task and the new task of the same type. However, in this study, the No Repetition group, who followed a similar procedure to the participants in Bygate’s study, produced longer fluent runs, a higher phonation/time ratio and shorter pauses in the same task than the new task. One possible reason for such difference between Bygate’s study and the present study could be that the No Repetition group in the present study had opportunities to reflect on the first task and notice erroneous forms before the second task in each session. The reflection on the first task could become internal repetition, “repetition which is encouraged by the demands of processing the input material and/or of preparing the intended task outcome” (Bygate, 2006, p. 173). In other words, the No Repetition group might have had some kind of repetition. Furthermore, they had noticed an erroneous form and correctly connected the form with its meaning/function; therefore, their consciousness of the connection might have been raised before the second performance. Such internal repetition and/or consciousness raised might then have influenced their repeatedly retrieving declarative knowledge of the connection in the second performance of a new task. Consequently, when they undertook the same task at the posttest, both processes of conceptualization (message generation) and formulation (grammatical encoding) might have been boosted, engendering more fluency than the new task which would have required them to pay more attention to, and be burdened by, conceptualization. This may indicate the effectiveness of reflection and/or noticing after the first performance on facilitating proceduralization at the same task at least.

In the new task at the posttest, on the other hand, no difference was found between the two groups in the length of pauses, the phonation/time ratio and the length of pauses. Next, looking at within comparisons to the pretest, both groups produced similar lengths of pauses and phonation/time ratios to the pretest. The Repetition group also produced similar lengths of fluent runs to the pretest. However, the No Repetition group produced shorter lengths of fluent runs than the pretest. In other words, no matter if the pictures in the posttest were new, the Repetition group produced a similar length of speech stretches and speech ratio to the pretest without additional time for pausing. However, the No Repetition group produced shorter speech stretches than the pretest although the time for pausing did not increase. Therefore, it can be said that a gain in fluency from pretest to posttest on the new task was not seen in either the participants who repeated or those who did not; however, the participants who did not repeat lost fluency at the new task probably because of the difficulty of the pictures in the task whereas those who repeated did not lose fluency. This may imply the effectiveness of task repetition on facilitating
proceduralization and transfer to a new task.

In the study by De Jong and Perfetti (2011), participants in the Repetition group spoke about the same topic three times whereas those in the No Repetition group spoke about different topics three times. The participants then had an immediate posttest one week after the last session with new topics. The study found that some participants in the Repetition group showed differences between the pretest and a new task at the immediate posttest regarding the length of pauses and the phonation/time ratio whereas others showed the difference regarding the length of fluent runs. In other words, in De Jong and Perfetti (2011), the Repetition group showed improvement from the pretest to the posttest, but the No Repetition group did not show improvement. However, in the present study, the No Repetition group showed regression from the pretest to the posttest, while the Repetition group did not show the regression. The Repetition group in this study may show another type of evidence for proceduralization. The participants in this study had lower proficiency than those in De Jong and Perfetti (2011) regarding the mean length of pauses ($d = .27$), the phonation/time ratio ($d = .96$) and the length of fluent runs ($d = 2.25$). Therefore, it may be implied that task repetition can facilitate proceduralization and transfer to a new task even if learners do not have a specific level of proficiency.

Hypothesis 2, that gains in accuracy from a pretest to a posttest would be larger for participants who repeated than for those who did not, was supported. At the pretest, both groups used a similar number of target forms, articles and verbs, and the ratio of erroneous uses of articles and verbs was also similar. In the same task at the posttest, the Repetition group produced fewer errors of articles and verbs than the No Repetition group although the two groups used a similar number of articles and verbs. Next, looking at within comparisons between the same task and the pretest, both groups made fewer errors of articles and verbs in the same task than the pretest although neither group used a lesser number of articles and verbs in the same task than the pretest. Therefore, it can be said that a gain in accuracy from the pretest to the posttest on the same task was seen in the participants who did not have task repetition, but the gain was larger for the participants who had task repetition. The result that the No Repetition group gained accuracy contrasts with what Bygate (2001) found. In his study, participants who had no task repetition did not improve accuracy in the same task at the posttest regarding the number of errors per t-unit. This difference between Bygate's study and the present study might have been caused by participants having had no opportunity to notice erroneous forms in his study. This may indicate the effectiveness of noticing before the second performance for facilitating proceduralization and increase in accuracy. Furthermore, looking at within comparison between the new task and the same task, both groups also made fewer errors of articles and verbs in the same task than the new task although neither group used a lesser number of articles and verbs in the same task than the new task. This may indicate the effect of priming, which is the phenomenon in which prior exposure to specific language forms or meanings facilitates a speaker's subsequent language production (Trofimovich & McDonough, 2011). Bygate (2001) also stated that doing the same
task may assist 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” (p. 29). In the present study, the Repetition group produced fewer errors of articles and verbs than the No Repetition group in the same task. This better accuracy by the Repetition group in the same task may indicate the greater effectiveness of task repetition on facilitating proceduralization compared to no task repetition.

In the new task at the posttest, the Repetition group produced fewer errors of articles and verbs than the No Repetition group although the two groups used a similar number of articles and verbs. Next, looking at within comparisons with the pretest, both groups used a similar number of articles and verbs in the new task. However, the Repetition group produced a lower ratio of erroneous articles and verbs in the new task than in the pretest while the No Repetition group produced a similar ratio of erroneous articles and verbs in the new task and the pretest. Therefore, it can be said that a gain in accuracy from pretest to posttest on the new task was seen in the participants who had task repetition. This may indicate the effectiveness of task repetition on facilitating proceduralization and transfer of increase in accuracy to speaking about new pictures. Yuan and Ellis (2003) proposed the trade-off effect between fluency and accuracy. However, the task repetition with noticing given in the present study seems to have helped the Repetition group perform both the same task and the new task accurately without sacrificing their fluency.

The aggregated results of these two hypotheses may indicate that proceduralization of linguistic knowledge occurred in this study. 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). Specifically, the increase of accuracy in the same context found in this study is related to items (1) and (2). The increase of accuracy in the new context found in this study is related to (3). Finally, the increase of fluency in the same context and the new context found in this study is related to (4).

In other words, proceduralization of linguistic knowledge in a speaking task may take place through the opportunity to notice erroneous forms after the first performance and then soon undertaking the next task. How much proceduralization is facilitated, that is, how much fluency and/or accuracy improves in doing the same task and/or a new task, will then be evident in the task undertaken after the noticing: either the same task or the same task type but with a new task.

5. Conclusion

The two major findings in this study were:
1. Fluency and accuracy in the same task will be improved from the present level through either task repetition with noticing or no task repetition with noticing. However, proceduralization of linguistic knowledge through task repetition will engender greater fluency and accuracy in the same task than having no task repetition.

2. Accuracy in a new task will be improved through task repetition with noticing, and proceduralization of linguistic knowledge through task repetition will lead to greater accuracy in the new task than having no task repetition.

“Proceduralization is considered a slow process that requires many encounters with the same items” (De Jong & Perfetti, 2011, p. 562). The many encounters with articles and verbs through task repetition with noticing in this study might have developed procedural knowledge enough for it to be used in the same task for improving fluency and accuracy and in the new task for improving accuracy. 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.

However, there are at least three limitations in this study. First is that the forms targeted in the study were not sufficiently focused. Verb errors included tense, word choice, and subject-verb agreement. Article errors covered all types of article uses. Therefore, it was not clarified if a specific feature of a form was used erroneously, the feature was noticed, and/or the form with the same feature was used repeatedly at the posttests. In the process of proceduralization, it is necessary for learners to experience first the form-meaning connection in an exemplar repeatedly, and later experience the same form-meaning connection in many exemplars repeatedly and then generalize the connection. A specific feature of a focused form should have been targeted in order to examine if task repetition can facilitate proceduralization. The second limitation is that it is not clear if the opportunity for noticing might have had any impact on proceduralization. In this study, there was no group which was not given any opportunity for noticing erroneous articles and verbs. Therefore, it was not possible to examine the effectiveness of the opportunities for noticing. The last limitation is that it is not clear if the two groups were at first equivalent in terms of (1) language proficiency related to the ability to notice errors through feedback by underlining and/or metalinguistic explanation, and (2) being at the developmental stage where proceduralization can be facilitated with/without task repetition.

Considering such limitations, further research is necessary to expand the potential of task repetition to facilitate the proceduralization of linguistic knowledge. First, it is necessary to examine the effects of various types of task repetition, such as repeating a task after a while (Bygate, 2001), or immediately repeating a task numerous times (De Jong & Perfetti, 2011; Lynch & Maclean, 2001). Second, it is necessary to examine the long-term effect of the combination of task repetition and noticing by a delayed posttest. De Jong and Perfetti (2011) gave a new task as an immediate posttest one week after the last training session and another new task as a delayed
posttest three weeks after the last training, so giving a delayed posttest might clarify whether the effectiveness was temporary or will be retained.

Notes

1. The sample size was decided by G*Power 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009). According to a priori analysis, the necessary total sample size was 28, with $1-\beta = 0.8115$.
2. The interpretations of $\eta_p^2$ were based on Bakeman (2005): $|.02|$ ≤ small < $.13$; $.13$ ≤ medium < $.26$; $.26$ ≤ large.
3. The interpretations were based on Cohen (1988): $r$ (|.10| ≤ small < .30; .30 ≤ medium < .50; .50 ≤ large) and $\Delta$ (|.20| ≤ small < .50; .50 ≤ medium < .80; .80 ≤ large).
4. The interpretations of $d$ were based on Cohen (1988): $.20$ ≤ small < .50; $.50$ ≤ medium < .80; .80 ≤ large.

References


