

# Reflection in second language writing: A longitudinal study of task repetition from a Complexity Theory perspective

Kyoko BABA

## **Abstract**

This study longitudinally investigates the effects task repetition with reflection had on the development of the L2 writing of EFL university students by looking at both group trends and individual patterns of development from a Complexity Theory perspective. Two classes of students (46 students in total) were engaged in a timed writing task 30 times over one year. One class (the Extended Reflection Class) wrote extended reflective comments immediately after writing every composition, while the other class (the Short Reflection Class) only wrote short personal comments. It was revealed that the most notable difference was found in the development of fluency between the two classes. The text length of the Extended Reflection Class became distinctively longer, while that of the Short Reflection Class remained largely the same. Case studies on three students who showed a different developmental pattern demonstrated that the frequency and timing of phase shifts differed between students, and the nature of each phase shift varied even within the same student. Nonetheless, those who experienced phase shift came to set goals for their next writing task and did some extra efforts even outside class when they realized that the task would be repeated.

## **Introduction**

Most things are repeated in second-language learning classrooms, and teachers may intuitively know that students learn a second language (L2) through repetition. Research on task based language teaching (TBLT) has shown that L2 speakers come to speak fluently by repeating the same task (i.e., task repetition). However, the attention of TBLT researchers has largely been directed toward relatively short-term changes in fluency. Their basic argument is that the cognitive capacity of learners is limited when they produce connected speech, so that they cannot focus on content and form simultaneously. They first pay attention to the content of their speech (this is sometimes called planning or rehearsal), and by performing the same task, they come to focus on form, which results in progress in speaking fluency (Bygate & Samuda, 2005). Then, what if learners are engaged in a writing, rather than a speaking task? Or, what if they repeat the same task over a longer period? These were the questions we

investigated in our previous studies (Baba & Nitta, 2010, in press), and continue to ask in the present one.

Baba and Nitta (in press) found that the simple repetition of a timed-writing task without corrective feedback over 30 weeks (one academic year) did have an impact on the changes in L2 learners' writing, but not in a predicted way. We predicted that L2 writers would improve fluency in their writing by repeating the same type of task, as indicated by TBLT research. However, their writing fluency (counted by the number of words per essay) remained stable as a whole. Conversely, we saw a significant change in the grammatical complexity of their writing.

The reason for the significant change was not totally clear, but we suggested that it might be attributed to the students' consciousness and awareness of their own writing. In the interview one student said that her consciousness and awareness of her own writing had changed. According to her, she tried to write longer at first, but as she repeated the task, she began to concentrate on grammatical structures, and attempted to use the structures that she had recently learned in class. Such attempts could be found in her later compositions. Based on this finding, we speculated that "once basic writers become conscious of their own writing and set some goals for writing improvement, the initial step for them to try might have been to concentrate on their use of grammar in writing" (Baba & Nitta, in press). In other words, our previous study has pointed to the significance of writers' reflective consciousness in their development of L2 writing. We also found that learners demonstrated various developmental patterns in different aspects of their writing (Baba & Nitta, 2010). We can be fairly certain that individual students vary in how they reflect on their own writing and also how their reflections affect their development. It is therefore important to monitor not only dynamic changes in a class as a whole, but also the unique developmental paths that each student may follow. Thus, the purpose of the present study was to longitudinally examine what effects task repetition with reflection had on the development of the L2 writing of EFL university students by looking at both group trends and individual patterns of development.

### ***Researching Change in Complexity Theory***

There has been a general assumption in second language education (SLE) research that change or development in L2 learners occurs linearly. L2 development has often been investigated with inferential statistics because of this underlying assumption. For instance, the effects of some interventions have been examined by conducting pre- and post-tests, and seeking out significant differences between the two. However, such an assumption is being challenged by a new group of researchers who maintain a perspective of complex dynamic systems (e.g., The "Five Graces Group," 2009). This perspective

is based on Complexity Theory (it may be referred to in a variety of ways, such as Chaos Theory, Complex Systems Theory and Dynamic Systems Theory, but I will use the term "Complexity Theory" after this), which aims for "the interdisciplinary understanding of reality as composed of complex open systems with emergent properties and transformational potential" (Byrne, 2005, p. 97). According to the theory, a research object is recognized as a system that consists of and relates to other systems. The system is always open to change, and transforms dynamically.

Various characteristics of a system have been explicated elsewhere (e.g., Larsen-Freeman & Cameron, 2008, The "Five Graces Group," 2009), but there are two features of a system that are especially pertinent to the present classroom study: *heterogeneity of agents in a system* and *non-linearity of a system*. These concepts seem to make more sense in real classrooms than the traditional positivistic or laboratory experiments. For instance, when researchers who assume linear development investigate the effect of some instruction, they let a group of students write a composition before and after the instruction. They then compare the compositions with statistics to see if the group underwent significant changes. However, teachers who are actually teaching an ESL class might not only want to know about changes in the class as a whole (let us say, a system) but also those in individual students (heterogeneity of agents). Hence, it is necessary to look both at the class and its agents (which also work as systems).

In addition, to assess whether some instruction has been effective or not for specific students at the end of a course would have few pedagogical implications for focal students, because it would be too late for them to benefit from the findings at that point. Therefore, teachers usually observe their students' performance at shorter intervals (such as every week or month), and analyze how the students have developed and will be able to develop further. Such observations inevitably are qualitative and longitudinal in nature. Changes in students' writing may not always take place in a lock-step manner (non-linearity of a system). A student may write a good composition one day, and an inferior one the next. Occasionally, s/he might use subjunctive mood correctly, but may fail to do so at other times. Complexity Theory throws light on this *change* itself, as Dörnyei (2009) states:

Social scientists tend to focus on well-defined and generalizable *variables* to describe the social world around them. A complexity/dynamic systems approach needs to shift the emphasis from this variable-centered, reductionist practice to studying how systems *change* in time. (p. 242)

This focus on change has a strong connection to what is happening in classrooms, where teachers (and occasionally students themselves) pay more attention to how the

students' L2 performance has changed and will change than to what any technically operationalized variable in their performance was like.

Changes in a system can be minor or major. Returning to the example above, the students may avoid using subjunctive mood at first, or use it with errors. However, at some point in their development, they may fully master the ability to use it. Students may experience "real progress" at a certain moment, and hardly revert to the previous state. This phenomena is called *phase shift*, *phase transition*, or *developmental jump* (I will use the term *phase shift* after this). During phase shift, a system undergoes radical changes, which leads it to a new mode or structure.

Studies that have been designed to investigate such changes should inevitably be longitudinal. However, second language educational research in general has lamented the paucity of longitudinal studies (Ortega & Byrnes, 2008; Ortega & Ibarra-Shea, 2005). In addition, not all kinds of so-called longitudinal designs have been suitable to investigate change. Reviewing methods of measuring change, Willett (1994) distinguished two research designs: two-wave and multi-wave. With the two-wave research design, researchers compare participants' performance "before and after" or "pre and post" treatments. This design does not allow researchers to examine how performance has changed during treatment. In L2 writing research, for instance, change in writing has often been investigated by assessing the difference between a pre- and a post-test (or writing before and after some treatment) (e.g., Ishikawa, 1995; Shaw & Liu, 1998). This design may reveal whether there treatment had an effect, but it will not tell *how* this eventuated.

In contrast, with multi-wave design, "each person is measured repeatedly over extended periods of time (Willett, 1994, p. 671)." This design was rarely used in SLE research before the advent of Complexity Theory mainly due to the lack of a theoretical framework that could support it and of research techniques that could enable it to be implemented (though there were some exceptions such as Mellow, Reeder, & Forster, 1996). However, we have been seeing an increase in such multi-wave studies in recent years, which have revealed some aspects of L2 development that have not previously been investigated. For example, Larsen-Freeman (2006) collected oral and written data four times over six months from five women living in an English-speaking country whose native language was Chinese in a seminal work using Complexity Theory. She found that the averages of four linguistic aspects (accuracy, fluency, vocabulary complexity, grammatical complexity) in their speech and writing ostensibly developed smoothly, but the developmental paths of the individuals varied widely in terms of which linguistic aspect developed first and their rates of development.

Collecting more writing samples over a longer period of time, Verspoor, Lowie, and van Dijk (2008) and Spoelman and Verspoor (2010) clearly described phase shift in

language development. Verspoor et al. (2008) collected 18 writing samples from one Dutch learner of English over three years, and found that there was a three-stage developmental process in changes of average word length in his writing. First, there was a stable trough period (average word length remained short), and then a flux period (the figure changed dramatically each time), and last a stable peak period (average word length stabilized at a higher level). The researchers claimed that the flux period demonstrated phase shift.

Similarly, Spoelman and Verspoor (2010) collected 54 writing samples over three years from a Dutch learner of Finnish as a foreign language, where phase shift was observed in the development of the number of noun phrases that contained more than three words. By means of a moving progmax-regmin graph method (see Methods section), the researchers found the value suddenly *jumped* later in the observation period (between the 44<sup>th</sup> and the 45<sup>th</sup> texts).

These observations on phase shift suggest that the learners may have experienced some qualitative and discontinuous change in their L2 development. These qualitative changes must reflect something important in the development of language learners. It is necessary to collect some qualitative data to scrutinize such changes in L2 writing. However, the studies thus far have not looked at the qualitative aspects of change in a system (however, see Larsen-Freeman, 2006). Thus, we collected not only writing samples repeatedly over one academic year in the present study but also the writers' accounts of reflective thinking about their own writing.

### ***Focus on Repetition***

Unlike TBLT research, L2 writing research has paid scant attention to task repetition. The two research areas that are closest to task repetition are research on the effect of corrective feedback and journal writing. In studies on corrective feedback, researchers have usually let participants write various compositions and return them with different types of feedback. Most of these have compared the effect of at least two types of feedback: correction of grammatical errors and content-focused comments. They have also compared these with no feedback. The main focus of these studies has, of course, been whether teachers' corrective feedback is effective or not (for reviews, see, e.g., Ferris, 2010; Guenette, 2007; Truscott, 2007), but some studies have found that even without corrective feedback, L2 writers made progress in some aspects of their writing. For example, in one of Chandler's (2003) studies, ESL learners wrote five essays over 14 weeks. One group of learners received corrective feedback, while the other did not. Chandler compared the first and the last essays, and found that the former group with corrective feedback had significantly reduced error rates in their writing, while the latter group had not. However, both groups improved fluency (based

on the reported time for writing). In fact, the latter group seemed to have enhanced their fluency to a greater extent than the group with corrective feedback.

Kepner (1991) found that the types of feedback did not matter when message related comments and surface error correction were compared. In Kepner's study, ESL university students wrote a journal biweekly during one semester, and received one type of feedback. As a result, two groups of students who received different types of feedback did not differ in the number of surface-level errors they made. In addition, Kepner also assessed the ideational quality of journals, and found that the group who received message related comments produced more propositions that contained higher level thinking (e.g., comparison and evaluation). It is necessary to interpret Kepner's finding with caution, because she only compared the two groups' compositions written each week toward the end of the semester, and did not examine the progress within each group; nor did she note the initial difference in the two groups. Yet, her study implies that other factors might have influenced students' writing more than the type of feedback did.

Duppenthaler (2004) also investigated L2 university students' journal writing with different types of feedback, but he employed a wider variety of indices of writing: the number of words, the number of clauses, the number of error-free clauses, the Flesch-Kincaid Readability Index, and four vocabulary measures. The types of feedback that three groups of students in his study received were: content-focused feedback, positive comments, and error-focused feedback. The students wrote 22 journals as a home assignment over one academic year. Duppenthaler compared the first four and the last four journals written by the three groups. He found that there was little difference between the three groups, and concluded that none of the groups demonstrated noteworthy changes in their journals. It is regrettable that he compared the first four and the last four journals separately, and did not examine the change within each group. Still, his study suggested that teacher's feedback on error correction may matter relatively little in journal writing.

Citing Duppenthaler's (2004) study, Casanave (2004) claimed that whether the teacher corrected errors or not was not as important as task repetition itself. Reflecting on her own teaching experience, she argued, "[m]y written corrections to some of the errors in students' journals seemed to have little effect on later journals, whereas the regular weekly practice did" (p. 94). Her argument is also based on her previous classroom research (Casanave, 1994), in which Japanese EFL students wrote journals once a week during the first two semester of an academic year and twice a week during the last semester. She compared three points in time (those of the first two, the middle two, and the last two journals) with a variety of measures, including the number of words per T-unit, the number of clauses per T-unit, and the length and rate

of error-free T-units. She summarized her findings: "In addition to write more fluently, some students wrote more accurately over time ...; some wrote on more mature topics; and some wrote with more detail, depth, and expressiveness" (Casanave, 2004, p. 74). Such changes occurred without corrections to grammatical errors.

Casanave's (1994) findings are largely supported by studies on journal writing by L2 students. Through repeated journal writing, students may develop the content, organization, vocabulary, overall quality of their writing (Liao & Wong, 2010), grammatical accuracy (Weissberg, 1998), and the use of grammatical morphemes (Peyton, 1990). Moreover, they may feel more confident and motivated in writing (Casanave, 1994; Duppenhaler, 2002; Liao & Wong, 2002). Personally communicating with their teacher may also empower students who have different life experiences and linguistic and cultural backgrounds (Peyton & Staton, 1993). Journal writing provides students with more opportunities to write freely with relatively less pressure and attempt new grammatical and linguistic features that they are working on.

The findings from the previous studies suggest that regardless of the types of feedback, repeating a similar task may be a crucial factor in L2 writers' changes. This supports the Complexity Theory perspective, as Larsen-Freeman (2009) puts it:

In fact, from a complexity theory standpoint, using a task more than once is what drives learning. When it comes to language learning, revising the same, or similar, territory again and again is essential. (p. 584)

Thus, the present study highlights task repetition itself, and explores how it might induce changes in systems both in L2 classes and in the individuals themselves.

### ***Reflections on Journal Writing***

Introspective methods such as think-aloud and stimulated recall interviews have been widely used in L2 writing research (Gass & Mackey, 2000). Such introspective methods are valuable tools that provide clues to the learner's internal processes and activities such as cognitive processes, communication strategies, intentions, and motivation. However, some researchers have recently advocated that verbalizing internal activities themselves offers learning opportunities to learners (Swain, 2006; Swain, Lapkin, Knouzi, Suzuki, & Brooks, 2009). They called this verbalization that was used as a learning tool *linguaging* and defined it as "the process of making meaning and shaping knowledge and experience through language" (Swain, 2006, p. 89). Linguaging can be both oral and written. Suzuki (in press) revealed that the written linguaging of revision processes facilitated the learners' successful revisions with improved accuracy.



Introspection as a learning tool has also attracted attention in journal writing research, but in a broader sense. Introspection in studies on languaging and those on self-explanation in cognitive psychology (e.g., Kastens, & Liben, 2007) merely means verbalizing what one is thinking or doing. However, the act of writing journals itself is often regarded as a reflection of the learner's life and classroom experience in journal writing research. The learner is expected to learn from such experience by re-interpreting various events through journals. Boud (2001) proposed that journal writing may reinforce reflective practice in the learner, and explained reflection as "a process of turning experience into learning, that is, a way of exploring experience in order to learn new things from it" (p. 10). He claimed that reflection can occur at three stages: *reflection in anticipation of events, reflection in the midst of action, and reflection after events*. First, learners imagine and anticipate future events through reflection and try to make the most of them by setting goals and preparing for them. Second, they become more aware of the situation in which the events are occurring, and undertake various actions more consciously through reflection (reflection at this stage seems similar to the verbalization above). Finally, they re-examine their experience of the events, re-confirm their feelings and emotions, and re-evaluate and re-interpret the meanings that the events have for them, sometimes by connecting the events with their past experiences. Unfortunately, SLE research thus far has mainly focused on reflection *in the midst of action*, but has scarcely investigated the effect of reflection *in anticipation of events or after events*.

The present study was carried out to employ reflective writing both as a research tool that would partly illuminate what was happening to learners during phase shift, and as a learning tool that would help writers pay attention to various aspects of writing and actively engage in writing tasks in such ways as Boud (2001) described. Students in one class in the study wrote extended and focused reflections by referring to Boud's three aspects of reflection, while those in the other class wrote short personal comments after they wrote their compositions. There were three research questions in this study:

- (1) How does student writing in two classes taught to EFL university students change in terms of fluency, lexical complexity, and grammatical complexity when they repeat the same writing task with or without extended and focused reflections over one year?
- (2) How can phase shift occur in individual writers? What are the differences between writers who experience phase shift and those who do not?
- (3) How are extended and focused reflections related to phase shift?



## Methods

### *Setting*

The data were collected from two classes in a course at a Japanese university over one academic year (30 weeks in total). This course was offered to first-year students as a compulsory program. The curriculum for the department required first-year students to intensively enhance four English skills, and the course was designed to be the basis of academic writing. The first-year students were assigned to a class according to their English proficiency at the beginning of the academic year, and each class of students took the same English skills classes. Therefore, they frequently saw and knew one another well. I was appointed as an adviser for the two classes (a kind of homeroom teacher), and met each student individually outside class. I was, therefore, fairly close to my students for a university instructor.

One class (called the Short Reflection Class, which was investigated in Baba & Nitta, 2010) preceded the other (called the Extended Reflection Class) by one year, but both classes were taught by me, and shared the same syllabus. I tried to teach them the same content in the same way. On the first day of the class, I told the students that I would use their compositions and reflective comments for research purposes, and asked them to sign a consent form (which I did not open until I had finished awarding final grades). At the beginning of each lesson, students wrote a composition within 10 minutes, and then their reflective comments (see below). Although they participated in the study on a voluntary basis, all of them had to engage in this task as part of fulfilling course requirements.

### *Participants*

All the students in the two classes belonged to the Department of English, and were majoring in English. Twenty-three students in the Short Reflection Class and 22 students in the Extended Reflection Class regularly attended the class until the end of the year, so the data for these 45 students were analyzed in the study. The students' mother tongue was Japanese, and they shared a similar educational background (they had had six years of formal English education before entering university). Their English proficiency level was low-intermediate. All the students had to take a TOEIC test in May and January (at the beginning and end of the academic year), and the average score for the two classes was 390.43 for the Short Reflection Class and 373.64 for the Extended Reflection Class in May, while it was 419.78 for the Short Reflection Class and 403.18 for the Extended Reflection Class in January. There were no significant differences between the two classes either in May ( $t(31.40)=-.97, p=.34$ ), or in January ( $t(43)=.99, p=.33$ ).

No students had ever engaged in such a timed-writing task as that used in the study

before the class either in English or in Japanese. In addition, the students had little experience writing in English before entering university, and they had little chance of writing in English outside the classroom even at university. These were advantageous conditions for observing changes in their writing through classroom writing activities, for there were few intervening factors.

### ***Task and Procedure***

The task used in this study was timed writing, in which students wrote a composition on a chosen topic for 10 minutes. The students were given a topic list with three different topics each time. For example, students could choose from: (1) Name your favorite game or sport and explain why you find it enjoyable; (2) What foreign country would you like to visit, and why would you like to go there?; (3) What is your purpose in studying at this university? We tried to select topics that were personal and required little background knowledge. All the topics were written both in Japanese and English. The same list of three topics was used for two weeks, and the students were told to write on the same chosen topic twice. Then, another new topic list was set for them the following week. The same topics were set for both classes in the same order. The aim of offering three topic alternatives was to compensate for differences in the students' individual experiences and preferences.

Immediately after they had written their compositions, students were asked to rate the difficulty of the topic so that we could check what influence the difficulty had on their writing. The average topic difficulties in the 30 weeks turned out to be 3.62 for both classes, and there were no differences ( $t(57)=-.09, p=.93$ ).

The students were also asked to write reflective comments on their writing in Japanese immediately after they had written their compositions. The students in the Short Reflection Class wrote a few sentences in a box entitled "personal comments." The content to be written was not specified. It usually took a few minutes for them to write the comments. On the other hand, those in the Extended Reflection Class filled out an extended reflection sheet. There were three sections on this sheet: (1) comments on today's composition, (2) comments in comparison with previous compositions, and (3) goals for future writing. By filling out these sections, the students reflected on their own performance on the day, considered how their writing had changed or not changed compared with previous tasks, and set various goals for the next writing task. There were four subsections in the first section on: (1) grammar and vocabulary, (2) organization and expression, (3) content, and (4) writing processes and strategies, and students wrote about how they had performed in terms of each aspect. Before they wrote the first composition, I explained how they were to write in each section, and presented some examples. It usually took about 10 minutes for the

students to complete the reflection sheet. All the comments referred to in this paper are translations by the author.

I wrote a few encouraging comments on all the compositions and returned them the following week. I tried to give similar feedback to both classes. As the aim of the feedback was to create the sense of an audience and to maintain students' motivation to write every week, I did not correct any errors. I returned my feedback after the students had finished writing their compositions, so they did not look at their previous compositions or my feedback while writing.

### ***Text Measurements***

To analyze the approximately 1350 compositions that were written by the students over one year, we used the Web-based computational tool *Coh-Metrix* (Graesser et al., 2004) available at <http://cohmetrix.memphis.edu/cohmetrixpr/index.html> (version 2.0). Spelling mistakes in the compositions were carefully corrected manually and also with a spell checker. This study focused on three aspects of text quality (fluency, lexical complexity, and grammatical complexity), so the three text measures below for each aspect were used.

***Fluency.*** Fluency was measured with the number of words per composition. This is a simple, but powerful tool to assess fluency. L2 writing research has revealed that it is one of the best measures to distinguish L2 writing proficiency levels (e.g., Jarvis, Grant, Bikowski, & Ferris, 2003).

***Lexical Complexity.*** The Measure of Textual, Lexical Diversity (MTLD) was used to assess lexical complexity (McCarthy & Jarvis, 2010). The measure is similar to the type-token ratio, but the type-token ratio is strongly influenced by text length, and is therefore not advantageous to compare texts of different lengths. MTLD was developed to overcome the major weakness with the type-token ratio. MTLD was not calculated with Coh-Metrix but a standalone tool that Phil McCarthy devised.

***Grammatical Complexity.*** We used sentence syntax similarity (STRUT, all sentences across paragraphs) (McCarthy, Cai, and McNamara, 2009) to measure grammatical complexity. STRUT represents the degree of similarity in the syntactic structures of sentences in a passage by comparing the syntactic trees of each pair of sentences. If a wider variety of different sentence structures is used in a passage, its STRUT value will decrease. However, to make STRUT visually comparable with the other two measures, the equation ( $y=1-x$ ) was used as McCarthy et al. (2009) did in their study.

### ***Analysis***

To visually capture trends in changes of L2 writing development, the study employed two graphical tools: progmax-regmin graphs and altitude lines graphs. A progmax-

regmin graph displays a very general developmental trend (van Greet & van Dijk, 2002). This graph can be drawn with a spread sheet program by fixing the first and last points and calculating the maxima and minima respectively within a certain window size (I used a moving window of 5 in this study). For instance, if there are 30 data points, the prograx-regmin graph is calculated as:

$$\begin{aligned} & \max (t1 \cdots t5), \max (t1 \cdots t6), \max (t1 \cdots t7), \cdots \max (t1 \cdots t30) \\ & \min (t30 \cdots t26), \min (t30 \cdots t25), \min (t30 \cdots t24) \cdots \min (t30 \cdots t1) \end{aligned}$$

The progmax-regmin graph is usually drawn with one person's data, but when I drew graphs for a class of students, I used the median from each week. Medians instead of averages were used because the median can represent a student who exhibited average performance in a particular week, while an average cannot represent a real student.

Whereas progmax-regmin graphs are drawn using the medians for each week, altitude line graphs indicate the range of variability in a class of students by plotting some students whose performance is located at various positions in a week (usually the maximum, 75<sup>th</sup> percentile, 50<sup>th</sup> percentile, 25<sup>th</sup> percentile, and minimum). Three positions were plotted in the present study, i.e., 75<sup>th</sup> percentiles, 50<sup>th</sup> percentiles, and 25<sup>th</sup> percentiles. The maxima and minima were removed from the graph because a few students occasionally demonstrated extraordinary performance, which disrupted the description of the overall characteristics of changes.

## Results

### *Change in Two Classes*

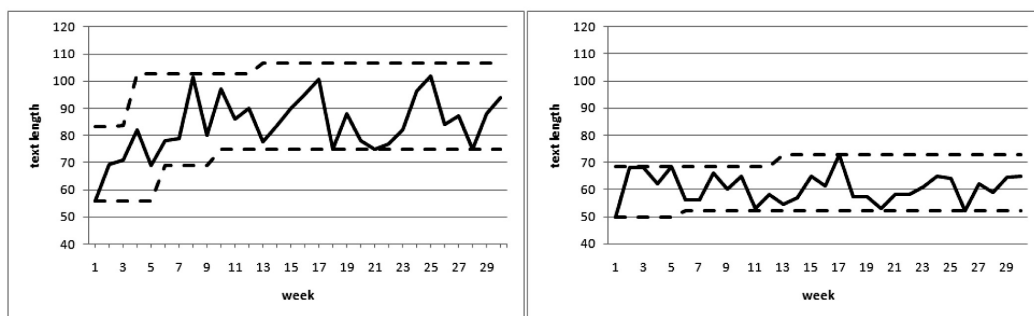


Figure 1. Moving progmax-regmin graph of change in text length (median) for Extended Reflection Class (left) and for Short Reflection Class (right).

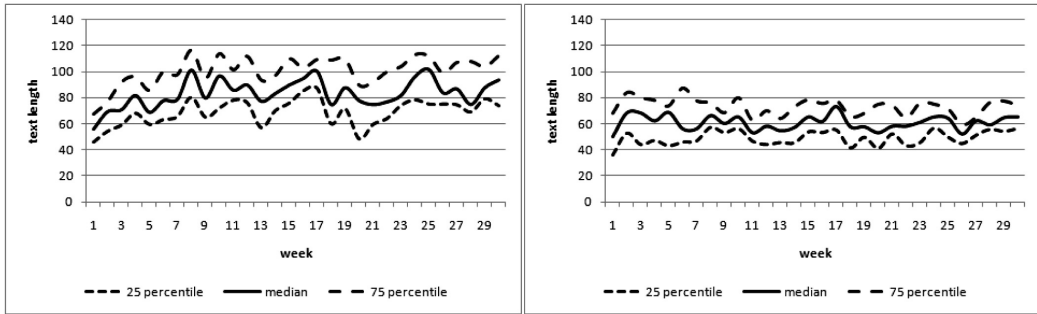


Figure 2. Altitude line graphs of text length for Extended Reflection Class (left) and for Short Reflection Class (right).

To answer the first research question, I will first describe the change in the two classes in terms of the three indices (fluency, lexical complexity, and grammatical complexity). In this analysis, "class" is regarded as a "system." To describe change in a system is like observing a flock of flying birds. It does not permit simplification or abstraction with averages, neither does it allow a mechanical judgment of whether or not writing quality has improved. Rather, it should involve the synthetic analysis of different data with a more or less subjective interpretation.

Figure 1 has a line graph of text length (the median for each week) and a progmax-regmin graph (dashed lines) for the Extended Reflection Class and Short Reflection Class. As we reported in our previous study (Baba & Nitta, 2010), the text length for the Short Reflection Class did not indicate an incremental trend, and shifted within a narrow and stable width of a band of fluctuations (roughly between 50 and 70 words). The text length for the Extended Reflection Class, on the other hand, sharply increased until the 8<sup>th</sup> week, and then dynamically fluctuated within a wider width of a band (roughly between 75 and 100 words). The altitude line graphs for the two classes in Figure 2 display the same tendency. The altitude line graph for the Short Reflection Class indicates a stable trend, while that for the Extended Reflection Class plots an incremental trend.

The progmax-regmin graphs in Figure 3 indicate the changes in the medians of MTLTD for the two classes. In contrast to text length, the two graphs display a similar incremental tendency. Yet, the fluctuations looks wider for the Short Reflection Class (although the median was exceptionally high in the 11<sup>th</sup> week for the Extended Reflection Class), especially in the latter half of the academic year. In addition, the median shifted to a higher level towards the end for the Short Reflection Class, while it was rather stable for the Extended Reflection Class. The altitude line graphs in Figure 4 may support this observation. As a whole, MTLTD moved at a slightly higher level for the Short Reflection Class than for the Extended Reflection Class, but the

overall tendency for the two classes was fairly similar.

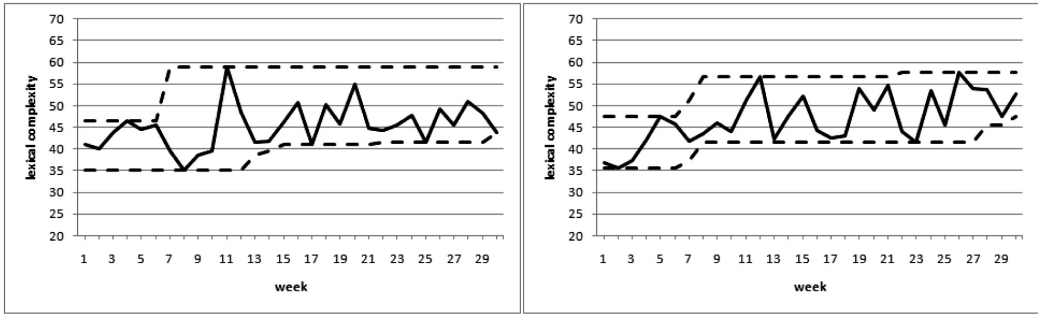


Figure 3. Moving progmax-regmin graph of change in MTLD (median) for Extended Reflection Class (left) and for Short Reflection Class (right).

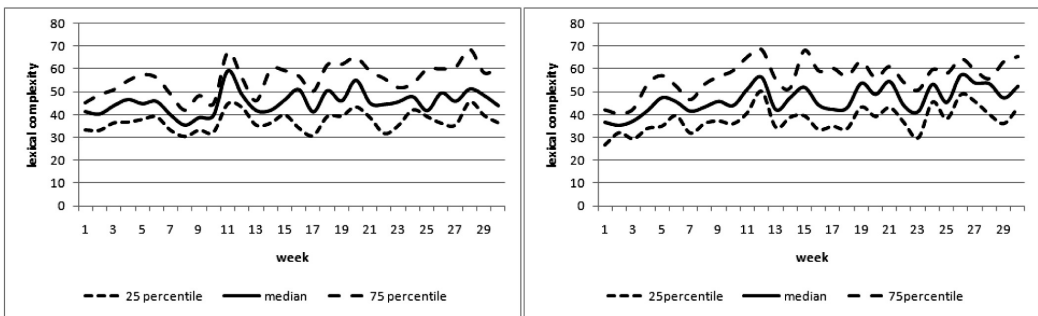


Figure 4. Altitude line graphs of MTLD for Extended Reflection Class (left) and for Short Reflection Class (right).

The change in STRUT indicates a different pattern. Unlike text length, the STRUT of the two classes shifted largely at similar levels. However, the progmax-regmin graphs in Figure 5 indicate that the median for grammatical complexity steadily increased for the Short Reflection Class, while that for the Extended Reflection Class remained relatively stable, except that it shows an incremental trend after the 23<sup>rd</sup> week. The altitude line graphs in Figure 6 also indicate that grammatical complexity for the Short Reflection Class increased rather smoothly throughout the academic year, while that for the Extended Reflection Class did not exhibit such a clear incremental trend. Nonetheless, it would be untrue to say that grammatical complexity for the Extended Reflection Class remained the same. When the 25<sup>th</sup> percentile line for the Extended Reflection Class is removed and plotted (Figure 7), a clear incremental trend appears just as in the median for the Short Reflection Class. This may mean that the inter-individual variability of grammatical complexity for the Extended Reflection Class decreased as the levels of poor performers increased. In our previous study (Baba & Nitta, 2010), we found that repetition of the writing task seemed to enhance the

minimum quality of the student's writing (the students began to write better even under difficult conditions). If we regard this phenomenon as a certain kind of development, it would be possible to say that grammatical complexity for the Extended Reflection Class underwent some changes, which looked like development.

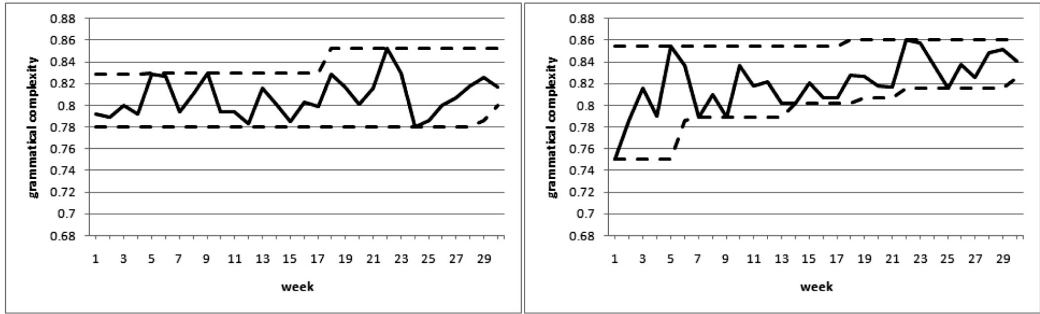


Figure 5. Moving progmax-regmin graph of change in STRUT (median) for Extended Reflection Class (left) and for Short Reflection Class (right).

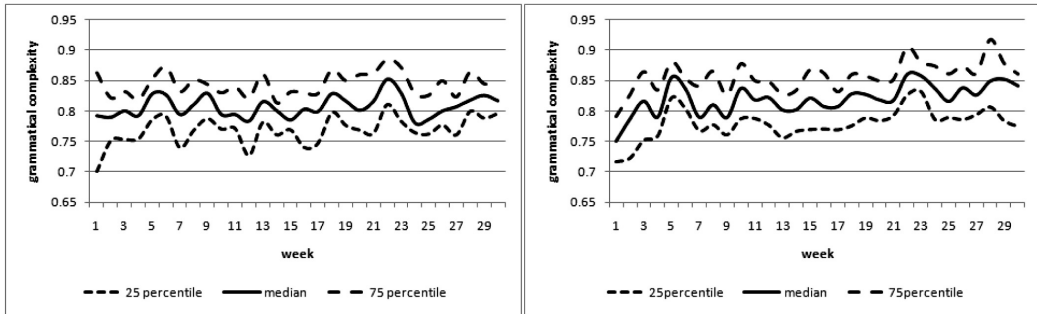


Figure 6. Altitude line graphs of STRUT for Extended Reflection Class (left) and for Short Reflection Class (right).

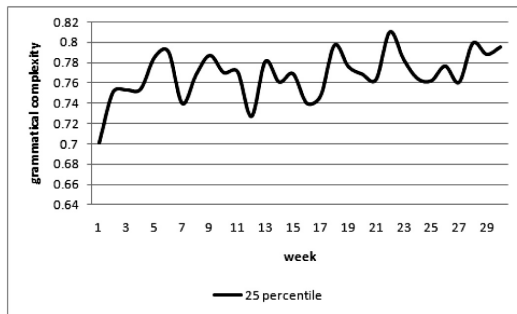


Figure 7. 25<sup>th</sup> percentile of STRUT for Extended Reflection Class.



### ***Changes in Three Students***

The analysis thus far has suggested that the most noticeable difference between the Extended Reflection Class and Short Reflection Class was the change in text length. Therefore, to answer the second and third research questions, I will focus on the text length by three students who were in the Extended Reflection Class. After carefully comparing the changes in the three indices and examining the reflective comments by all the students in the Extended Reflection Class, I found that almost all of them demonstrated some sign of a developmental trend in at least one index. However, the changes in some students were more dramatic than those in others. In what follows, I will discuss two students, Kei and Michi (pseudonyms) who experienced a substantial change in their writing, and one student, Asako (pseudonym), who did not seem to experience such a change in any of the three writing aspects (fluency, lexical complexity, or grammatical complexity).

#### ***Kei***

Figure 8 plots the changes in text length in Kei's writing over one academic year. As the moving progmax-regmin graph shows, the lengths of her compositions started from about 60 words, and reached some 160 words toward the end. This change seemed to go through a few different stages, as the graph has a stepped shape. To make these stages more evident, I added a line graph of a moving standard deviation of residuals (window size=5) to Figure 9 (van Geert, & van Dijk, 2002). This line represents the magnitude of change with the developmental trend omitted. That is, a higher value in this line graph indicates a period of marked changes regardless of the raw values (as the text lengthened, the moving standard deviations increased, but this effect of raw values was removed by using residuals). Conversely, low values indicate periods with small fluctuations.

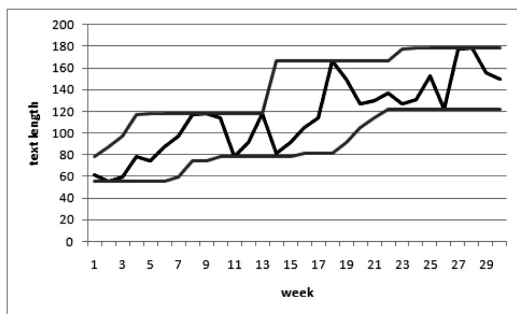


Figure 8. Moving progmax-regmin graph of change in text length in Kei's writing.

The moving standard deviation in the residuals line in Figure 9 has a few peaks and

valleys. In Kei's case, the first small peak suggests that her compositions became lengthy as she got used to timed writing tasks and she stayed at this level for a short period. This phenomenon was observed in many students. After that, there were three peaks (the academic year finished right at the highest point of the third peak). I drew three lines at each peak, which signals a period of intense change. A valley came after this transition period, which suggests that Kei advanced to a new stage and stayed there for some time. The progmax-regmin graph indicates that her change is stepped, and the three lines are located at the beginning of each staircase. It is likely, therefore, that some phase shift occurred at these points. The first phase shift occurred in about the 8<sup>th</sup> week, and its transition period ended in around the 11<sup>th</sup> week. Then, the second phase shift occurred in about the 14<sup>th</sup> week, and Kei moved to the next stage in about the 20<sup>th</sup> week. The last phase shift occurred in about the 24<sup>th</sup> week, and it seems that Kei kept experiencing moderately great changes afterwards.

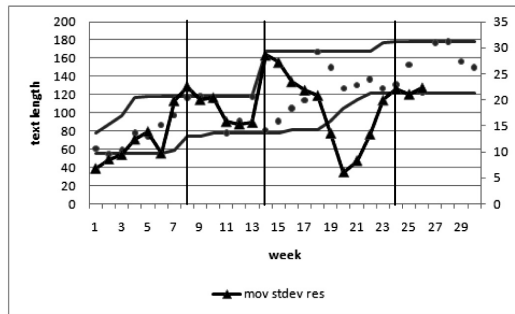


Figure 9. Magnitude of change (detrended) in text length in Kei's writing.

What was happening during the phase shifts? Why did Kei undergo them? Her compositions and reflective comments partly answer these questions. Kei was a diligent and motivated student. Compared to other students, she wrote a relatively large volume of comments on the reflection sheet, and her comments were detailed and concrete. For example, she wrote in the grammar and vocabulary section in the 4<sup>th</sup> week that "I started to learn vocabulary with a TOEIC textbook, so I wanted to use these words. I'm so pleased when I encounter words I know; what do you think? I think I'll come to love the TOEIC textbook. My grammar is confusing. It would be useful to memorize sentences from the book." She then wrote in the organization and expression section, "I used phrases, especially idioms, today. I felt thrilled when I could include idioms in my sentences. I would like to become a walking dictionary." The comments were quite copious for such a small space.

Whereas Figures 8 and 9 only present the text length, her compositions demonstrated some qualitative changes around the phase shift periods. As mentioned in the

Methods section, I provided my students some encouraging feedback, and always wrote either "Good" or "Very Good" below their compositions. I only awarded them with "Very Good" when I was particularly impressed with a composition, so I rarely awarded this. However, Kei received several of these. The first time I awarded her with a "Very Good" was in the 7<sup>th</sup> and 8<sup>th</sup> weeks, which precisely correspond to the first phase shift in Figure 9. The following are her 5<sup>th</sup> and 8<sup>th</sup> compositions.

When I choose friend common hobbies. it's important to talk for a long time. And funny. Kind is so good. Singing good is so happy. Because I go to Karaoke many time. Like festival. Because I love festival and it more enjoying. Believer, I always to believe people. But in my friend is not believe is in it. I'm so sad. Believer will be my best friend. I guess people look like. (Week 5, 72 words, topic: What characteristics do you regard as important in a person you would choose as a friend?)

I really respect my parents. Most is my dad. He is pioneer of Survival Game. He brought it in Japan of first person. And "bb dan [toy bullet]" is his introduced. Do you know "bb dan"? It has very popular in Japan in these days. Children for Adult was use it for Model Gun. As you know, Survival Game also relative to Model Gun. Because my father is famous person in the Model Guns society. Many people respect to him. But he isn't pridely. And no use a lot of money. So usually and usually. I respect it. (Week 8, 98 words, topic: What are the most important skills and/or values that you have learned from your parents?)

In the 5<sup>th</sup> week, her composition sounds fragmented and disconnected. Most sentences are incomplete or ungrammatical. The response is inconsistent, which starts from hobbies, and suddenly jumps to a festival, and to whether her friends believe her or not. Her 8<sup>th</sup> composition, on the other hand, sounds more coherent, even though there are grammatical errors here and there. In her reflective comments, she wrote, "I think I could write a good composition today, even though I repeated myself." I replied to her comments, "I agree. There was an explanation about your father, and it led to why you respect him. It's readable." She also commented on her grammar and vocabulary, "I tried to use idioms, like *as you know*. I have been conscious of grammar and vocabulary since last week [probably in other English classes]."

Kei's first phase shift did not occur out of the blue. Her reflective comments imply that there was some kind of prelude to it. During the first three weeks, most comments she wrote were negative, stating what she could not do or some difficulty

she had experienced in writing. Then in the fifth week, she emphasized that she enjoyed writing. She wrote in the content section, "Probably because the topic was good for me, I really enjoyed writing. I wanted more time to think before, but I didn't feel like that today. I felt for the first time that I would like to write more. I thought that it is important to think carefully and write, but it is also important to spontaneously move my hands." In comparison with previous writing she also wrote, "I felt more and more excited. Different from the previous composition, I felt I wanted to write more. I look forward to writing next week," and she even thanked me for the fact that she felt this way. Then in the 7<sup>th</sup> week, she recognized that she had done a good job as previously mentioned, and was pleased that the number of words had increased.

During the second phase shift (approximately between the 14<sup>th</sup> and 19<sup>th</sup> weeks), she seemed to attempt many things in her writing. For instance, she "set a goal to write at least 5 words per sentence" (Week 14), "decided to write a plot of the episode first, and then write details" (Week 15), and "tried to clarify *when, who, what, how*" (Week 16). She stated she "could use the conjunctions that I had learned during the summer vacation, and noticed the importance of adverbs and adverbial phrases such as *in fact* and *recently* because they are easy to use" (Week 17). She also wrote that she "used a lot of examples with *looks like, like, and example*" (Week 18). Then in the 19<sup>th</sup> week, I awarded her a "Very Good" again. In that week, she mentioned that she used the words that she learned in other classes, and was also excited that her mind was filled with what she would like to write about. Further, she was pleased that she could write about many things that she wanted to write about. After that, her writing became more fluent and the content became thoughtful, and she received "Very Goods" every week until Week 23. This may have meant that she got used to writing compositions of this length, and the quality of her writing stabilized to some extent.

Toward the end of the academic year, during which she still seemed to be experiencing the last phase shift, she began to pay attention to organization for the first time. In the 27<sup>th</sup> week, she wrote, "I would like to use a more effective organizational structure," and then in the following week, she attempted to divide her composition into three paragraphs (she had learned how to arrange paragraphs in my class). However, this attempt ended in failure as she reflected, "I put *First*, but forgot to write *Second and Finally*, which is regrettable." Although her attempt to create paragraphs was unsuccessful, her compositions in the 27<sup>th</sup> and 28<sup>th</sup> weeks were the longest (both were 179 words). This did not hold for all students, but in Kei's case, her efforts to write well-organized compositions might have resulted in lengthier compositions.

One thing that was noteworthy in Kei's reflective comments throughout the

academic year was that she tried to utilize knowledge that she had acquired outside class or to seek such knowledge and writing techniques. As previously mentioned, she often used words that she had learned for the TOEIC test or in other English classes. In the 13<sup>th</sup> week, she even referred to a TV program, and wrote, "In the program, they said that when you do not know how to say something in English, you may want to rephrase it into a simpler expression in Japanese, and then translate it into English again. I was not sure how to say *Shokyokuteki* [passive], so I wrote *not outgoing*." These comments suggest that she devoted a great deal of effort to this writing task, which may be partially attributable to the fact that she enjoyed the task as she had written in the 5<sup>th</sup> week.

### **Michi**

Like Kei, Michi seemed to experience a few phase shifts over one academic year. She was one of the most motivated students, and wrote the longest compositions in the class. Although Kei wrote many comments on the reflection sheet, Michi wrote even more. Figure 10 shows the changes in the lengths of Michi's compositions and her progmax-regmin graph. Her compositions started with some 100 words, and reached about 250 words toward the end. Whereas this differed from Kei's developmental pattern, the length of Michi's compositions also seemed to shift in stages.

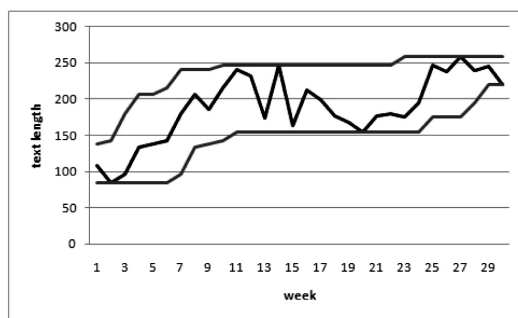


Figure 10. Moving progmax-regmin graph of change in text length in Michi's writing.

The moving standard deviation of the residuals line in Figure 11 makes these stages more explicit. Michi also underwent some initial changes as Kei had experienced until she got used to the timed writing tasks in about the 8<sup>th</sup> week. Except for this initial change, Michi seemed to go through two phase shifts, which are indicated by the vertical lines. In the first phase shift, there was a large fluctuation period that started around Week 11 and settled to a new stage around Weeks 18 or 19. The second phase shift began in Week 23, and ended in about Week 25, which seemed to lead Michi to another stage.

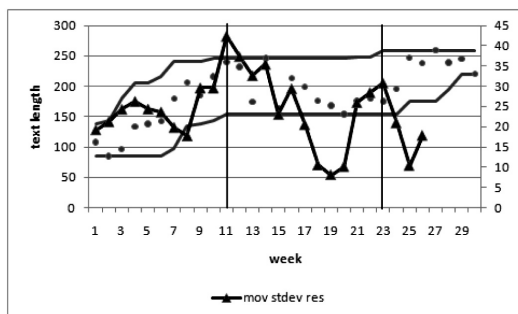


Figure 11. Magnitude of change (detrended) in text length in Michi's writing.

Similar to Kei, there seemed to be a preliminary phase in Michi's mind, and it took her some time to fully rise to a new level. Michi's first phase shift can be characterized by her acquiring the ability to create paragraphs. Whereas Kei became conscious of paragraphs at the very end of the academic year (and did not seem to have mastered paragraphs), Michi first referred to "organization" in the 8<sup>th</sup> week (about three weeks prior her first phase shift). In that week, she wrote, "the organization was horrible." She lamented again in the following week, "I couldn't think of organization!" These comments suggest that she started to recognize the organizational problem in her writing. Then in the 10<sup>th</sup> week, she did not organize paragraphs, but wrote that she had intended to do so, "I couldn't make paragraphs, because if I had done so, I would have run out of space [in fact, she wrote on all the lines, and even in the margins]. Next time, I will ask for another sheet, and I wish I could write with paragraphs in mind."

After these three weeks, she tried to make paragraphs for the first time in the 11<sup>th</sup> week, but this attempt was not entirely successful. She wrote, "I started with *First*, but failed to put *Next*, so my composition turned out badly." Finally in Week 12, she divided her composition into four paragraphs. However, probably because she lacked confidence, she asked me "Are my paragraphs all right?" on the reflection sheet. In Week 13, she realized that she had mastered paragraphs, and wrote, "I learned how to make paragraphs." She then created paragraphs for three weeks. However, in the 17<sup>th</sup> and 18<sup>th</sup> weeks, she failed to do so again and wrote, "I messed up sentences without thinking of paragraphs (Week 17)" and "I couldn't think about organization. I couldn't think about paragraphs, either (Week 18)." Finally, in the 19<sup>th</sup> week, she organized paragraphs again and wrote, "I thought about organization before I started writing." From then on, she completely acquired the concept of paragraphs, and used them each time. The following excerpts are from Michi's compositions before and after the first phase shift.

If my doctor told me that I had only a few months to live. I want to say "Thank you" to everyone who I met before!!! Of course my family, friends, teachers and more a lot!! I want to write many many letter for everyone!!! And I go to various place!! I want to always smiling and always happy!! I want to travel to Hawaii and Guam! That place is my memorial place!! And I want to meet some teacher in Florida!! I want to go to Florida with my friend and I want to meet Larissa!! who is my favorite teacher in Eckerd College. Larissa is so kind and her smile is so cute!! Her body is so big!! but She can running fast!! I like to eat many sweets and food! I am heavy eating girl! So, I want to many delicious foods and sweet in various place!! Everyday is so important!! I spend everyday unintentionally but I want to spend every day unintentionally but I want to spend more important than before. I want to say everyone who I met  
(Week 9, 183 words, topic: If your doctor told you that you had only a few months to live, how would you alter your way of life?)

I think "good parents" is so difficult to tell. Just "kindness person" is not good person of course. "Parents" is great person I think! Because "mother" cook lunch everyday to her son and daughter!! She have to get up early!! Between cooking, washing and cleaning, "mother" is very very busy!! But she do all things everyday!! That is very good!! This person is Just "good mother".

"Father" is more difficult to think than "mother". Many children go shopping and dinner with them father, but he is not always with them, because he works everyday. He wakes up early, eat breakfast, change the clothes and go to his office. He back to his home at night that day!! That is so hard!!! But our father are managing almost everyday! I expect "father".

Mother cooks children's meal and father works everyday to earn money for his family's life... This is a "good parents!" I think!!

They have a dignity and kindness mind. They always think child  
(Week 19, 164 words, topic: What do you think are the essential characteristics of a good parent?)

Michi's composition in Week 9 sounds like an inventory of her wishes. Her writing suddenly jumps to the teacher she likes, and then to her favorite food. It seems that she was writing off the top of her head. However, her Week 19 composition is better organized, even though it is almost as long as the one in Week 9. She wrote about her mother in the first paragraph, and about her father in the second paragraph. In the third paragraph, she tried to summarize what good parents are. Each paragraph



works as a semantic unit, and the connection between each sentence sounds more natural compared to the Week 9 composition.

Michi's second phase shift seems to be associated with affective factors. Unlike the first phase shift, her writing does not seem to have changed qualitatively in a straightforward manner. Rather, her compositions became lengthy because she was firmly resolved to make them so. One week before her second phase shift (Week 22), she wrote many pessimistic comments, such as "I don't feel I'm making progress," "I couldn't write well today, either." "My writing was bad this time as well as the previous time." In the next week, she still wrote some negative comments on her writing. However, at the same time, she stated; "I can't explain myself well, but *I just have to work hard in order to make advances*, instead of just *wishing to make advances*." In the 24<sup>th</sup> week, she wrote comments such as those where she was persuading and encouraging herself: "While I was writing, I strongly wished I could write a better composition," "I would like to keep up and remain undaunted," "It was better when I autosuggested myself that I could do it," and "I would like to love this timed-writing task more."

What was notable in Michi's attitude was that she often wrote some kind of mottos or goals on the reflection sheet, and she used them to motivate herself as she did during the second phase shift. For example, she wrote "Work day in day out," "Don't feel pressed," "Do my best," and "Stimulating sentences!" Most important of all, she enjoyed and tried to make herself enjoy this task. In the 4<sup>th</sup> week, she wrote to herself for the first time, "Enjoy writing!" and made similar comments three times. She also wrote that she enjoyed writing five times.

In addition, Michi seemed to make efforts directed toward this task outside class, as Kei did. For instance, she wrote that she tried to memorize the words that she learned in other classes, and used them for this task. In the 14<sup>th</sup> week, she mentioned that she remembered some expressions that she had learned when she took part in a short skit in a speaking class, and was pleased that she could use them in her writing. Similarly, in the 27<sup>th</sup> week, she wrote that she had found a Website to learn English and was using it to promote her reading abilities and to build up her vocabulary. She wished she could apply what she had learned there to the timed writing task. Moreover, in Weeks 24 and 25, she received some advice from her friend, who was not her classmate but close to Michi, and acted on it.

### ***Asako***

In contrast to Kei and Michi, Asako did not seem to experience any phase shifts, judging from the changes in the text length of her compositions (Figures 12 and 13). The moving standard deviation of the residuals line in Figure 13 suggests that Asako's

compositions became lengthier as she got used to the tasks until Weeks 8 and 9, which was similar to what Kei and Michi had experienced. Asako then seemed to undergo two periods of major fluctuations as the vertical lines indicate. However, as the moving progmax-regmin graph shows, even after these periods with large fluctuations, there did not seem to be any stepped changes, which could be seen in Kei's and Michi's graphs.

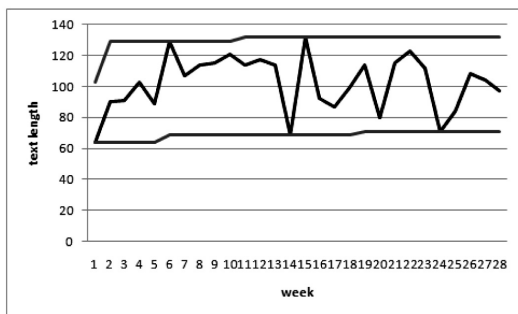


Figure 12. Moving progmax-regmin graph of change in text length in Asako's writing.

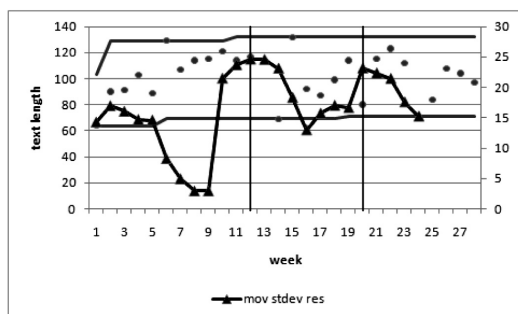


Figure 13. Magnitude of change (detrended) in text length in Asako's writing.

Asako was a diligent student, and as a whole, I thought her writing ability was above average in the class. The following compositions were written by Asako when she became familiar with the timed writing task (Week 7) and toward the end of the academic year (Week 29), which are similar in length. Although she made various grammatical errors in both, she had a fairly good command of grammatical structures for a student in my class even in the 7<sup>th</sup> week. In both compositions, she mentioned something or someone she liked, and recounted her own episodes. The flow of the sentences is not particularly awkward in either composition. Unfortunately, however, qualitatively different changes are hard to identify between the two compositions.

If I was alone for a week, I read a lot of books. For example, fantasy,

mystery etc. Especially I want to read "Gone with the wind". Because I watched a video last years. It's very interesting so I thought want to read it. However. there is nothing at the nearly library. I like read books very much. When I was primary student, I almost read at school library. Everyday I went to there, I borrow 2 or 3 books. Next day I returned that books and I borrow new books. However when I was high school student, I didn't read a lot of books.

(Week 7, 105 words, topic: If you were alone for a week, what books (or music) would you select to read (or listen to)? Explain about them (it).)

If I could have a conversation with a famous person, I want to talk with Gacto Tamura. he is a famous figure-skater. If I talk with him, I will learn several jump and spin. In fact, I talk with famous figure-skater. For example, Miki Ando, Mai Asada, Mao Asada are friend with me. Because When I was elementary school student, we took part in same club. Especially, I played with Mao. Another person, I talked with Takahiko Kozuka. However I talked with him a little. Because he was very shy, and me, too. So, If we talked him, my friend talked a lot alone.

(Week 29, 104 words, topic: If you could have a conversation with a famous person (living or dead), whom would you choose? Discuss.)

Interestingly, Asako's comments on the reflection sheet were usually short, and as she repeated the task, they became even shorter. She basically wrote what she did. For instance, after she wrote the Week 7 composition above, she wrote, "I wrote about the books that I had read when I was small," and after the Week 29 composition she wrote, "I wrote about the people that I had talked with as well." She always wrote some goal in the section for "goals for future writing," but her goals were usually vague or too abstract. Some of her goals were "[I should] enrich the content," "To write more easily comprehensible sentences," "and To be able to write more fluently." They provide a stark contrast to the goals that Kei and Michi described, which were more detailed and concrete. After I had carefully read all of Asako's compositions and reflective comments, I gained the impression that she was seriously engaged in this task, but that she distanced herself from it. She was not very enthusiastic about the task, and probably did not think about it once she had finished writing the compositions (she never mentioned she had done anything for it outside the class).

## Discussion

The first research question in this study was "How does the writing ability of two classes of EFL university students change in terms of fluency, lexical complexity and grammatical complexity when they repeat the same writing task with or without extended and focused reflection over one year?" The graphical analyses of the progmax-regmin graphs and the altitude line graphs suggest that the most notable difference between the Short Reflection Class and Extended Reflection Class was the change in text length. While the text length of the Short Reflection Class remained at the same level throughout the academic year, that of the Extended Reflection Class became distinctively longer. There was not such a significant difference in the other two aspects (lexical complexity and grammatical complexity) between the two classes, even though grammatical complexity increased more obviously in the Short Reflection Class. These results support some previous studies that demonstrated task repetition without corrective feedback such as journal writing has some effect on L2 writing development (e.g., Casanave, 1994; Kepner, 1991; Liao & Wong, 2010).

It is difficult, however, to assess what effect extended reflection has based on these results, because different aspects of writing influence one another. Some aspects may expand because another aspect has or has not developed. It may be possible to assume that grammatical complexity in the Extended Reflection Class did not significantly increase because their texts became longer. Students in the Extended Reflection Class might have paid more attention to fluency and devoted their efforts to improving this at the expense of grammatical complexity. Another possibility is that the measure of grammatical complexity (STRUT) that was used in this study assessed the degree of similarity between sentence structures in a composition, so when students wrote more sentences, it is likely that they used the same structures more frequently, which decreased grammatical complexity. Nonetheless, it may be that there were differences between the two classes, and that these differences were likely to result from whether students wrote extended reflections after the timed writing task or not. Baba and Nitta (2010) support this claim in that another university English class, whose instructor and course objectives were different from those of the two classes in the present study, found a very similar developmental pattern to that in the Short Reflection Class after they had engaged in a timed writing task exactly in the same way.

Because the difference between the two classes appeared most clearly in text length, I focused on fluency in the writing of three students to investigate Research Questions 2 and 3, "How can phase shift occur in individual writers? What are the differences between writers who experience phase shift and those who do not?" and "How are extended and focused reflections related to phase shift?" First, I examined writing by

two students (Kei and Michi) who seemed to have experienced some phase shifts over the academic year. It was revealed that the frequency and timing of phase shifts differed between these two students, and the nature of each phase shift varied even within the same student. Except for the first short period of fluctuation, which was observed in many students until they became used to the task, Kei underwent three moderate phase shifts, while Michi experienced one large and one moderate phase shift. In Kei's case, the first phase shift occurred when she started to enjoy the task. Her second phase shift was promoted simultaneously with trial and error choices and in the third phase shift she started to try organizing paragraphs. Michi, on the other hand, worked on paragraphs during the first phase shift, which lasted for a relatively long period of time. Her second phase shift occurred after the change in her affect and motivation. These results suggest that even in the same class with the same teacher, the same teaching content, and a similar L2 proficiency level, individual students were likely to follow unique developmental paths, as has been reported in previous research (Casanave, 1994; Larsen-Freeman, 2006) and emphasized in Complexity Theory (e.g., Larsen-Freeman & Cameron, 2008).

The students' reflective comments suggested that there were also similarities in the process of phase shift in the broad sense. Previous studies have only looked at phase shift with numerical values, and have observed that a specific feature of L2 writing suddenly *jumped* at some point (Spoelman & Verspoor, 2008) or found that there is a long period of fluctuation before a writing feature settles at a higher level. In addition to such observations of dramatic changes in numerical values, the present study found that individual L2 writers have a preparation period. Before each phase shift, both Kei and Michi reported some changes in their motivation toward the task or some problems and difficulties that they were faced with. They also wished to overcome these problems. They then took some action to improve their writing, which seems to have resulted in the extensive fluctuations in text length. Better and stabler performance only eventuated after that fluctuation period. That is, phase shift may be in a well prepared state before the actual *jumps* in numerical values with the change in the writer's consciousness or motivation toward the task.

The present study also looked at one student, Asako, who did not seem to experience phase shift. The results suggested that there were two principal differences between Asako and the other two students who underwent phase shift. First, Asako's attitude toward the task was less enthusiastic than that of Kei and Michi. As mentioned in the previous section, Asako earnestly engaged in the task, but it probably was not something that she really wanted to develop. The task might have left her cognizance every time she finished writing. In contrast, Kei and Michi gradually made various efforts outside class to progress in the task. When I interviewed Kei at the end of the

academic year, she said, "This task influenced how I studied in other English classes." Kei and Michi made use of their resources including textbooks for other classes, TV programs, and their friends in the task. It seems that Kei and Michi kept these in mind more frequently in their student lives than Asako.

The second difference between those who experienced phase shift and those who did not is concerned with the third research question. The number and intensity of reflective comments that Asako and the other two students wrote were quite different. This could be one reason why she did not undergo phase shift. Although overgeneralizations should not be made from the analysis of a small sample, the results from the study point to the effect writing reflective comments had. Summarizing what some students (including Kei and Michi) said in the interview, reflective comments were useful in two ways. First, writing reflective comments, according to them, made them aware of problems in their own writing. One student told me that reflective comments reminded her of what she could not do on that day, and made her conscious of her strengths and weaknesses. Michi also said, "By writing reflective comments, I noticed that I could not use something well, like a grammatical structure, in my writing, which led me to be aware of that structure and to use it in other English classes." All the students that I interviewed agreed that they felt frustrated with their lack of an extensive vocabulary every time they wrote reflective comments, and also felt that they needed to enhance this.

Second, and more importantly, writing goals and mottos in reflections seemed to impel students into action to improve their performance in the task. As described in the previous section, Michi often wrote goals and mottos on the reflection sheet, and encouraged herself to take some action. Kei also said in the interview that she sometimes noticed the gap between what she wrote (goals and plans) and what she was actually doing, asking herself "Am I expending as much effort as I write here?" According to her, although this awareness did not always drive her to take action, it at least created incentive for her to do so.

The two advantages of reflection mentioned by the students correspond to Boud's (2001) *reflection in anticipation of events* and *reflection after events*. They evaluated their performance after writing compositions, and paid attention to the areas in which they needed to make progress. They then established goals and what they intended to change, wishing to overcome weaknesses. What is noteworthy here is that there were such advantages of reflection because the students *repeated* the writing task. If they engaged in the task once or only a few times, they would not have recognized the advantages of reflection. They knew they would repeat the same task repeatedly (in fact, it seems that they really understood this fact after several weeks), so they analyzed their weaknesses and set goals and plans for their own development from a

long-term viewpoint.

This study was not without its limitations. First, it examined three aspects of L2 writing (fluency, lexical complexity, grammatical complexity), but did not look at accuracy. To analyze multiple drafts from students in two classes, I resorted to a computational tool, so it was not feasible to study accuracy, or to use more detailed indices that would have required manual analyses (e.g., number of words per T-unit or noun phrase). In analyzing the three students' writing, I mainly focused on text length. In future research, however, it might be useful to conduct manual analyses of more features in compositions. Second, the present study collected data from students over one academic year. However, one year may not be long enough to observe critical moments in the development of some students. Even though it would be impossible to collect data from a particular class of students for more than two years, following a small number of students over longer periods would be beneficial.

### Conclusion

This study revealed that the weekly repetition of a timed-writing task brought some changes in L2 student's writing regardless of whether they wrote extended reflective comments or not. Yet, the existence or non-existence of extended reflective comments had a different effect on their writing. In particular, there was a notable difference in text length between two classes, with and without extended reflective comments. This difference was then examined through case studies on three students. The graphical analyses suggested that two students experienced phase shifts, while the remaining student did not. This result was corroborated by further analyses of their compositions and reflective comments, and it was revealed that the graphical analyses were useful for identifying phase shifts, i.e., decisive changes in the development of L2 writing ability. Such analyses were impossible without a multi-wave research design, which pointed to the significance of repeated data collection with relatively short intervals.

The study also described what was happening during phase shifts in reference to the students' internal activities for the first time. Although individual students may follow unique developmental paths depending on a number of factors (e.g., affect, motivation, ways of thinking, L2 writing proficiency, and even one's circle of friends), it has been suggested that one prerequisite for phase shift seems to be "repetition" of the task. Students only seemed to set goals for their next writing task and became more conscious of their own writing development after they realized that the task would be repeated for a fairly long time and understood what this repetition meant to them. Of course, not all of the students benefited from the task used in the study (like Asako). As Holmes and Moulton (1995) showed, some student may disfavor a particular task such as journal writing. Consequently, I am not arguing that the timed writing tasks



were effective for all the students. The timed writing tasks appeared to be well received by most of my students, but there may be more effective tasks for different groups of L2 learners. The point I wish to emphasize is that regardless of the type of task, repeating a particular task may be a critical factor in L2 development.

The findings from this study imply that a Complexity Theory approach may contribute to further understanding of the nature of L2 writing development. The present study did not look at the *environment* that encompasses systems and also interacts with them. However, it has been considered significant in the Complexity Theory approach, for it is necessary to "observe the operation of the whole system and the interaction of the parts rather than focusing on specific units in it" (Dörnyei, 2009, p. 240). Therefore, in future research, we intend to delve into the interactions between students in class as well as classroom dynamics.

### Acknowledgements

We are grateful to the Ministry of Education, Sports, Culture, Sports, Science, and Technology of Japan (Research Grant #21520643) for supporting this research.

### References

- Baba, K., & Nitta, R. (2010). Dynamic effects of task type practice on the Japanese EFL university student's writing: Text analysis with Coh-Metrix. *Proceedings of the 23rd International Florida Artificial Intelligence Research Society Conference*.
- Baba, K., & Nitta, R. (in press). Dynamic effects of repeating a timed writing task in two EFL university courses: Multi-element text analysis with Coh-Metrix. In P. M. McCarthy & C. Boonthum (Eds.), *Applied Natural Language Processing and content analysis: Identification, investigation, and resolution*. Hershey, PA: IGI Global.
- Boud, D. (2001). Using journal writing to enhance reflective thinking. *New Directions for Adult and Continuing Education*, 90, 9-17.
- Bygate, M. and Samuda, V. (2005). Integrative planning through the use of task-repetition. In R. Ellis (Ed.), *Planning and task performance in a second language*. Amsterdam: John Benjamins.
- Byrne, D. (2005). Complexity, configurations and cases. *Theory, Culture & Society*, 22(5), 95-111.
- Casanave, C. P. (1994). Language development in students' journals. *Journal of Second Language Writing*, 3(3), 179-201.
- Casanave, C. P. (2004). *Controversies in second language writing: Dilemmas and decisions in research and instruction*. Ann Arbor: University of Michigan.
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of L2 student writing. *Journal of Second Language Writing*, 12, 267-296.
- Duppenthaler, P. (2004). A Study of the Effect of Three Different Types of Feedback on Writing: Part 2 - Data Analysis. *Tezukayama Gakuin Daigaku Kenkyu Ronju, Bungakubu [Tezukayama Gakuin University Studies and Essays, Faculty of Literature]*, 39, 1-27.
- Dörnyei, Z. (2009). Individual differences: Interplay of learner characteristics and learning

- environment. *Language Learning*, 59(s1), 230-248.
- Ferris, D. R. (2010). Second language writin research and written corrective feedback in SLA. *Studies in Second Language Acquisition*, 32, 181-201.
- Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum.
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-Metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, 36(2), 193-202.
- Guenette, D. (2007). Is feedback pedagogically correct? Research design issues in studies of feedback on writing. *Journal of Second Language Writing*, 16, 40-53.
- Holmes, V. L., & Moulton, M. R. (1995). A contrarian view of dialogue journals: The case of a reluctant participant. *Journal of Second Language Writing*, 4(3), 223-251.
- Ishikawa, S. (1995). Objective measurement of low-proficiency EFL narrative writing. *Journal of Second Language Writing*, 4(1), 51-69.
- Jarvis, S., Grant, L., Bikowski, D., & Ferris, D. (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing*, 12, 377-403.
- Kastens, K. A., & Liben, L. S. (2007). Eliciting self-explanations improves children's performance on a field-based map skills task. *Cognition and Instruction*, 25(1), 45-74.
- Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second-language writing skills. *The Modern Language Journal*, 75(3), 305-313.
- Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics*, 27(4), 590-619.
- Larsen-Freeman, D. (2009). Adjusting expectations: The study of complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, 30(4), 579-589.
- Larsen-Freeman, D., & Cameron, L. (2008). *Complex systems and applied linguistics*. Oxford: Oxford University Press.
- Liao, M.-T., & Wong, C.-T. (2010). Effects of dialogue journals on L2 students' writing fluency, reflections, anxiety, and motivation. *Reflections on English Language Teaching*, 9(2), 139-170.
- McCarthy, P. M., & Jarvis, S. (2010). MTLTD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381-392.
- McCarthy, P. M., Cai, Z., & McNamara, D. S. (2009). Computational replication of human paraphrase assessment. In C. H. Lane & H. W. Guesgen (Eds.), *Proceedings of the 22nd International Florida Artificial Intelligence Research Society Conference* (pp. 266-271). Menlo Park, CA: The AAAI Press.
- Mellow, J. D., Reeder, K., & Forster, E. (1996). Using time-series research designs to investigate the effects of instruction on SLA. *Studies in Second Language Acquisition*, 18, 325-350.
- Ortega, L., & Byrnes, H. (Eds.). (2008). *The longitudinal study of advanced L2 capacities*. New York: Routledge.
- Ortega, L., & Ibarra-Shea, G. (2005). Longitudinal research in second language acquisition: Recent trends and future directions. *Annual Review of Applied Linguistics*, 25, 26-45.
- Peyton, J. K. (1990). Dialogue journal writing and the acquisition of English grammatical morphology. In J. K. Peyton (Ed.), *Students and teachers writing together: Perspectives on journal writing* (pp. 67-97). Alexandria, VA: Teachers of English to Speakers of Other Languages.
- Shaw, P., & Liu, E. T.-K. (1998). What develops in the development of second-language writing? *Applied Linguistics*, 19(2), 225-254.

Reflection in second language writing: A longitudinal study of task repetition from a Complexity Theory perspective (Kyoko BABA)

- Spoelman, M., & Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: A longitudinal case study in the acquisition of Finnish. *Applied Linguistics*, 31(4), 532-553.
- Suzuki, W. (in press). Written languaging, direct correction, and second language writing revision. *Language Learning*.
- Swain, M. (2006). Verbal protocols: What does it mean for research to use speaking as a data collection tool? In M. Chalhoub-Deville & C. A. Chapelle & P. Duff (Eds.), *Inference and generalizability in applied linguistics: Multiple perspectives* (pp. 97-113). Amsterdam: John Benjamins.
- Swain, M., Lapkin, S., Knouzi, I., Suzuki, W., & Brooks, L. (2009). Languaging: University students learn the grammatical concept of voice in French. *The Modern Language Journal*, 93(1), 5-29.
- The "Five Graces Group". (2009). Language is a complex adaptive system: position paper. *Language Learning*, 59(s1), 1-26.
- Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, 16, 255-272.
- van Geert, P., & van Dijk, M. (2002). Focus on variability: New tools to study intra-individual variability in developmental data. *Infant Behavior & Development*, 25, 340-374.
- Verspoor, M., Lowie, W., & van Dijk, M. (2008). Variability in second language development from a dynamic systems perspective. *The Modern Language Journal*, 92(2), 214-231.
- Weissberg, R. (1998). Acquiring English syntax through journal writing. *College ESL*, 8(1), 1-22.
- Willett, J. B. (1994). Measurement of change. In T. Husen & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (2nd ed., pp. 671-678). Oxford: Pergamon.