Narrative Task-Type Repetition and Changes in Second Language Use in a Classroom Environment: A Case Study

Masanori MATSUMURA
Kazuyo KAWAMURA
Armand AFFRICANO

Reprinted from
THE EDUCATIONAL SCIENCES
Journal of the Faculty of Education and Human Sciences
Yokohama National University
No.10, FEBRUARY, 2008
Narrative Task-Type Repetition and Changes in Second Language Use in a Classroom Environment: A Case Study

Masanori Matsumura  Yokohama National University
Kazuyo Kawamura   Kogakkan University
Armand Affricano  Kansai Gaidai University

Abstract
Second language researchers have investigated into how aspects of learner language are affected by particular tasks under particular conditions. However, most of the previous studies were limited in the sense that they only looked at one snapshot of a learner’s performance. This study attempts to gain insights into second language development that may have resulted from task-based learning over an extended period of time. Changes in the productive skills of one learner are reported that were brought about by repetitive use of the narration task over three weeks in a classroom environment. The learner gained in accuracy and certain aspects of fluency not only in the familiar narrative task but also in his impromptu speech, though his language did not become more complex. When these results were compared with the performance in the same narration and speech tasks by another learner who had repeated decision-making tasks, several interesting contrasts were identified. The results may suggest that there are aspects of learner language that are upgraded by the repetition of specific types of tasks.

Task-based language teaching (TBLT) has been the center of attention in second language (L2) pedagogy as a methodological innovation, and also as a target of research among L2 researchers.1 In its essence, a task is a classroom activity that requires learners to use a second language in a meaningful context in order to attain a (nonlinguistic) goal, during which the learners are supposed to be involved in the same cognitive processes as in natural language use (cf., Ellis, 2003). TBLT is expected to provide the learners with a rich “holistic second-language experience” (Willis, 1996, p. 135) in the sense that they are provided with opportunities to be exposed to as well as to produce the target language, and also have opportunities for grammatical consciousness-raising as a post-task activity in many TBLT options.

Part of the recent academic interest in task-based instruction may have grown, as Ellis (2003) points out, out of the expectation that the task is the “shared construct” (p. 34) on which the interests of L2 researchers and practitioners meet. This is noteworthy especially when we look back on the history of disconnection between L2 research and L2 teaching practices. The ultimate goal of the joint endeavor of practitioners and researchers who are interested in tasks and task-related learning/teaching should be to design a task-based syllabus in which tasks are sequenced according to their difficulty, complexity, and expected developmental gains. In order for this goal to be attained, information has to be obtained as to what type of task has what effect and how the conditions for task implementation affect the process.

Previous task-related research may be categorized into two broad types.2 One is the study of interaction among learners, or between learners and the instructor, when they are engaged in a particular type of task. Most of the studies in this tradition are heavily influenced by Long’s Interaction Hypothesis (e.g., Long, 1981, 1996), and have been carried out with the assumption that input of good quality and/or corrective
feedback (i.e., clarification request, confirmation check, recast, etc.) given in the course of negotiation of meaning should promote L2 acquisition. In those studies, learners’ task performance has been analyzed in terms of how much meaning negotiation and corrective feedback is generated in their completion of a particular task under a particular condition (e.g., Berwick, 1991; Duff, 1986; Newton, 1991) and how it results in better comprehension and/or acquisition (e.g., Mackey, 1999).

The second type of task-related research is the study of task-generated language production. The precursor to this line of research may be found in the study of variation in the 1980s (e.g., Tarone, 1983), which showed that the language produced by L2 learners varies according to the amount of attention paid during speech to formal features of the target language, exactly as revealed by sociolinguists such as Labov (e.g., 1972) in regard to first language production. Recently, many researchers have been vigorously investigating how L2 learners’ performance varies according to the type of task (e.g., narrative, decision-making, opinion-exchange, etc.) they are engaged in and the conditions for task implementation (e.g., with or without chances for planning time), in terms of such dimensions as fluency, accuracy, and complexity of their language. The present study may be located among this second group of studies, but tries to go further into a relatively unexplored research area, i.e., the potential for TBLT to develop the productive skills of the learners in the long run, rather than the influence of task type or task condition on a one-time learner performance. Specifically, we attempt to gain insights into how a repetitive use of the narrative task causes changes in the fluency, accuracy, and complexity levels of learner language. Before reporting the design and the results of the present study, in the next section, we will review relevant previous studies on task-based production.

REVIEW OF PREVIOUS STUDIES

As noted above, the study of task-based production has been trying to uncover how the conditions for task implementation and the design features of the tasks themselves influence aspects of learner language. Many studies dealing with implementational variables have addressed the effect of planning. It has been reported in the literature that when learners are afforded extra time prior to a task activity, the complexity and fluency of their language increase (e.g., Ortega, 1999). On the other hand, if they are allowed to plan on-line (i.e., to take as much time as they need for task completion) the learners use the opportunity mainly for increasing the accuracy of their language (Yuan & Ellis, 2003). Skehan and Foster (1999) reported that in the narration task, it is only when the storyline was structured (i.e., has a basic predictable sequence of events) that a chance for on-line planning made the learner language more accurate. Mehnert (1998) found that different lengths of pre-task planning had different effects on learner’s L2 production. A short (one minute) preparation time helped learners to increase accuracy, with little further improvement with the provision of additional time. Fluency increased proportionally as a function of the length of the time given. Complexity increased only when a considerable length of planning time (ten minutes) was available. Other studies dealing with implementational variables have investigated the effects of the introduction of a surprise element (Foster & Skehan, 1997), the effects of the organization and management of the time for pre-task planning (Foster & Skehan, 1999). See Ellis (2003) for an excellent review of these and related studies.

The research on task design variables has examined how inherent characteristics of a task influence learner performance. For example, in their papers in the 1990’s, Skehan and Foster (e.g., Foster & Skehan, 1996) looked at learner performance in three different types of tasks: the personal (information conveying) task, the narrative task, and the decision-making task. More recently, developing Skehan’s (1992) theorizing, Skehan (2001) abstracted out five task features (i.e., familiarity of information, dialogic vs. monologic task,
degree of structure, complexity of outcome, and transformations) and analyzed how each of these parameters determine L2 production by the learner. In many of the studies in this tradition, the design of the experiment is notably complex with many variables processed simultaneously, so that the results are often considerably difficult to interpret. However, it may be possible to list several of the major findings to date:

- The monologic narrative task lowers the levels of complexity and accuracy of learner language, particularly in the condition of telling while watching a video as opposed to telling after watching. (Skehan & Foster, 1999).
- Learner language tends to be accurate and complex in the decision-making task (e.g., Duff, 1986; Newton & Kennedy, 1996).
- Closed tasks (i.e., the tasks that have but single solutions) do not motivate learners toward complex language use (e.g., Brown, 1991; Tong-Fredericks, 1984).
- Among open tasks that allow multiple solutions, learners’ utterances are more complex in divergent tasks than in convergent tasks (Duff, 1986).
- Tasks that have a firm structure invite fluent language use (Skehan & Foster, 1999).
- Tasks that require a complex outcome motivate learners to produce complex language. (Skehan, 2001)

Thus, there is now a collection of significant findings on how a learner’s performance is affected by inherent task characteristics and conditions for task completion. Importantly, however, a developmental perspective is missing in the previous discussion: A majority of studies have been concerned only with performance, or one-time language use, rather than the development of performance skills over time. Arguably, a one-time good performance and continued, consistent progress in the long run are two different things. The fact that the accuracy, complexity or fluency level is low in a one-time performance of a particular task by no means predicts what will happen regarding the learner’s language when the same type of task is repeated over an extended period of time. It would not be surprising if repeated struggles with a task of the proper level of difficulty for a given learner helps them develop the skills necessary for its completion, just as athletes increase their physical capacities by working out under adequately overloaded conditions, even though the performance level is low on the first trial. Thus, it would be important to inquire into the long-term potential-stretching function of the repeated use of task.4

This, however, does not imply that no research has been done that employs a developmental perspective. In a series of studies, Martin Bygate and his colleagues (e.g., Bygate, 1996, 1999, 2001; Bygate & Samuda, 2004) have been investigating how task repetition (or “task rehearsal” in their terminology) influences learner language.5 Referring to Levelt’s (1989) speech production model, these researchers assume that when learners repeat the same type of task (task-type repetition), or exactly the same task (task repetition), the first encounter functions as one form of planning, so that the processes of conceptualization and formulation will be easier in a subsequent performance.

Bygate’s (2001) research is of special importance in the present context because of its elaborateness as well as its similarity and relevance to the study that is to be reported below. See Table 1 for a schematization of Bygate’s research design (his Table 2.1). In his experiment, Bygate divided his participants (overseas students studying at the University of Reading, Britain) into three groups: a narrative group, an interview group, and a control group. During the first week of the ten-week experiment, the participants in every group were engaged in one narrative task and one interview task. After that, the learners in the narrative group met on a fortnightly basis, once in each of the
Table 1 Bygate’s (2001) experimental design

<table>
<thead>
<tr>
<th>Group</th>
<th>Treatment 1</th>
<th>Treatment 2</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Week 1</td>
<td>Week 3</td>
<td>Week 5</td>
</tr>
<tr>
<td>Narrative</td>
<td>Nar. 1</td>
<td>Nar. 2&amp;3</td>
<td>Nar. 3&amp;4</td>
</tr>
<tr>
<td>Group</td>
<td>Int. 1</td>
<td>Int. 2&amp;3</td>
<td>Int. 3&amp;4</td>
</tr>
<tr>
<td>Interview</td>
<td>Nar. 1</td>
<td>Int. 1</td>
<td>Nar. 1&amp;6</td>
</tr>
<tr>
<td>Group</td>
<td>Int. 1</td>
<td>Int. 2&amp;3</td>
<td>Int. 3&amp;4</td>
</tr>
<tr>
<td>Control Group</td>
<td>Nar. 1</td>
<td>Nar. 1&amp;6</td>
<td>Int. 1&amp;6</td>
</tr>
<tr>
<td></td>
<td>Int. 1</td>
<td>Int. 1</td>
<td></td>
</tr>
</tbody>
</table>

*Note:* Nar. = Narration; Int. = Interview


third, fifth and seventh weeks, and were engaged in two narration tasks on each occasion. One of the two tasks they encountered in the fifth and seventh weeks were recycled from the previous treatment session, and the other was a newly added task. The interview group followed the same procedure with interview tasks. The learners were given neither feedback nor instructions throughout because the researcher considered these to be ‘noises’ that possibly mask the genuine effect of task and/or task-type repetition. Meanwhile, the control group did nothing at all. Finally, in the tenth week, all the participants in the three groups were asked to complete two narrative tasks and two interview tasks. One task in each type was identical to the one administered in the first week, the purpose of which was to see the effects of task repetition. The other was a new task of the same type, with which the effects of task-type repetition were expected to be shown.

Bygate’s (2001) major findings were: (1) In all the three groups, learner performance improved in terms of complexity but not in terms of accuracy in their completion of the same task that they had experienced ten weeks before; (2) Only when the interview group was engaged in the same interview task that had been encountered ten weeks before did the learners’ language become more fluent; (3) The effect of task-type repetition was minimal even when the learners in the experimental groups had practiced particular types of tasks (narration and interview, respectively) several times over weeks. In short, the learners performed significantly better on some measures on the second trial of an old task, but improvement was not evident when they were engaged in a new task of the same type. Thus, Bygate’s (2001) hypothesis was supported in that task repetition was effective in several regards. However, his hypothesis was not supported in terms of the effect of task-type repetition. This part of his results is somewhat puzzling, for if, as Bygate and Samuda (2004) note, “task repetition can have an impact on the processing, and not just on the product” (p. 45) of speech, it should be expected that the revised processing strategies help the learner carry out a new task of the same type.

At least a part of the results that Bygate (2001) obtained may be attributable to the design of his experiment. First, in his experiment, treatment sessions (his “Treatment 2”) were distributed over the course of eight weeks between the pretest (his “Treatment 1”), and the final data collection. It would not be entirely surprising, under the condition of such a limited encounter with the task, that the treatment did not have a significant impact on learner’s processing mechanisms, as Bygate (2001) himself acknowledges (p. 43). This is all the more likely when we consider that the participants were learning English in Britain, where the language is the medium of daily communication, and, thus, were exposed to, and had chances to speak, English on a daily basis. It would be natural, then, that the amount of English they experienced outside of
the classroom during the ten-week experiment overrode the amount of English they encountered in the experimental sessions, a point that Bygate himself also acknowledges. It may not be unreasonable under these conditions that Bygate (2001) did not obtain clear evidence of the effects of his treatment, especially of task-type repetition.6

From a pedagogical perspective, there is a concern for the ‘ecological validity’ of Bygate’s (2001) results. As noted above, there were no accompanying pre-task activities/guidance or post-task feedback/instruction given to the learners throughout the entire length of his experiment. This by no means reflects the context in which tasks are used in a classroom. Tasks are to be used in instructional practices and, thus, it is natural that a task activity is embroidered by pre- and post-task activities, including post-task focus on form in most TBLT options (cf., Long & Robinson; Willis, 1996, 2004; Willis & Willis, 2007). The way tasks were used in isolation from their additional components by Bygate (2001) makes it difficult to gain insights into how tasks work in real-world classroom situations.7

Gass, Mackey, Fernandez and Alvarez-Torres (1999) are also negative toward the effects of task-type repetition, while acknowledging the effects of task repetition. They examined the effect of three consecutive uses of an identical narration task on learners’ L2 production. The learners were rated as having improved in general terms on their third performance by English-speaking raters. Improvement was also found in regard to several specific grammatical features and vocabulary. However, the effect of task repetition did not transfer to a subsequent occasion on which the learners carried out a new task of the same type. One interpretation of this is that the impact of the three-time repetition of a task had not been significant enough to cause changes in the internal processing mechanisms of the learners, to the extent of being displayed in performing a new task of the same type.

Thus, while the studies have reported the evidence of the impact of task repetition, they have failed to show that the effects transfer to a new task of the same type. However, it may be still reasonable to expect that effects of task-type repetition will emerge if a certain conditions (most critically the intensity of repetition) are met. As reviewed above, performance studies have demonstrated that a particular type of task encourages a particular language use, or a particular processing. Repeated engagement in the particular type of task, then, will expectedly develop a particular “processing strategies” for speech production in the learner, as a repetitive engagement in a particular sport will develop particular athletic skills and particular muscle groups. A further prediction may be that particular aspects of L2 productive skills fostered by a repeated experience with a particular type of task are transferred to and manifested in performances on different tasks. This may seem obvious but can by no means be taken for granted. If different tasks require different processing strategies, it is possible that the mechanism established by the repetitive use of a task is not reliably drawn on in carrying out another task, as skills or muscles developed by a sport is sometimes useless in another sport. Thus, the cross-task transferability of the task effect has to be empirically established.

THE PRESENT STUDY

The general design
In the brief review of the literature presented above, we pointed out that the mainstream task studies have mainly, though not exclusively, been concerned with L2 production or use, rather than with development. Those studies that have addressed task and task-type repetition had their design problems that made some of their conclusions open to scrutiny. In this circumstance, we report a semi-longitudinal case study and discuss the effects of task-type repetition on the development of L2 production skills. In order to circumvent
Bygate’s (2001) limitations (see the previous section), in the present study:

1. The treatment sessions were more intensive than in Bygate’s (2001) study: three times a week for a three-week period, hence the label “semi-longitudinal.”
2. Participants were Japanese learners studying English as a foreign (as opposed to second) language in university classrooms in Japan, outside of which they were not exposed to the target language.
3. In order to meet the criterion of ecological validity, tasks were used in normal language classes, in which each lesson was organized around a task.

A narration task was used in the study. The use of the narration task is popular in the language classroom and has also been widely used for research purposes (e.g., Foster & Skehan, 1997, 1999; Yuan & Ellis, 2003). Bygate (1999) claims that the narration task invites “linguistically denser talk” (p. 206), the expectation being that it contributes to L2 development. The advantages of using the narrative task in L2 research are also reiterated by Kawauchi (2005). Her point is that such monologic tasks as narration are cognitively demanding because the learners cannot solicit help from their conversational partners, and, thus, that the very fact that it is monologic makes it desirable for assessing L2 oral productive skills of individual learners. Referring to Ortega (1999), Kawauchi further emphasizes the point that the story-telling task effectively limits the range of individual variation in language use (p. 148).

The present study attempts to explore the effects of repeated use of the narrative task in terms of the changes in the degrees of accuracy, complexity, and fluency of the language of one learner (named Yūta), by analyzing his narrative performance before and after nine lessons. In each of the nine lessons, narration tasks were used as a main component of the lesson, with accompanying pre-task and post-task modules. This is one of the points of departure from Bygate (2001), in which tasks were used in isolation, with no add-on activities. See below for more details of the classroom management procedures of the present study.

One of the concerns of the present study is transferability of the effects (if any) of task-type repetition. As argued at the end of the previous section, if the extensive use of one type of task really has an impact on the learner’s language-generating mechanism, we may well expect that the effect is not restricted to the kind of task that the learner has experienced. With this in mind, in addition to the narration task, Yūta was asked to make a speech of more than one minute on given topics in the pre-test and post-test sessions. (The topics were different on the two occasions.)

Further, in order to gain clearer insights into the effects of the repetitive use of the narration task, it was desirable to compare Yūta’s gains/losses with those of another learner who was engaged in a different type of task. For this purpose, we also report data from one additional learner (named Kōtarō) who attended a class in which dialogic decision-making (argumentation) tasks were used for an equivalent three-week period. He was summoned to the pre- and post-test sessions to perform the same narration and speech tasks that were administered to Yūta. Below, we present data and discuss the results of the study by comparing the features of the language produced by the two learners.

**Participants**
The present study is a case study in nature, though it embodies an experimental operationalization and the pre-test, post-test design. As mentioned above, we look at the performance of one Japanese learner of English as compared with another on two occasions, prior to and following three weeks of task-based lessons in which particular type of task was repeated. Both of the participants were first-year male students (age: 18) in the International Language Department of Kansai Gaidai University, located in Osaka, Japan. The
TOEFL (PBT) scores for both of the participants were around 480, which means that they were advanced beginners. Both started learning English at the age of twelve and had learned in dominantly grammar-based programs before entering university. During the weeks of the present study, they did not attend any other classes that aimed at enhancing their oral proficiency in English. Nor did they participate in any activity outside the classroom that may have helped them expand their speaking abilities. This means that any possible developmental gains can be attributed to their participation in the present study.

Materials

The topics for the narrative and decision-making tasks were chosen from published ELT textbooks. As our concern was task-type repetition rather than task repetition, different tasks were used in different lessons. The narration topics for the two tests and the nine lessons were all based on the episodes in *Only in America* and the *Grapevine* series, both published by Oxford University Press. All the episodes depicted scenes from people’s lives and were five to ten minutes long. For the decision-making tasks that were to be used in the class that Kōtarō attended, several units from *Let's Start Talking, Can’t Stop Talking* and *The Non-Stop Discussion Workbook*, all authored by George Rooks and published by Heinle & Heinle, were used, with minor adaptations for Japanese students in some cases.

The topics for the impromptu speeches were “My Strong Points” for the pre-test, and “Requests to the University President” for the post-test. Immediacy, or the relevance to the students, was the reason for the choice of these topics. They were expected to be of equivalent cognitive demand for the learners. Table 2 is a comprehensive list of the topics of the tasks that Yūta and Kōtarō carried out in the two tests and the nine lessons.

<table>
<thead>
<tr>
<th>Table 2 Design of the study and task topics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Yūta</strong></td>
</tr>
<tr>
<td>Pre-test</td>
</tr>
<tr>
<td>NR1: The Websters (<em>OA</em>)</td>
</tr>
<tr>
<td>SP1: My Strong Points</td>
</tr>
<tr>
<td>Lesson 1</td>
</tr>
<tr>
<td>Lesson 2</td>
</tr>
<tr>
<td>NR3: The Wedding (<em>G2</em>)</td>
</tr>
<tr>
<td>Lesson 3</td>
</tr>
<tr>
<td>Lesson 4</td>
</tr>
<tr>
<td>NR5: Strange Encounter (<em>OA</em>)</td>
</tr>
<tr>
<td>Lesson 5</td>
</tr>
<tr>
<td>NR6: At the Doctor’s (<em>G2</em>)</td>
</tr>
<tr>
<td>Lesson 6</td>
</tr>
<tr>
<td>NR7: Office Blues (<em>OA</em>)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Lesson 7</td>
</tr>
<tr>
<td>NR8: Good Morning, Greenstown (<em>OA</em>)</td>
</tr>
<tr>
<td>Lesson 8</td>
</tr>
<tr>
<td>NR9: Dennis Cook’s Party (<em>G3</em>)</td>
</tr>
<tr>
<td>Lesson 9</td>
</tr>
<tr>
<td>NR10: The Artist (<em>OA</em>)</td>
</tr>
<tr>
<td>Post-test</td>
</tr>
<tr>
<td>NR11: A Day in the Life of Dennis Cook (<em>G1</em></td>
</tr>
<tr>
<td>SP2: Requests to the University President</td>
</tr>
</tbody>
</table>

Note: NR=Narration task; SP=Speech; DM=Decision-making task; *OA=* *Only in America*; *G1*= *Grapevine 1*; *G2*= *Grapevine 2*; *G3*= *Grapevine 3*; *CST*= *Can’t Stop Talking*; *LST*= *Let’s Start Talking*; *NSDW*= *The Non-Stop Discussion Workbook*
Procedures

Pre-test: On the first day of the study, Yūta and Kōtarō were asked to come to the pre-test in turn with their respective partners, who were nominated by the experimenter from among their classmates. In a room that was specially set up for recording, Yūta and Kōtarō, in their respective sessions, first watched a short video with recorded sound. The video was, then, immediately replayed, but with the sound off. During the second viewing, the participants were required to give an “on-line narration of the story.” That is, the learners had to describe the events happening on the screen while following the progress of the story. In this way, the conditions for the task were equivalent to what Skehan and Foster (1999) called “watch, watch and tell” in their experiment.

The rationale for this particular implementation of the narration task was related to the assumption that adequate overloading would assist development. Having had a chance to watch the video before they narrated the story, the learners were familiarized with the story content but, without time for planning before the second viewing, they supposedly had not have time to prepare forms to be used effectively in their task performance. Nor did they have the luxury of time for on-task planning due to the on-line nature of the task. Thus, the task was supposed to be a considerable challenge for the learners to the extent that it met our research purposes. Yūta and Kōtarō narrated the first half of the episode to their respective partners, and the roles were reversed in each pair for the latter half.

After completion of the narration task, the learners were asked to make a speech on what they thought to be their own strong points, without the opportunity for prior planning. For both the narration and the speech, no pre-task guidance or post-task form-focused activities were provided in this pre-test session.

Lessons: During the following three weeks, as classroom activities in the university classes, Yūta was engaged in nine narrative tasks in nine lessons, while Kōtarō was administered nine decision-making tasks. There are different proposals among researchers as well as practitioners and material developers as to how TBLT classes should be organized. Nunan (2004) claims that tasks should be introduced step by step, with the teacher providing useful vocabulary, schemas on the task topic, and grammar in a bottom-up fashion before getting down to the task. Other researcher-practitioners such as Willis (1996, 2004; also see Willis & Willis, 2007) strongly argue against pre-task grammar focus, claiming that free communication would be impeded by the preceding instruction. The present study followed Willis’s sequence of pre-task, task cycle, and post-task language focus (see the left-most column of Table 3 below for her framework). As Ellis (2003) and other researchers point out, her framework is prescriptive in the sense that it has no empirical basis. Nevertheless, to the knowledge of the present authors, it is a major TBLT option that is widely referred to and that is compatible with the tenet of “focus on form” (cf., Long, 2007; Long & Robinson, 1998, etc.). The central claim of focus on form is that attending to form in the context in which communicative intent is already clear and transparent is helpful in promoting acquisition of the form. Another reason for the non-provision of pre-task grammar instruction in this study was related to the fact that the participants were Japanese university students who had learned pedagogical English grammar extensively throughout their secondary education and, thus, had a substantial knowledge base to be referred to in task completion, so that a pre-task introduction of grammar was simply unnecessary (see Swan, 2005, for a related discussion).8

The class that Yūta attended was first introduced to the video topic of the day and then was given minimal guidance on useful lexical items as a pre-task activity. The introduction was kept at a minimum in order to encourage unrestricted language use in the following task activity. The procedures for narration itself were the same as in the pre-test as explained above: Each learner narrated either the first or the second
half of the story during the second viewing of the video to his or her partners. After the task activity, the learners in the class were given ten minutes to discuss in English the storyline and the expressions to be used effectively in their narration with their partners of the day, while referring to a printout carrying several cuts from the video scenes that were distributed to guide the discussion. During the discussion, the learners were encouraged to ask questions freely to the teacher (Armand Affricano). This phase was intended to correspond to Willis’s (1996, 2004) “Planning,” a subcomponent of her “Task Cycle.” Then, based on the outcomes of their joint effort, the learners were told to narrate the story again to each other. (This was intended to serve the function of the “Report” phase, another subcomponent of Willis’s Task Cycle.) Each lesson concluded with a brief review of the grammar and vocabulary that were relevant to the completion of the day’s task (Willis’s “Language Focus”). Each lesson took approximately sixty minutes. Yūta worked with nine different partners, one in each of the nine classroom sessions. As our main concern was task-type repetition rather than task repetition, the stories used in the nine lessons (and the test sessions before and after them) were different, as was shown in Table 2 above.

The regular university class that Kōtarō attended was assigned a series of decision-making tasks during the three week period. After minimal initial guidance, the students in the class were given a handout, which carried a list of five to seven persons or items with accompanying descriptions of each, and were asked to decide the order of priority with their partners according to the instructions given on the sheet. As an example, in the task entitled Which university do you want to attend? the learners were asked to imagine that they were going to study at a university located in the United States. On a printout was a list of six universities and the information including the number of students at each university, the number of teachers, the quality rating of the English department, local climate, average cost of living in the area, the tuition for the university, and other details. The students were first asked to come up with their own order of preference for the day’s topic in five to ten minutes and, then, they worked in pairs to negotiate the order of priority in order to reach a joint conclusion, or the “pair rank” within a time limit of ten minutes (Task Activity). After completing this, they cooperatively prepared for an effective presentation of their ideas (Planning). Finally, the teacher had the learners change partners and report the conclusion that they and their ex-partner had reached to the newly nominated partner. (Report). The entire process took about sixty minutes, including the final vocabulary and grammar review (Language Focus). Kōtarō worked on different task topics with different partners in nine lessons. Table 3 illustrates how the components of Yūta’s and Kōtarō’s classroom activities correspond to the stages of Willis (1996, 2004).

**Post-test:** After the three weeks of nine lessons in their respective classes, Yūta and Kōtarō were asked to come in turn to the post-test with their respective partners (the same students as in the pre-test). They performed an on-line narration, following the exact same procedures as in the pre-test. That is, Yūta and Kōtarō narrated half of the video story to their respective partners (and the partners narrated the remaining half.) The story used in the post-test was different from that in the pre-test in order for the effects of task-type repetition to be seen. However, the two stories were similar in their contents (i.e., a description of a day of a person), in order to ensure that they pose a similar level of cognitive demand to the learners. They were also asked to make an impromptu speech. The topic of the post-test speech was “Requests to the President of the University.”
## Table 3 Compatibility between the components of the lesson in the present study and Willis’s (1996, 2004) framework

<table>
<thead>
<tr>
<th>Willis’s Stages</th>
<th>Narration Class</th>
<th>Decision-Making Class</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-task</strong></td>
<td>Introduction to the task topic and vocabulary</td>
<td>Introduction to the task topic and vocabulary</td>
</tr>
<tr>
<td></td>
<td>First video viewing</td>
<td>Decision of the order (individual)</td>
</tr>
<tr>
<td><strong>Task Cycle</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Task Activity</td>
<td>On-Line narration</td>
<td>Decision of the order (in pair)</td>
</tr>
<tr>
<td>Planning</td>
<td>Confirmation of video content (10 minutes; in pair)</td>
<td>Confirmation of video content (10 minutes; in pair)</td>
</tr>
<tr>
<td>Report</td>
<td>(Planned) narration</td>
<td>Presentation of the conclusion</td>
</tr>
<tr>
<td><strong>Language Focus</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analysis</td>
<td>Grammar and vocabulary review</td>
<td>Grammar and vocabulary review</td>
</tr>
<tr>
<td>Practice</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Analysis

The recorded samples of the pre-test and post-test narration and the speeches by the two learners were carefully transcribed and were coded into AS-units, and were analyzed in terms of fluency, accuracy, and structural complexity of the language. The AS-unit is a unit for analysis of speech samples (hence “AS”) that was developed by Foster, Tonkyn and Wigglesworth (2000) in an attempt to overcome the problems with the c-unit analysis (cf., Pica, Holliday, Lewis & Morgenthaler, 1989) mainly by taking the intonational features of speech into consideration. The unit is defined by its advocates as “a single speaker’s utterance consisting of an independent clause or subclausal unit, together with any subordinate clause(s) associated with it.” (Foster, Tonkyn & Wigglesworth, 2000, p. 365)

The inherent nature of on-line narration requires several special considerations in the coding of the relevant samples. First, in on-line narration, the length of pauses is not a reliable measure of fluency because, as Skehan and Foster (1999) note, “[p]ausing in these circumstances is more likely to be a reflection of the need to keep pace with the story as it unfolds” (p. 106) rather than an indication of a breakdown in the speech formulation process. Instead, we focus on the “repair fluency” (Skehan, 1998b) for the narrative data. The repair fluency in this paper is defined as the rate of “genuine,” or legitimate components of sentences (i.e., the parts other than disfluency) in one entire sample as evaluated by the total number of syllables. Thus:

\[
\text{(1) The repair fluency score} = \frac{\text{Number of legitimate syllables}}{\text{Total number of syllables}} \times 100
\]

The higher the score, the more fluent the learner is estimated to be on this measure. In evaluating the speech performance, in which there arises no need for the learners to wait for the scene to change, we can take temporal variables into consideration of fluency. Therefore, in addition to the repair dimension, we calculated the time spent on producing genuine sentence elements (i.e., excluding disfluencies and filled and unfilled pauses) per ten seconds in each sample in order to obtain the “breakdown fluency” (Skehan, 1998b) score. Thus:

\[
\text{(2) The breakdown fluency score} = \frac{\text{Time spent on producing legitimate components of sentence}}{\text{Total speaking time}} \times 10
\]

The higher the score, the more fluent the learner would be considered to be on this breakdown fluency.
Accuracy is, as in a majority of studies, operationalized in this study as the percentage of the number of error-free clauses to the total number of clauses in one sample. Thus:

\[
(3) \text{The accuracy score} = \frac{\text{Number of error-free clauses}}{\text{Total number of clauses}} \times 100
\]

As our primary concern in this paper is grammatical accuracy in oral performance, inappropriate lexical choice (e.g., *washing his teeth* for *brushing his teeth*) were not counted as errors. Utterances comprised only of one major word (e.g., *Oh, kitchen!* in the context of a man coming down to the kitchen) were excluded from the measure of accuracy in the present study because their inclusion would result in a false rise of the accuracy rate for a particular sample. Thus, our analysis in this dimension is commensurate with the level 2 of AS-unit analysis (Foster, Tonkyn & Wigglesworth, 2000). We were not concerned with the tense of verbs when considering the grammaticality of the clause in the narrative performance, for it was both possible to report the events on the screen as happening in real time and as having happened in the past.

Finally, the complexity of learner language was evaluated in terms of a subordination ratio, i.e., the mean number of clauses per an AS-unit. Thus:

\[
(4) \text{The complexity score} = \frac{\text{Number of clauses}}{\text{Number of AS-units}}
\]

As in the calculation of accuracy, the AS-units comprised of only one word were not counted. Lexical complexity as evaluated by such measures as type-token ratio was not considered in this study because it is likely that, in the narration task, the individual story that the learner describes largely determines the range and frequency of lexical items to be used, making any measure of lexical complexity inappropriate for the evaluation of learner’s gain/loss between the two tests in which the learners narrated different stories. The type-token ratio may also be affected by the speech topic in speech performance. Thus, it is not very illuminating to look at changes in the vocabulary use of learners between the two speeches in which the learners were given different topics. In fact, a preliminary analysis did not find any notable pattern of changes across learners or the two tests in terms of the lexical type-token ratio.

The authors mutually checked their codings, and, after extensive discussion, reached an agreement on every detail. External native English-speaking informants were also consulted regarding the grammaticality of several clauses in the samples. The calculations were also mutually checked.

**Results**

Tables 4 through 7 and Figures 1 through 7 show the results of the pre- and post-test narration and speech by the two learners:

*Fluency*: Yūta stays at almost the same level in regard to his narrative repair fluency between the pre- and post-tests (0.03% gain), while Kōtarō loses 2.49% (Table 4 and Figure 1). In the speech post-test, while Kōtarō stays at about the same level as in the pre-test (0.44% loss), Yūta gains a sizeable 9.17% (Table 4 and Figure 2). As for the breakdown fluency in the speech performance, both learners score more or less higher in the post-test than in the pre-test (Table 5 and Figure 3). The percentile gains for Yūta and Kōtarō in this regard are 5.46% and 6.67% respectively.
Table 4 Development of repair fluency (percentage of valid sentence components in terms of the number of syllables)

<table>
<thead>
<tr>
<th>Test</th>
<th>Participant</th>
<th>Pre-Test</th>
<th>Post-Test</th>
<th>Gain</th>
<th>Percentile Gain (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narration</td>
<td>Yūta</td>
<td>92.08</td>
<td>92.11</td>
<td>+0.03</td>
<td>+0.03</td>
</tr>
<tr>
<td></td>
<td>Kōtarō</td>
<td>79.93</td>
<td>77.94</td>
<td>-1.99</td>
<td>-2.49</td>
</tr>
<tr>
<td>Speech</td>
<td>Yūta</td>
<td>86.14</td>
<td>94.04</td>
<td>+7.90</td>
<td>+9.17</td>
</tr>
<tr>
<td></td>
<td>Kōtarō</td>
<td>75.59</td>
<td>75.26</td>
<td>-0.33</td>
<td>-0.44</td>
</tr>
</tbody>
</table>

Table 5 Development of breakdown fluency (percentage of valid sentence components in terms of speech time)

<table>
<thead>
<tr>
<th>Test</th>
<th>Participant</th>
<th>Pre-Test</th>
<th>Post-Test</th>
<th>Gain</th>
<th>Percentile Gain (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech</td>
<td>Yūta</td>
<td>8.06</td>
<td>8.5</td>
<td>+0.05</td>
<td>+4.42</td>
</tr>
<tr>
<td></td>
<td>Kōtarō</td>
<td>7.05</td>
<td>7.52</td>
<td>+0.47</td>
<td>+6.67</td>
</tr>
</tbody>
</table>

Table 6 Development of accuracy (percentage of error-free clauses)

<table>
<thead>
<tr>
<th>Test</th>
<th>Participant</th>
<th>Pre-Test</th>
<th>Post-Test</th>
<th>Gain</th>
<th>Percentile Gain (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narration</td>
<td>Yūta</td>
<td>53.85</td>
<td>62.21</td>
<td>+8.36</td>
<td>+15.52</td>
</tr>
<tr>
<td></td>
<td>Kōtarō</td>
<td>66.67</td>
<td>48.84</td>
<td>-17.83</td>
<td>-26.74</td>
</tr>
<tr>
<td>Speech</td>
<td>Yūta</td>
<td>57.74</td>
<td>60.87</td>
<td>+3.13</td>
<td>+5.42</td>
</tr>
<tr>
<td></td>
<td>Kōtarō</td>
<td>70.83</td>
<td>60.00</td>
<td>-10.83</td>
<td>-15.29</td>
</tr>
</tbody>
</table>

Table 7 Development of complexity (number of clauses per AS-unit)

<table>
<thead>
<tr>
<th>Test</th>
<th>Participant</th>
<th>Pre-Test</th>
<th>Post-Test</th>
<th>Gain</th>
<th>Percentile Gain (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narration</td>
<td>Yūta</td>
<td>1.13</td>
<td>1.18</td>
<td>+0.05</td>
<td>+4.42</td>
</tr>
<tr>
<td></td>
<td>Kōtarō</td>
<td>1.42</td>
<td>1.54</td>
<td>+0.12</td>
<td>+8.45</td>
</tr>
<tr>
<td>Speech</td>
<td>Yūta</td>
<td>1.62</td>
<td>1.15</td>
<td>-0.47</td>
<td>-29.01</td>
</tr>
<tr>
<td></td>
<td>Kōtarō</td>
<td>2.00</td>
<td>2.14</td>
<td>+0.14</td>
<td>+7.00</td>
</tr>
</tbody>
</table>

**Accuracy:** Yūta and Kōtarō pattern differently in their gain/loss in the percentage of error-free clauses in the narrative tests (Table 6 and Figure 4). Kōtarō’s score has dropped sharply from the pre-test to the post-test (26.74% loss), whereas Yūta’s score has risen considerably (15.52% gain). Kōtarō’s accuracy level dropped in the speech post-test, too (15.29% loss). Yūta scores slightly more, though not to a great extent, in the post-test speech (1.45% gain) (Table 6 and Figure 5)

**Complexity:** The mean number of clauses per AS-unit in their narration has increased both for Yūta and Kōtarō (Table 7 and Figure 6). Yūta’s gain (4.42%) is modest compared with Kōtarō’s (8.45%). In the speech, the scores of the two learners pattern differently (Table 7 and Figure 7). Yūta’s post-test complexity
Figure 1 Repair fluency in pre- and post-test narration

level is much lower than in the pre-test (29.01% loss), while Kōtarō scores higher in the post-test (7.00% gain).

DISCUSSION
In this section, changes in fluency, accuracy, and complexity of the language produced by Yūta and Kōtarō will be discussed, based on the data presented in the previous section. Further, we will make brief comments on two topics of relevance: the individual differences of the learner that may affect the effects of task-type repetition, and the accuracy-complexity trade-off.

Fluency
As shown in Figure 1, even after a nine-time repetition of the narration task, the disfluency rate in Yūta’s language was almost the same as in the pre-test. However, somewhat mysteriously, he remarkably reduced the rate of disfluency in his post-test speech performance, as Figure 2 shows. A possible interpretation of this may be that though his fluency in this dimension had been potentially developed, the task of on-line narration was so taxing that the acquired potential was latent in the post-test narrative task performance and manifested itself only in a possibly less demanding speech task. This may suggest a complex nature of the effects of task-type repetition. Kōtarō’s pre-test and post-test scores, on the other hand, are almost comparable both in his narration and speech, suggesting that repetitive use of the decision-making task did not assist him in reducing disfluency.

As for breakdown fluency in the speech performance, Kōtarō as well as Yūta exhibit some gain in their scores from the pre-test to the post-test. In reference to his failure in reducing disfluency, Kōtarō’s gain is considered to be mainly because of the reduction of the length of pause (be it filled or unfilled). That is, his gain in breakdown fluency was mainly a result of his reducing the time for, not the amount of, fragments. Of course, this needs to be confirmed with a larger population, but the results from our two learners suggest that narrative task-type repetition is effective in reducing the amount of disfluency, which is apparent only in a task other than on-line narration, while repetitive use of the decision-making task may help the learner decrease the length of time for “stop and think.”
There is a clear and sharp contrast in the pattern of gain/loss in the accuracy scores between the two learners in their narrative task performance (Figure 4). Yūta gained considerably from the pre-test to the post-test (15.52% gain), while Kōtarō suffered a substantial loss in his score (26.74% loss), leading to Yūta outperforming Kōtarō in the post-test. Note that this is the only instance of the learner with an inferior starting score outstripping the learner with an initial higher score in the post-test. In reference to his complexity scores (Figures 6 and 7), the key to understanding Kōtarō’s loss of accuracy may lie in an interaction between the accuracy and complexity dimensions. It is possible that the repeated encounters with the decision-making task had vested Kōtarō with the strategy of prioritizing complexity over accuracy, and that his attentional resources have been exhausted in gaining complexity. We will come back to the issue of accuracy-complexity trade-off in a later section.

In the speech data, Yūta’s gain is not as remarkable as in his narrative task performance. One reason for this may be his relative unfamiliarity with the speech task. However, there may be another reason for his relatively small gain in the speech test: the topic of the post-test speech. As explained above, the participants reported their strong points in the pre-test speech, while in the post-test speech they were asked to articulate their requests to the university president. In designing the experiment, the two topics were intended to be of equivalent cognitive load, both soliciting the learners’ personal views on issues of immediate relevance (either to themselves or to their circumstances). On afterthought, however, the post-test speech topic...
may have posed a greater challenge to the learners. To manage the post-test topic, it may have been necessary for them to think somewhat seriously about societal problems around the university and/or to make sociolinguistic considerations for an imaginative situation of talking to the university president. In consideration of this as well as the considerable decline in Kötarō’s speech scores, it may be able to say that Yūta did a good job.  

**Complexity**

The results show that both Yūta and Kötarō gained in their complexity scores in the narrative performance (Figure 6). Kötarō’s gain was greater than Yūta’s. This is remarkable when we consider that Kötarō did not practice the narration task. Arguably, this may be due to the ability developed by the repetitive use of the decision-making task being carried over to the unfamiliar narrative task performance. If this is the case, the contribution of the repetitive use of the decision-making task to complexity development should be considerable. The decision-making task creates the need for reasoning, requiring the learners to state their own points of view while countering or accepting the ideas expressed by their interlocutors. It may be reasonable that complex language use required in these processes makes considerable contributions to the complexity development in the long run. By contrast, it seems that the narrative task repetition did not stretch Yūta’s potential to a large extent in the complexity dimension. In the narration, a sequence of simple sentences often suffices to describe somebody doing something on a screen and, thus, there arises little need for the learner to elaborate with subordinate clauses. Moreover, even if the learner should want to be more complex, he or she may well not be able to do so under the severe time pressure created by the on-line nature of the task. “Simple and accurate” may be the strategy, be it conscious or subconscious, that Yūta has developed in order to cope with the demand of on-line narration.

The potential of the decision-making task in developing complexity is also suggested by the changes in Kötarō’s complexity scores between the pre- and post-test speeches, which was also not a task that he had repeated (Figure 7). Kötarō’s 7% gain can be considered to be substantial, especially when we call to mind the above-mentioned demanding nature of the post-test speech. Yūta’s considerable regression, on the other hand, may be attributable partly to the relatively small impact that the repetitive use of the narration task had in this regard, and partly to the difficulty he had in coping with the post-test speech topic. Taking all things into consideration, the results of the present study suggest the possibility that the extensive use of the decision-making task promotes the ability to produce complex language, while the effect of narrative task repetition in this regard is negligible. As argued above, it may be possible to attribute the accuracy loss for Kötarō in his narration to his gain in this dimension.

**Individual differences of the learner and the effect of task-type repetition**

It is evident in Figures 1 through 7 that at the time of the pre-test, Yūta and Kötarō were different in regard to the aspects of language that they prioritize. On the one hand, it is shown in the fluency part of the results (Figures 1-3) that Kötarō spoke more hesitantly than Yūta both in his narrative as well as speech pre-test performance, with his utterances being embroidered by many fragments that were immediately replaced or reformulated. On the other hand, Kötarō’s pre-test utterances were generally more accurate and more complex than Yūta’s (Figures 4-7). These facts mean that the two learners had notably different L2 speech styles: Kötarō speaks wastefully, staggering along, with many disfluencies that he immediately reformulates into correct, elaborate sentences, while Yūta speaks with less effort, but with many flaws. This contrast calls for a discussion on the individual differences of learners as a factor potentially affecting the results of task-type repetition, and a discussion on “learner-task matching.” That is, it is possible that at least some
of the results reported and discussed above were a function of particular learner characteristics. What if Kōtarō had repeated the narration task and Yūta had repeated the decision-making task? Unfortunately, the present study cannot provide an answer to the question. A suggested direction for future research would be an investigation into the three-way interaction of task-type, individual differences, and the aspects of learner language affected by task-type repetition.

The accuracy-complexity trade-off

The final topic of the discussion in this paper is the interaction between accuracy and complexity. According to Skehan (e.g., 1998, 2001; also see Skehan & Foster, 2001), accurate use of language is a consequence of the learner’s “being conservative,” not daring to take risks to use the structures and vocabulary that are on the cutting-edge of their current knowledge. Complexity, on the other hand, is supposedly a result of the learner exploring and elaborating on his or her language in order to express his or her ideas. Referring to the L2 as well as psychology literature (e.g., VanPatten, 1990), Skehan maintains that accuracy and complexity compete for attentional space in the L2 production process and that giving priority to either one of them inevitably deteriorates the other. The results of the studies by Skehan and others generally support this position. However, other researchers like Robinson (e.g., 2001), who support the multiple resource view of attention, are not content with this theorizing. Referring to the claims by Neumann (1996) and others, Robinson argues that cognitively complex tasks require the learner’s attention to the extent that accuracy and complexity of language are increased.

Skehan and Robinson are both arguing about accuracy-complexity trade-off in L2 use rather than L2 development, which is the concern of the present study. From Yūta’s and Kōtarō’s data, repetition of a particular type of task seems to be effective in the development of either accuracy or complexity, hence suggesting the existence of trade-off in developmental terms. On the one hand, Yūta, who had repeated the narration tasks improved in the accuracy but not the complexity scores. On the other hand, Kōtarō, who had repeated the decision-making tasks gained only in complexity. This may suggest that the impact of the task-type repetition is selective, with different types of tasks having their unique roles in affecting the development of specific aspects of the production skills.

It is possible that the trade-off is relative to the proficiency of the learner. In L2 speech production, it may be natural that less proficient learners exhaust their attentional resources in pursuit of the quality of either one of accuracy or complexity, while more advanced learners are resourceful enough to pay due regard to both.17 Developmentally, too, more advanced learners (at least more advanced than the two in the present study) may find ways to manage the conflicting demands that a task poses to them, and may progress along both dimensions.18 This should be explored in future research.19

SUMMARY AND CONCLUDING REMARKS

We have reported gains (and in some cases losses) in the scores of the two learners who had repeated a specific type of task for a three-week period in terms of the fluency, accuracy, and complexity of their language. It has been shown that Yūta’s and Kōtarō’s L2 use did change after repeating particular types of tasks. One interpretation of the fluency-related results in the present study may be that narrative task-type repetition helped Yūta develop the ability to speak with less disfluency, though the potential was not discernible in the demanding on-line narration task. The narrative task-type repetition also seemed to be useful in reducing the pause length and the time spent on disfluency in Yūta’s language. For Kōtarō, on the other hand, repetitive use of the decision-making task seemed to contribute to the reduction of the length of
time for pauses and fragments but not to the reduction of the amount of disfluency. As for accuracy and complexity, the obtained data showed that the narrative task-type repetition was effective in promoting the accurate use of language, while complexity increased rather by repetitive use of the decision-making task, hence the accuracy-complexity trade-off. In sum, the results from the two learners reported in the present study point to the possibility that there is a task-type effect, which the previous studies (e.g., Bygate, 2001; Gass, Mackey, Fernandez & Alvarez-Torres, 1999) failed to identify. Also, a possibility has been suggested that the aspects of productive skills developed in a repeated encounter with a particular type of task are transferable to a new task.

However, it is important to keep in mind that this study is exploratory in nature, looking only at the development of one learner who was engaged in narrative task-type repetition, as compared with another who was administered decision-making tasks. As noted above, larger-scale studies will have to see whether the present results are idiosyncratic to our learners or represent a general truth of L2 development. In the future studies, inquiry may also be made into how the effects of a certain task type interact with such factors as the proficiency levels and other learner attributes (e.g., analytic vs. holistic).

ACKNOWLEDGEMENTS
The authors would like to thank Antoine Stebbins for his insightful comments on earlier drafts of this paper as well as for his editorial advice.

NOTES
1 In this paper, we use the term “second language” as a general term to refer to languages learned after the mother tongue. Where a distinction is necessary between second and foreign languages, it will be so noted.
2 Of course, there has been task-related research done in other academic contexts than those outlined here. For example, recent years have witnessed the emergence of task-related studies that have sociocultural underpinnings. See Chapter 7 of Ellis (2003) and chapters in Lantolf and Thorne (2006) for reviews of the L2 studies carried out within this theoretical framework.
3 Some studies investigated how a particular type of task influenced the use of specific linguistic features such as articles (Tarone & Parrish, 1988), but, recently, researchers have favoured more general measures for the evaluation of learner language like overall fluency, accuracy, and complexity scores. See Ellis (2005) for a discussion of the limitations and dangers in examining specific features of learner language.
4 This is not to say that the performance studies have no developmental implications. Lantolf (2005) points out that the very fact that a learner can use a certain feature, be it with some aids and/or only in some conditions, means that “it is within their ZPD [zone of proximal development]” (p. 345).
5 Studies investigating the effects of task repetition that are not be referred to in this paper include Yule, Powers and McDonald (1992) and Németh and Kormos (2001). The former discusses the effects of task-type repetition in terms of the development of communicative efficiency and the latter from the pragmatic perspective.
6 Other factors that Bygate (2001) suspects may have affected his results are the proficiency levels of the learners and their familiarity especially with the interview type of task (p. 43).
7 In this context, it is worthwhile to note that Ellis (2005) makes an urgent request that “there is a clear need
for shift from laboratorylike studies to the careful evaluation of both its implementational and the learning outcomes in real classrooms” (p. 725).

8 This may sound odd, given, as noted, that the participants are advanced beginners. However, one of the notable characteristics of typical Japanese learners of English is a discrepancy between rich knowledge of English grammar and limited communication skills, which may be attributable to the way they have learned English in secondary education.

9 In fact, Yūta’s pre-test narration was demarcated by some extremely long pauses, which were likely signs of his waiting for the events worth reporting to come up. Disfluency is composed of false starts, repetitions, reformulations, and replacements. See Foster and Skehan (1999) for their definitions.

10 Counting errors in lexical choice does not change the general picture for accuracy. For reference, the accuracy scores for Yūta would be 46.15 and 60.61 in the narrative pre- and post-tests, and 53.19 and 56.52 on the speech pre- and post-tests when lexical as well as grammatical errors were taken into account; Kōtarō’s scores would be 55.56 and 37.21 on the narrative pre- and post-tests, and 70.83 and 60.00 in the speech pre- and post-tests.

11 In fact, the length of filled and unfilled pauses per ten minutes in Kōtarō’s speech was 0.81 seconds in the pre-test and 0.40 seconds in the post-test. On the other hand, pause length in Yūta’s speech was almost constant across the two tests: 1.18 seconds per ten seconds in the pre-test and 1.14 seconds per 10 seconds in the post-test.

12 One may argue that Yūta’s accuracy gain is simply due to the fact that he had more potential space for progress created by his initial lower level of achievement. However, this is not likely to be the case. Kōtarō’s 66.67 percent accuracy in the pre-test narration can by no means be regarded as a “ceiling.” Moreover, the interpretation cannot explain the results for complexity.

13 One question that may arise concerning the laborious nature of the post-test speech is why it affected less on fluency than on the accuracy and complexity of learner production. We leave this point open for future research.

14 Some increase in Kōtarō’s disfluency rate in the narrative as well as speech post-test performances (Figures 1 and 2) may also be interpreted as the cost he had to pay to gain complexity. This, however, may be amenable to other interpretations. For example, it may be a manifestation of his overall negligence in fluency (note his relatively low starting scores in fluency).

15 Skehan and Foster (2001) make some passing notes on the relevance of individual learner differences to L2 task performance in their discussion of the tension existing among fluency, accuracy and complexity (pp. 190-191). But they do not elaborate on the issue.

16 This is similar to, but is different from, one of the points that Skehan and Foster (2001) make. Drawing a qualitative distinction line between native speakers and the L2 learners, they argue that native speakers with fully developed, automatized linguistic ability are free from the accuracy-complexity trade-off, while L2 learners have to prioritize one, even when they are at a fairly advanced level. There should be no reason, however, for denying that even L2 learners can control both dimensions as they approach the near-native level (if the level is attained at all).

17 Another potential variable is inherent task difficulty. It is highly probable that tasks with greater cognitive demand tax the learner’s attentional resource, and there thus emerges the trade-off.

18 It may be possible to interpret the results in a different way. That is, development may proceed “cyclically,” with the task-type repetition influencing accuracy and complexity at different stages of
learner development. It might be that Yūta was at the stage of development where his accuracy is sensitive to any task experience at the time of the study, while Kōtarō was at the stage where his complexity was ready for a leap. However, this is not a plausible scenario because, as noted, the two learners were comparable in their TOEFL scores (around 480, PBT).

REFERENCES


