Toward an Empirical Model of EFL Writing Processes: An Exploratory Study

MIYUKI SASAKI
Nagoya Gakuin University, Japan

The present study investigated EFL learners’ writing processes using multiple data sources including their written texts, videotaped pausing behaviors while writing, stimulated recall protocols, and analytic scores given to the written texts. Methodologically, the study adopted a research scheme that has been successfully used for building models of Japanese L1 writing. Three paired groups of Japanese EFL writers (experts vs. novices, more- vs. less-skilled student writers, novices before and after 6 months of instruction) were compared in terms of writing fluency, quality/complexity of their written texts, their pausing behaviors while writing, and their strategy use. The results revealed that (a) before starting to write, the experts spent a longer time planning a detailed overall organization, whereas the novices spent a shorter time, making a less global plan; (b) once the experts had made their global plan, they did not stop and think as frequently as the novices; (c) L2 proficiency appeared to explain part of the difference in strategy use between the experts and novices; and (d) after 6 months of instruction, novices had begun to use some of the expert writers’ strategies. It was also speculated that the experts’ global planning was a manifestation of writing expertise that cannot be acquired over a short period of time.

The present study investigated the writing processes of Japanese EFL (learning English in a non-English-speaking environment) learners with three different levels of second language (L2) writing ability (professional, and more- and less-skilled) both cross-sectionally and longitudinally. Although an increasing number of studies have recently begun to report details of L2 learners’ cognitive processes while writing in L2 (see Cumming, 1998, for examples), very few studies to date have examined the writing processes of EFL learners with different levels of L2 writing proficiency, and the changes in their writing processes over time. The present study is also unique in its use as the main data source of stimulated recall protocols produced by the participants while watching their own writing behaviors on video immediately after they finished writing, which made it possible to induce writing process data even from those
participants who would have been too distracted by the traditionally used think-
 aloud data collection method to produce meaningful data.

BACKGROUND

Studying the composing process has been a major focus of L2 writing research for the past several decades (Cumming, 1998; Krapels, 1990; Silva, 1993). Basically following the designs of first language (L1) composition process studies, researchers have investigated various aspects of L2 writing processes for different groups of participants. Of particular interest to the present study were those studies that examined micro-level writing processes without interference, such as planning enforcement by the researchers (e.g., Friedlander, 1990). One of the earliest of those studies was by Zamel (1983, p. 169), who analyzed “think-
aloud” protocol data collected while six “advanced” ESL (learning English in an English-speaking environment) students were completing “a course-related writing task.” Among these six students, four were identified as “skilled” and two as “unskilled” based on experienced readers’ “holistic assessments” (p. 172) of their writings, and these students spent 4 to 18 h (divided into several sessions) explaining what they were thinking about as they wrote. Raimes (1985) also examined concurrent think-aloud data collected from eight unskilled (determined by holistic measures of their essays) ESL students who wrote a narrative within a 65-min class period. A much larger scale study was Cumming’s (1989) investigation of 23 French-speaking college students’ English writing processes using their written texts and think-aloud data. The study was notable for (a) its application of multivariate statistical analyses, which was made possible by the relatively large sample size, (b) its comparison of the students’ writing processes for three different tasks (letter writing, summary, and argumentation, 1 to 3 h each), and (c) the introduction of controlled variables of L1 writing expertise and L2 writing proficiency. Using the participants’ decision statements in the think-aloud data, Cumming especially focused on four aspects of writing the students attended to while writing (language use, discourse organization, gist, and procedure for writing) and five categories of problem-solving behaviors (heuristic searches with and without resolution, problem resolution, problem identification, and knowledge telling). More recently, Bosher (1998), basically using Cumming’s (1989) coding systems, examined the L2 writing processes of three Southeast Asian ESL college students with different educational backgrounds. Bosher’s study was unique in that she used, as alternative data to think-aloud protocols, stimulated retrospective protocols collected from the participants who recalled their writing processes while watching their own videotaped writing behaviors.

These previous studies that examined part of or the entire process of L2 writing commonly found that (a) skilled L2 writers were similar to their L1
counterparts in that they tended to plan more, revise more at the discourse level, and spend more time on the given task while having their ideas and the generated forms interact freely (e.g., Cumming, 1989; Raimes, 1987; Zamel, 1982, 1983); (b) unskilled L2 writers were similar to their L1 counterparts in that they tended to plan less and revise more at the word and phrase level (e.g., Raimes, 1985, 1987; Zamel, 1983), but they were different from their L1 counterparts in that they were relatively less concerned about surface level revisions (e.g., Raimes, 1985, 1987) and in that they showed more commitment to the given assignment (e.g., Raimes, 1985, 1987); (c) there might be a “writing expertise,” which is independent of L2 proficiency, affecting L2 writing (e.g., Bosher, 1998; Cumming, 1989; Raimes, 1985, 1987); (d) students’ attention patterns and problem-solving behaviors while writing differed according to their L1 writing expertise and the type of tasks they were involved in (e.g., Cumming, 1989); and (e) using L1 while writing in L2 can improve the quality of the L2 writing (e.g., Lay, 1982).

Because researchers have realized that L2 writers’ strategies are similar to those used for L1 writing, many studies from the late 1980s on have also compared the same participants’ L1 and L2 writing processes. It should be noted that here again think-aloud protocol data were the main data sources of analysis for most studies. In these studies, the participants’ L2 is mostly English (but see Cumming et al., 1989 and Whalen & Ménard, 1995 for exceptions), but their L2s greatly vary. For example, Jones and Tetroe (1987) compared six college-level Spanish-speaking ESL students’ planning behaviors while these students wrote two English and one Spanish descriptive expositions. Arndt (1987) compared six Chinese postgraduate EFL students’ processes of writing expositions in L1 and L2 (completed within 1 h each). Similarly, Skibniewski (1988) compared three college-level Polish EFL students’ processes of writing expository essays in L1 and L2. In contrast to Jones and Tetroe’s or Arndt’s study, Skibniewski could compare the differential effects of writing expertise on the three students’ L1 and L2 writing processes because they had distinctly different L2 writing skills (they were identified as skilled, average, and unskilled). Similarly, Cumming et al. (1989) employed 14 English-speaking college students with different writing expertise to compare their summary writing processes in English and French. Using Cumming’s (1989) coding scheme, Cumming et al. specifically focused on the participants’ problem-solving behaviors. Finally, two more recent and larger scale studies were completed by Uzawa (1996) and Whalen and Ménard (1995). Whalen and Ménard analyzed 12 English-speaking participants’ planning, evaluation, and revision strategies at three different levels of discourse (pragmatic, textual, and linguistic) while writing argumentative texts in their L1 and L2 (French) within a maximum of a 2-h period for each. On the other hand, Uzawa compared 22 Japanese ESL students’ processes of writing first drafts of descriptive expositions in Japanese (30 min) and in English (1 h), as well as their processes of translating a magazine article from Japanese into English (1 h).
addition to comparing overall characteristics of each writing process, Uzawa compared attention patterns employed for the three types of writing.

In spite of the conspicuous individual differences reported by many of these comparative studies (e.g., Arndt, 1987), they also have commonly found that (a) L1 and L2 writing strategies, whether the writers were skilled or unskilled, were basically similar, which indicates that L1 writing strategies can be transferred to L2 writing (e.g., Arndt, 1987; Cumming et al., 1989; Jones & Tetroe, 1987; Moragne e Silva, 1988; Skibniewski, 1988; Uzawa, 1996; Whalen & Ménard, 1995); (b) compared with their L1 writing processes, students’ L2 writing processes, especially the higher-order cognitive operations, were negatively affected by lower linguistic proficiency (e.g., Moragne e Silva, 1988; Whalen and Ménard, 1995); and (c) the quality of written L2 texts is more strongly associated with the quality of the students’ L1/L2 writing strategies rather than with their L2 proficiency (e.g., Cumming et al., 1989; Jones & Tetroe, 1987; Whalen & Ménard, 1995).

Although these previous studies have made significant contributions to the field, they are also limited in several ways. First, they investigated mainly ESL learners whose educational backgrounds were typically heterogeneous, and whose L2 proficiency was high enough so that they could receive their education in their L2. Even when EFL learners were examined, their L2 proficiency tended to be high (e.g., Arndt, 1987; Skibniewski, 1988). Second, even though some studies included “skilled” versus “unskilled” contrasts, virtually no studies have included a “novice” versus “expert” contrast where “expert” were those who used L2 writing for professional purposes.¹ Furthermore, many previous studies have employed cross-sectional designs only, lacking developmental perspectives. Including multiple perspectives where novice writers are compared with expert writers as their ultimate possible goals of achievement, or where the novice writers are compared before and after a certain period of writing instruction with other intervening variables controlled, is crucial for building a more comprehensive and dynamic model of L2 writing processes.

Another limitation of the previous studies of L2 writing processes is their almost exclusive use of think-aloud protocols as the main data source (but see the above description of Bosher, 1998 as an exception). Although collecting concurrent verbal reports is an effective way to obtain real-time data on the participants’ writing processes (Ericsson & Simon, 1993), it entails various inherent problems (Smagorinsky, 1994), some of which are especially relevant to the present study. First, it is very difficult for some potential participants to produce “think-aloud” data while writing in L2. It appears even more difficult when they are asked to speak in their L2 (e.g., Raimes, 1985, 1987) because many L2 writers often think in their L1 while writing (e.g., Cumming, 1989; Cumming et al., 1989; Uzawa, 1996). Moreover, even when participants were allowed to speak in any language, some expressed difficulty with the task. For example, Whalen and Ménard (1995), who seem to have allowed the participants to choose the language they spoke in, admitted that 10 potential participants (compared to the 12 who actually produced
the data for the study) could not perform this difficult task, and were thus excluded from the study. Finally, even if researchers can manage to obtain analyzable data from participants (see Hayes & Flower, 1980, p. 9, characterizing the nature of analyzing protocol data as “following the tracks of a porpoise”), there is always the danger of “reactivity.” Previous empirical studies (e.g., Janssen, van Waes, & van den Bergh, 1996; Stratman & Hamp-Lyons, 1994) have reported that the think-aloud condition appeared to have significantly affected the quality and content of the participants’ cognitive activities while writing. If we hope to generalize the results of our studies to a larger population in more natural L2 writing situations, it seems important to employ other data collecting methods that can provide access to both the aspects of writing processes and groups of participants that could not be studied using think-aloud procedures.

The present study was thus motivated by these limitations of the previous studies. It examined writing behaviors of three types of writers (professional, and more- and less-skilled student writers) with similar cultural and educational backgrounds, both cross-sectionally and longitudinally (over 6 months), using multiple data sources collected through a less disruptive method than the think-aloud technique. The method has been successfully used for building models of Japanese L1 writing (e.g., Anzai & Uchida, 1981; Handa, 1985). It was similar to the one used in Bosher (1998) in that the participants produced recall protocols while watching their videotaped writing behaviors, but it was different from Bosher’s method in that the participants could choose the language(s) in which they produced the protocols, and in that the data were coded by a coding scheme specifically developed for this type of data. The data analyzed in the present study included the participants’ written texts, their pausing behaviors while writing, stimulated recall protocols, and analytic scores given to the written texts.

The present study also complements the results of two preceding product-oriented studies (Hirose & Sasaki, 2000; Sasaki & Hirose, 1996). Sasaki and Hirose investigated factors that might influence the quality of 70 Japanese university students’ English expository writing and found that the three factors of students’ L2 proficiency, L1 writing ability, and metaknowledge of L2 expository writing (e.g., how to achieve unity and coherence in a paragraph) significantly explained the students’ L2 writing ability variance. We also found that good writers were significantly different from weak writers in terms of their attention to overall organization while writing in L1 and L2, writing fluency in L1 and L2, confidence in L2 writing for academic purposes, and experiences of regularly writing more than one paragraph in an L2 in high school. Based on these results, Hirose and Sasaki further examined the teachability of two of these explanatory factors, metaknowledge of L2 writing and regular L2 writing experiences. Thus, one group of students \( n = 43 \) with a similar background to that of Sasaki and Hirose was taught metaknowledge in their English writing classes over 12 weeks whereas the other group \( n = 40 \) was assigned to keep English journals at least 4 days a week in addition to receiving the same
metaknowledge instruction as given to the first group in their English writing classes over 12 weeks. In this longitudinal study, we found that teaching the metaknowledge to the students significantly improved their metaknowledge, but not their L2 writing ability in general. In contrast, the instruction of metaknowledge combined with regular journal writing significantly improved mechanical aspects (e.g., spelling, punctuation, capitalization, paragraphing), but not the overall quality of their L2 writing. The results of these product-oriented studies have their own importance, but the question of how these results came about still remains. The present study is an attempt to answer this question, although in a rather exploratory manner. It thus should be considered a precursor to a more confirmatory study with a larger sample size for the purpose of eventually building an empirical model of EFL writing processes.

**METHOD**

**Participants**

Three different groups of Japanese EFL learners (a total of 12) participated in the present study: the expert writer group \((n = 4)\), the more-skilled student writer group \((n = 4)\), and the less-skilled student writer group \((n = 4)\). The more- and less-skilled writer groups were also combined to form a group of “novice writers” \((n = 8)\). These four groups were compared both cross-sectionally and longitudinally in three pairs: the expert writer group versus the novice writer group, the more-skilled student writer group versus the less-skilled student writer group, and the novice writer group before and after 6 months (two semesters) of process writing instruction (see Table 1).

**Expert Writers Versus Novice Writers**

The experts were all professors of applied linguistics (two men and two women) with a mean age of 40.5 years. Their research interests included educational philosophy applied to EFL education and L1/L2 reading and writing (only one of them specialized in L2 writing). Although two of them had earned

<table>
<thead>
<tr>
<th>TABLE 1</th>
<th>The Relationship Among the Participant Groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experts ((n = 4))</td>
<td>Novices ((n = 8))</td>
</tr>
<tr>
<td>Novices I before instruction ((n = 4))</td>
<td>Novices II after instruction ((n = 4))</td>
</tr>
</tbody>
</table>
graduate degrees (two M.A.s and one Ph.D.) in American universities, they had learned English mainly through formal education in Japan until they graduated from universities. They were judged as “expert EFL writers” because their professional work included regularly writing English research papers while their life was anchored in Japan, a non-English environment.

The novices were all 18-year-old college freshmen (four men and four women), majoring in British and American studies. They had studied English for 6 years through formal education in Japan. None of them had been abroad longer than 1 month. They were judged to be “novice writers” because the results of a background questionnaire (see Appendix of Sasaki & Hirose, 1996) indicated that none of them had received much L2 writing instruction, including instruction on matters such as “organizing a paragraph centered on one main idea” or “developing a paragraph so that the readers can follow it easily.” Based on the results of an argumentative writing assignment that was different from the task used for the present study, half of the students (two men and two women) were randomly selected from the top one-third of a sample of 45 students (two classes combined), and the other half (two men and two women) from the bottom one-third (the mean of the summed scores by two raters according to Jacobs, Zinkgraf, Wormuth, Hartfiel, & Hughey, 1981, ESL Composition Profile was 152.3 out of the possible total of 200, and the SD was 9.5 whereas the mean of the bottom group was 95.3 and the SD 15.7).

More-Skilled Versus Less-Skilled Student Writers

The four novices with relatively high writing ability were also compared, as more-skilled student writers, with the other four with relatively low writing ability as less-skilled student writers. Although the two groups were selected on the basis of their composition scores as explained above, their overall English proficiency also turned out to be different. The more-skilled group’s mean score on the Comprehensive English Language Test for Learners of English (CELT) (Form A; see Harris & Palmer, 1986) was 163.5 (SD = 17.25) whereas the less-skilled group’s mean score was 126.5 (SD = 11.39). These scores indicate that the students’ English proficiency ranged from low- to mid-intermediate, which was somewhat lower than the participants’ proficiency level in the preceding study by Sasaki and Hirose (1996).

Novice Writers (Novices I) Versus the Same Writers After 6 Months of Process Writing Instruction (Novices II)

The eight novice writers were also compared before and after 6 months (two semesters interrupted midway by a 2-month summer vacation) of process writing
instruction. Although the instruction was basically intended to employ a “process approach” (Silva, 1990, p. 15), it also incorporated some “current–traditional” (Silva, 1990, p. 13) aspects because metaknowledge that could be acquired through a current–traditional approach had been found to be significant in explaining L2 writing quality of a similar group of students (Sasaki & Hirose, 1996). The eight novice students in the present study received instruction on paragraph writing in English with 37 other students once a week for 90 min. The instructor (the researcher) taught them process writing strategies such as planning and revising, based on Bereiter and Scardamalia’s (1987, p. 245) ideas of “Promoting the development of mature composing strategies,” using Hashiuchi’s (1995) Paragraph Raitingu Nyuumon [Introduction to Paragraph Writing], a composition textbook with special emphasis on process writing. Thus, in the first class, the students were told that writing is an interactive process between what they write and what they want to write, and that such a process is cyclical starting with planning and followed by writing and revising. Furthermore, in each chapter of the textbook, the students first learned rhetorical patterns such as comparison, cause–effect, and expressing opinions, and then were instructed to write a similar paragraph themselves. Before they started to write, they discussed the following points with their instructor and peers:

1. What is the purpose of this writing?
2. Who are the readers of this writing?
3. What content should I include to make it more effective?
4. How should I express the content to make it more effective?

After they wrote the first drafts, most drafts were read by their peers or the instructor. When time allowed, these readers gave comments on which points of the writing were interesting or appealing and which were difficult to understand. Then after receiving the returned comments, the students were told to revise their writing with the idea of unity and cohesion in mind (see Appendix A for the actual questions the students were asked to consider during their revision).

Data Collection Procedure

I followed a slightly revised version of the data collection technique used by Anzai and Uchida (1981). The technique was originally developed for investigating Japanese children’s L1 writing process as an alternative data collection method to the concurrent think-aloud procedure. In Anzai and Uchida, 10 (five male and five female) second-, third-, fifth-, and sixth-grade children (a total of 40) individually wrote a one-page composition (about 400 characters) entitled “Friends” with a female observer carefully recording their writing behaviors, including pauses longer than 2 s, and the time lengths of these behaviors. In addition, on two occasions, the observer asked each child how much planning
and what kind of plan s/he had made before starting to write; Just after these children started to write the first word and just after they finished writing the whole text (the questions were worded slightly differently each time). Although this intervention may have distracted the children during their writing, Anzai and Uchida explained that they did so for the purpose of checking the consistency of the children’s accounts given on two different occasions. Asking the questions on the first occasion was also useful for obtaining data on the children’s thinking process before starting to write when their memory was still fresh. After the second question session, the same observer asked each child what s/he was thinking about during the pauses that the observer had identified while s/he was writing. I decided to use the technique for the present study after having found in a pilot study with five participants from a similar sample population that it could be successfully used to investigate Japanese EFL students’ writing processes. Based on the results of this pilot study, however, I decided to videotape the participants’ hand movements on the paper while writing, as an additional aid for them to remember what they were thinking about at each pause they made in their writing (see Bosher, 1998; DiPardo, 1994).

Another EFL writing specialist and I collected the data. The four experts and the eight students wrote an argumentative composition according to Prompt 1:

**Prompt 1:** There has been a heated discussion about the issue of school uniforms in the readers’ column in an English newspaper. Some people think that wearing high school uniforms is a good custom, whereas others believe high school students should be given an opportunity to choose what they wear. Now the editor of the newspaper is calling for readers’ opinions. Suppose you are writing for the readers’ opinion column. Take one of the positions described above, and write your opinion within 30 minutes (Original in Japanese; see Appendix B).

The eight student writers then wrote according to Prompt 2 after the 6-month instruction period ended:

**Prompt 2:** There has been a heated discussion about the custom of non-Christian Japanese celebrating Christmas in the readers’ column in an English newspaper. Some people think that it is a good custom, whereas others believe we should abandon such a custom. Suppose you are writing for the readers’ opinion column. Take one of the positions described above, and write your opinion within 30 minutes (Original in Japanese; see Appendix B).

I selected argumentative topics because they had been used in the two preceding studies with similar populations (e.g., Hirose & Sasaki, 2000; Sasaki
& Hirose, 1996), and because they could be expected to demand “more complex processing” (Grabe & Kaplan, 1996, p. 121) than other types of writing (e.g., narratives), and thus were expected to distinguish the experts from the novices more clearly. All compositions were scored by two EFL writing specialists independent of the present study, according to Jacobs et al.’s (1981) English Composition Profile. In order not to make the raters aware of which topic was written first, all compositions were rated on one occasion after having been completely mingled with no indication of when they were written, or which participant groups they belonged to. The inter-rater correlation (Pearson correlation coefficients) for the content subscore was 0.91, the organization subscore 0.90, the vocabulary subscore 0.86, the language use subscore 0.88, the mechanics subscore 0.66, and the total score 0.91. I judged that these correlations were acceptable for this small sample.

To collect the data, we asked the participants to come individually to a room and write the compositions in a quiet atmosphere. We asked them to finish writing within about 30 min but told them they could take more time if they wanted to. Consequently, some of them took longer than 30 min to finish, but all of them finished writing within 1 h. The relatively short time allocation was chosen for two reasons. First, more than 80 percent of the 45 students from among whom the present student participants were chosen finished a similar task on a different occasion within 30 min when they were given unlimited time to finish. Second, the results of the pilot study showed that setting the 30-min time limit would be useful for making the following video-watching session manageable short (approximately 2 h). It was found in the pilot study that the video-watching session could make participants too tired to think properly if it exceeded 2 h. Furthermore, the participants were not allowed to use dictionaries when they wrote because the pilot study results showed that when they were allowed to use dictionaries, many of the less proficient students ended up just copying sentences directly from dictionaries instead of using them more appropriately.7

Before the participants started to write, we obtained permission to videotape them while they were writing. We then began to videotape them with the camera mainly focused on their hands and pencil movement. As in Anzai and Uchida (1981), we waited until the participants started to write the first words of the composition before asking them several questions about their planning, such as what they were trying to write at that time, and whether they had decided what they were going to write in the beginning, in the middle, and in the end. The participants were not told that they would be questioned right after they started writing. When they answered the questions, they were told that their answering time would not be included in the 30 min allocated for writing. After the first question session, we let them continue writing until they finished. Right after they finished, the participants were again asked, with slightly different wording, whether they had planned the beginning, the
middle, or the ending part of their composition before they started to write down the first word.

After this second question session, the researcher and each participant together watched the participant’s writing process on videotape. On the videotape, every time the participants stopped writing for a period longer than 3 s, the researcher asked them to explain, either in Japanese or English what they had been thinking about. This continued until they had finished the entire process of writing on the tape. All participants gave their accounts in Japanese except for some English expressions used or considered for use in the compositions (see the examples below). As for the clarity of their videotaped behaviors, in most cases we had a clear view of which part of the texts the participants were working on, which helped the participants’ recall a great deal. The participants’ accounts were all tape-recorded and subsequently transcribed.

From such writing sessions, I obtained and analyzed the following data:

1. Written texts and drafts
2. Time spent before starting to write
3. Occurrences of pauses at the “Episode” Boundaries
4. Participants’ retrospective accounts while watching their videotaped writing performance of what they were thinking about when they stopped writing (“pauses” longer than 3 s).

Coding of the Protocol Data

I basically followed Anzai and Uchida’s (1981) coding system to encode the participants’ retrospective protocol data. Based on their careful and well-designed empirical study, Anzai and Uchida identified six categories (planning, retrieving, generating ideas, verbalizing, rereading, and others) and several subcategories for classifying the participants’ writing strategies (see a detailed description in Appendix C). I added the two additional categories of “translating” and “evaluating” to suit the particular nature of the present data. It is noted here that “translating” in the present study was operationally defined as “expressing the generated ideas (whether it may be in the form of L1 or not) into L2,” not as “expressing the generated ideas into L1” as was used in Hayes and Flower’s (1980) L1 writing model. Because the distinctions among the subcategories under the same category were often important for characterizing each group’s strategy use (e.g., the difference between “global planning” and “local planning”), the present analysis mainly focused on the level of subcategories for analyzing the data.

Before I coded all data, data randomly selected from 3 out of the 20 protocols (15% of the total sample population) were coded by another trained applied linguist using the same categories. The inter-coder agreement was 0.91 (75 out of 82 coding decisions on strategy classifications). Having judged that the agreement rate was acceptable, I coded the rest of the data myself.
Dividing the Written Texts into Episodes

In order to determine the relative locations of pauses that the participants made while writing, the written texts were divided into “episodes” (see Appendix D for examples of the episode boundaries). Following Anzai and Uchida (1981, p. 42), an episode was operationally defined as “a semantically coherent chunk consisting of one or more sentences.” I divided all texts into episodes after having established reliability (32 out of 37 decisions on episode boundaries; 86.5%) with another trained applied linguist on seven texts selected randomly out of the 20 (35% of the total sample).

RESULTS AND DISCUSSION

Because the sample size of the present study was small, applying parametric statistical procedures to the results was not appropriate. I applied nonparametric procedures using SPSS Version 6.1 (SPSS, 1994) when applying them seemed appropriate and necessary. The other results are presented for descriptive purposes only.

Reliability of the Protocol Data

It was found that the participants’ responses to the same questions asked on two different occasions (immediately after they wrote down the first word, and after they finished the entire composition) were not different in content although some of the participants’ responses were more elaborate after they had finished writing as in Example 1.

Example 1: Excerpts from Novice 8’s accounts about whether she had decided what she was going to write in the end

*Just after she started to write the first word of the composition:*

Miyuki (the researcher): Have you decided what you are going to write in the end?
Novice 4: No.
Miyuki: You haven’t decided it yet?
Novice 4: No.

*Just after she finished writing the composition:*

Miyuki: Is the last part of your composition what you wanted to write before you started to write?
Novice 4: No, I wrote it on the spot.
Miyuki: Yes?
Novice 4: I wrote what came up in my mind (when I wrote the last part).

Thus, their accounts were judged to be internally consistent, and thus reliable for further analyses.

Composition Scores

Table 2 presents the means and standard deviations for the composition scores (according to Jacobs et al.‘s 1981 ESL Composition Profile) for the experts and the Novices I and II (after the 6 months of instruction). There was clearly a great gap between the quality of the experts’ composition and that of Novices I who wrote for the same prompt. In fact, the results of the Wilcoxon Mann Whitney test indicate that the experts’ subscore for content, organization, vocabulary, and language use as well as the total score (z values for these subscores and total score were $-2.66$, $-2.68$, $-2.71$, $-2.68$, and $-2.65$, respectively) were significantly different from those of Novices I even after the $\alpha$ level was adjusted to 0.0083 by a Bonferroni correction for multiple comparison (Tabachnick & Fidell, 1996).

As for the differences between Novices I and II, the novice students appear to have made some improvement in terms of vocabulary and language use (because the prompts for Novices I and II were different, it was not possible for the gain made in the content subscore to be seen necessarily as “improvement”). This improvement was in fact a little larger than that of Hirose and Sasaki’s (2000) participants who made a significant improvement only in terms of mechanics as the result of a 3-month period of metaknowledge plus journal writing instruction. However, the results of the Wilcoxon Matched-Pairs

<table>
<thead>
<tr>
<th>Variable (Total Possible)</th>
<th>Experts $M(SD)$ ($n=4$)</th>
<th>Novices I $M(SD)$ ($n=8$)</th>
<th>Novices II$^b$ $M(SD)$ ($n=8$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total (200)</td>
<td>190.00(3.56)</td>
<td>130.71(17.26)</td>
<td>136.5(13.47)</td>
</tr>
<tr>
<td>Content (50)</td>
<td>57.75(1.5)</td>
<td>41.43(5.56)</td>
<td>43.75(5.28)</td>
</tr>
<tr>
<td>Organization (40)</td>
<td>38.50(0.58)</td>
<td>27.57(5.41)</td>
<td>27.31(3.67)</td>
</tr>
<tr>
<td>Vocabulary (40)</td>
<td>38.00(0.82)</td>
<td>25.57(5.41)</td>
<td>27.00(2.33)</td>
</tr>
<tr>
<td>Language Use (50)</td>
<td>46.50(0.58)</td>
<td>30.57(5.71)</td>
<td>32.63(2.83)</td>
</tr>
<tr>
<td>Mechanics (10)</td>
<td>9.25(1.50)</td>
<td>7.00(0.82)</td>
<td>7.25(0.71)</td>
</tr>
</tbody>
</table>

Notes:

$^a$ These scores are sums of the two raters’ scores.

$^b$ Novices I and II are the same students before and after the 6 months of instruction (see Table 1).
Signed-Ranks test indicate that none of these differences were statistically significant (the $z$ values for organization, vocabulary, language use, mechanics subscores, and the total scores were $-0.63$, $-0.97$, $-1.53$, $-0.94$, $-1.41$, and $-1.02$, respectively). Although the findings suggest that process-writing instruction like that given in the present study may have the potential to improve students’ L2 writing, the true effect of such instruction should be investigated in a further study with a larger sample size.

**Fluency**

The participants’ writing fluency was measured in terms of the two indices of mean total number of words written in the texts (quantity) and mean number of words written per minute (speed; see Table 3). The results of the Wilcoxon Mann Whitney test indicate that the experts wrote significantly longer texts, and wrote significantly faster than the novices ($z = -2.65, p = .008$ for both the quantity and the speed; the $\alpha$ level was adjusted to 0.025 by a Bonferroni correction for multiple comparison). In other words, the experts were much more fluent than the novices. Similarly, the more-skilled students appear to be more fluent than the less-skilled students although the differences were not large enough to be statistically significant ($z = -1.77, p = .08$ for the quantity; $z = -0.35, p = .72$ for the speed). As has been found in previous studies (e.g., Hirose & Sasaki, 1994; Sasaki & Hirose, 1996), such a difference in fluency was probably caused by the difference in their L2 proficiency (see also the Writing Strategies section below). In contrast, after 6 months of instruction, the novices as a whole wrote neither a greater quantity nor faster. This suggests that process writing instruction of such a span may not help students improve their writing fluency.

**Time Spent Before Starting to Write**

The experts spent a longer time ($M = 6$ min and $37$ s; 20.5% of the total writing-task time) before starting to write than the novices ($M = 4$ min and $37$ s; 13.6% of the total writing-task time).

<table>
<thead>
<tr>
<th>TABLE 3</th>
<th>Two Indices of Fluency for Each Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Experts</td>
</tr>
<tr>
<td>Mean total number of words written</td>
<td>269</td>
</tr>
<tr>
<td>Mean number of words written per minute</td>
<td>8.98</td>
</tr>
</tbody>
</table>

Notes:

\textsuperscript{a} Because one of the participants lost her text, the remaining seven texts were examined here.  
\textsuperscript{b} The more- and less-skilled students were the novices before the 6 months of instruction began.  
\textsuperscript{c} Because one of the participants lost her text, the remaining three texts were examined here.
12.4% of the total time). Similarly, the more-skilled students spent a longer time ($M = 7$ min and 35 s; 18.3% of the total time) than the less-skilled students ($M = 1$ min and 39 s; 4.6% of the total time). After the 6 months of instruction, the eight novice students as a whole spent a longer time ($M = 7$ min; 20.2% of the total time) before starting to write than previously ($M = 4$ min and 37 s). These results are consistent with those of Bereiter and Scardamalia (1987, Chap. 1), who investigated the relationship between L1 writing expertise and the time spent before the participants started to write. It appears that, as in the case of L1 writing, planning time is related to L2 writing expertise. Moreover, the present results suggest that instruction can influence students’ planning time.

**Occurrences of Pauses at the Episode Boundaries**

Table 4 presents the mean total number of “episodes” that appeared in the five groups’ composition texts, and the mean ratios of pauses that occurred at the episode boundaries. The locations of the pauses were identified using both the videotaped writing processes and the protocol data. The total number of episodes, together with the total numbers of words used (see Table 3), indicate one aspect of complexity of the participants’ compositions (Anzai & Uchida, 1981). It appears that the experts wrote texts with more complex development than the novices and that the more-skilled students wrote more complex compositions than the less-skilled students. After 6 months of instruction, the novices as a whole were able to write compositions with a greater number of episodes.

As for the ratio of pauses occurring at the episode boundaries, the novices tended to pause at each boundary more often than the experts. In fact, the less-skilled students stopped at each episode boundary without fail. This suggests that the novices tended to stop and think about what they were going to write each time they finished writing one coherent chunk, whereas the experts tended to plan more globally. It appears that the novices employed a kind of “what

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean total number of episodes</th>
<th>Mean ratio of pauses that occurred at the episode boundaries (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experts ($n = 4$)</td>
<td>7.25</td>
<td>43.8</td>
</tr>
<tr>
<td>Novices I ($n = 8$)</td>
<td>4.14</td>
<td>93.1</td>
</tr>
<tr>
<td>More-skilled studentsa ($n = 4$)</td>
<td>4.75</td>
<td>88.8</td>
</tr>
<tr>
<td>Less-skilled studentsa ($n = 4$)</td>
<td>3.33</td>
<td>100</td>
</tr>
<tr>
<td>Novices II ($n = 4$)</td>
<td>5.88</td>
<td>55.3</td>
</tr>
</tbody>
</table>

*Note:* a The more- and less-skilled students were subgroups of Novice I.
next strategy” (Bereiter & Scardamalia, 1987), which was employed by many inexpert L1 and L2 writers in previous studies (e.g., Cumming, 1989; Grabe & Kaplan, 1996; Uzawa, 1996). After 6 months, however, the novices of the present study, as a whole, did not stop at the episode boundaries as often as before. This suggests that learners’ writing styles can be altered with 6 months of instruction.

### Writing Strategies

#### Frequencies

Table 5 presents the mean total number of writing strategies (token frequency) used by each group (see Appendix C for the description of these strategies). The participants used about 30 to 35 strategies on average for completing the task, except for the Novices II (after 6 months of instruction) group. The fact that both the experts and novices used a similar numbers of strategies incidentally suggests that the method employed in the present study was capable of eliciting data from these two groups equally well. This may suggest the superiority of this method over the think-aloud method that has been criticized for eliciting a greater amount of data from better writers or experts because they also tended to have better verbal ability (Flower et al., 1992; Hayes & Nash, 1996).

Another thing to be noted in Table 5 was that the standard deviation for the experts (23.57) was much larger than those for other groups. This indicates that there were greater individual differences among the experts than among the novices. Such individual differences could be due to the relatively smaller sample size of the expert group, but it might also mean that each expert tended to have unique patterns of strategy use that were different from those of other experts. Cumming (1989) reported a similar result in that the experts in his study differed in strategy use according to the fields they specialized in.

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Experts (n = 4)</td>
<td>32</td>
<td>23.57</td>
</tr>
<tr>
<td>Novices I (n = 8)</td>
<td>35.75</td>
<td>13.46</td>
</tr>
<tr>
<td>More-skilled students (n = 4)</td>
<td>36.25</td>
<td>11.44</td>
</tr>
<tr>
<td>Less-skilled students (n = 4)</td>
<td>35.25</td>
<td>17.06</td>
</tr>
<tr>
<td>Novices II (n = 8)</td>
<td>18.25</td>
<td>8.41</td>
</tr>
</tbody>
</table>

**Notes:**

- The total numbers of strategies means the total number of tokens of the strategies used. Thus, if the same strategy was used three times, it was counted as three instead of one.
Finally, after the 6 months of instruction, the number of strategies used by the novices decreased by almost half, from 35.75 to 18.25 (the difference was statistically significant; $z = -2.24$, $p = .003$ by the Wilcoxon Matched-Pairs test).

**TABLE 6**
Relative Distribution of Use of the Eight Writing Strategies

<table>
<thead>
<tr>
<th></th>
<th>Experts, $n$ (%)</th>
<th>Novices I, $n$ (%)</th>
<th>More-skilled students, $n$ (%)</th>
<th>Less-skilled students, $n$ (%)</th>
<th>Novices II (After a 6-month instruction), $n$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning total</td>
<td>7.8 (23.9)</td>
<td>11.4 (34.2)</td>
<td>9.8 (28.7)</td>
<td>13 (39.7)</td>
<td>5.8 (31.4)</td>
</tr>
<tr>
<td>Global planning</td>
<td>1.3 (8.7)</td>
<td>0.1 (0.4)</td>
<td>0.3 (0.7)</td>
<td>0 (0)</td>
<td>0.9 (7.3)</td>
</tr>
<tr>
<td>Thematic planning</td>
<td>0.3 (0.4)</td>
<td>1.1 (3.4)</td>
<td>1 (2.9)</td>
<td>1.3 (3.8)</td>
<td>0.3 (1.0)</td>
</tr>
<tr>
<td>Local planning</td>
<td>4.8 (11.5)</td>
<td>7.5 (22.6)</td>
<td>6 (17.8)</td>
<td>9 (27.4)</td>
<td>3.9 (19.5)</td>
</tr>
<tr>
<td>Organizing</td>
<td>1.3 (2.6)</td>
<td>0.4 (1.1)</td>
<td>0.8 (2.1)</td>
<td>0 (0)</td>
<td>0.1 (0.4)</td>
</tr>
<tr>
<td>Conclusion planning</td>
<td>0.5 (0.7)</td>
<td>2.3 (6.9)</td>
<td>1.8 (5.1)</td>
<td>2.8 (8.6)</td>
<td>0.6 (3.3)</td>
</tr>
<tr>
<td>Retrieving total</td>
<td>1.3 (2.9)</td>
<td>2.4 (6.4)</td>
<td>4 (10.2)</td>
<td>0.8 (2.6)</td>
<td>0.8 (3.5)</td>
</tr>
<tr>
<td>Plan retrieving</td>
<td>1.3 (2.9)</td>
<td>1.3 (3.9)</td>
<td>2 (6.2)</td>
<td>(0.5)</td>
<td>0.8 (3.5)</td>
</tr>
<tr>
<td>Information retrieving</td>
<td>0 (0)</td>
<td>1.1 (2.5)</td>
<td>2 (4.0)</td>
<td>(0.3)</td>
<td>1.0 (0)</td>
</tr>
<tr>
<td>Generating ideas total</td>
<td>0.3 (0.7)</td>
<td>0.8 (2.2)</td>
<td>1.5 (4.4)</td>
<td>0 (0)</td>
<td>0.1 (0.6)</td>
</tr>
<tr>
<td>Naturally generated</td>
<td>0 (0)</td>
<td>0.1 (0.3)</td>
<td>0.3 (0.5)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Description generated</td>
<td>0.3 (0.7)</td>
<td>0.6 (2.0)</td>
<td>1.3 (3.9)</td>
<td>0 (0)</td>
<td>0.1 (0.6)</td>
</tr>
<tr>
<td>Verbalizing total</td>
<td>12.3 (49.0)</td>
<td>4.1 (11.8)</td>
<td>6 (17.7)</td>
<td>2.3 (5.8)</td>
<td>3.1 (15.3)</td>
</tr>
<tr>
<td>Verbalizing a proposition</td>
<td>0.5 (0.8)</td>
<td>1.1 (3.1)</td>
<td>1.3 (3.7)</td>
<td>1 (2.5)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Rhetorical refining</td>
<td>11 (46.6)</td>
<td>2.1 (6.2)</td>
<td>4 (12.0)</td>
<td>0.3 (0.4)</td>
<td>2 (10.0)</td>
</tr>
<tr>
<td>Mechanical refining</td>
<td>0.8 (1.5)</td>
<td>0.8 (2.0)</td>
<td>0.8 (2.0)</td>
<td>0.8 (1.9)</td>
<td>1 (4.3)</td>
</tr>
<tr>
<td>Sense of readers</td>
<td>0 (0)</td>
<td>0.1 (0.5)</td>
<td>0 (0)</td>
<td>0.3 (1.0)</td>
<td>0.1 (1.2)</td>
</tr>
<tr>
<td>Translating</td>
<td>2.5 (4.6)</td>
<td>12 (32.2)</td>
<td>9 (25.1)</td>
<td>15 (39.2)</td>
<td>5.4 (32.6)</td>
</tr>
<tr>
<td>Rereading</td>
<td>6 (14.0)</td>
<td>1.9 (5.9)</td>
<td>2 (5.3)</td>
<td>1.8 (6.5)</td>
<td>2.6 (14.0)</td>
</tr>
<tr>
<td>Evaluating total</td>
<td>0.5 (1.1)</td>
<td>2.5 (5.5)</td>
<td>3.5 (7.1)</td>
<td>1.5 (3.9)</td>
<td>0.4 (2.3)</td>
</tr>
<tr>
<td>L2 proficiency evaluation</td>
<td>0 (0)</td>
<td>0.6 (1.4)</td>
<td>0 (0)</td>
<td>1.3 (2.7)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Local text evaluation</td>
<td>0.3 (0.7)</td>
<td>1.5 (3.1)</td>
<td>3 (6.2)</td>
<td>0 (0)</td>
<td>0.3 (0.6)</td>
</tr>
<tr>
<td>General text evaluation</td>
<td>0.3 (0.4)</td>
<td>0.4 (1.1)</td>
<td>0.5 (1.0)</td>
<td>0.3 (1.2)</td>
<td>0.1 (0.5)</td>
</tr>
<tr>
<td>Others total</td>
<td>1.5 (3.9)</td>
<td>0.8 (2.0)</td>
<td>0.5 (1.6)</td>
<td>1 (2.3)</td>
<td>0.1 (0.3)</td>
</tr>
<tr>
<td>Resting</td>
<td>0 (0)</td>
<td>0.4 (0.9)</td>
<td>0.3 (0.9)</td>
<td>0.5 (0.9)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Questioning</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0.3 (0.4)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Impossible to categorize</td>
<td>1.5 (3.9)</td>
<td>0.3 (0.9)</td>
<td>0.3 (0.7)</td>
<td>0.3 (1.0)</td>
<td>0.1 (0.3)</td>
</tr>
</tbody>
</table>

Notes:

* The ratio of each strategy represents the mean of the ratios of that strategy, instead of the ratio of the mean number of the strategy to the total mean number of strategies used. This procedure was employed to avoid the influence of one participant’s excessive use of a particular strategy (i.e., the “weight” of one strategy use varies from participant to participant because the total number of strategies used were different), and to equally reflect each participant’s use of that strategy relative to the total number of the strategies used. Thus, the ratios can differ even when the corresponding mean numbers of the strategies used by the same group were the same (e.g., the ratios of plan retrieving and information retrieving for the more-skilled students).
Signed-Ranks test). This phenomenon can be interpreted in several ways. The students might have learned prescribed ways to approach writing and applied them to this task with less trial-and-error. Or, as can be seen in Table 6, because the substantial part (6.6 out of 17.5, i.e., 37.7%) of this frequency decrease was caused by the decrease of the students’ use of the “translation” strategy (12 for Novices I and 5.4 for Novices II), it might have been related to the students’ English proficiency improvement, which could have enabled them to use translation less frequently. Yet another reason might be that after two semesters of instruction, the students might be thinking less, or with less effort, about their writing while composing, which may not be very desirable as a result of strategy-based instruction (cf. Cohen, Weaver, & Li, 1998). The true reasons for this decrease in strategy use should be investigated in further studies. One clear point here is that we are assured again that students can change their use of writing strategies as a result of instruction, which concurs with the results of previous L1 and L2 studies (e.g., Anzai & Uchida, 1981; Bereiter & Scardamalia, 1987; Cohen et al., 1998).

Table 6 shows the relative distribution of the eight categories and the subcategories used by each group (see Appendix C for detailed explanation of each category with examples). The top five subcategories (or categories with no subcategory) for each group are as follows:

Experts

1. Rhetorical refining, 46.6 percent
2. Rereading, 14.0 percent
3. Local planning, 11.5 percent
4. Global planning, 8.7 percent
5. Translating, 4.6 percent

Novices

1. Translating, 28.6 percent
2. Local planning, 22.6 percent
3. Rhetorical refining, 6.2 percent
4. Conclusion planning, 6.9 percent
5. Rereading, 5.9 percent

More-skilled students

1. Local planning, 17.8 percent
2. Translating, 25.1 percent
3. Rhetorical refining, 12.0 percent
Regarding the experts-versus-novices contrast, we first notice that translating the generated idea into English took a great amount of attention for the novices, but not for the experts. The fact that the novices had to stop and think about translation so often may explain the fluency difference between the experts and novices. Lack of L2 proficiency probably hindered the novices from writing longer and faster. Such a finding conforms with other studies (e.g., Cumming, 1989, Pennington & So, 1993; Sasaki & Hirose, 1996), supporting the hypothesis that L2 proficiency explains a substantial portion of L2 writing competence.

In contrast to the novices, the experts used “rhetorical refining” most frequently. As can be seen in Example 2, the experts often knew English well enough to think of several English expressions to convey their meanings, but they spent time choosing the most appropriate ones (the participants’ accounts in all examples below were originally given in Japanese except for the English words in quotation marks, and the accounts were translated into English by the author).

Example 2: Rhetorical Refining

I was wondering whether I should use “merits and demerits” or “pros and cons” (to express the idea of “good points and bad points”) (Expert 1).

Cumming (1989) reported that L2 writers with L1 writing expertise also demonstrated a similar behavior of paying special attention to word/phrase
choices. The L2 experts’ behavior in the present study exemplified above supports Cumming’s subsequent speculation that such behavior is a manifestation of what Gardner (1983, p. 73) called “linguistic intelligence” that transcends L1 and L2 differences.

Another aspect to note regarding the differences between the experts and novices is that the experts used more global planning (8.7%) than the novices (0.4%) whereas the novices used more local planning (22.6%) than the experts (11.5%). “Global planning” is planning detailed, overall organization of the composition as in Example 3.

Example 3: Global Planning

Well, I have to support one of the given positions, but, to tell you the truth, I am not for or against either of these ideas, but in order to write this (composition) today, I had to decide which position to take. That was the first thing to decide. So I decided which position I would take. Then, I thought about how I could develop my argument supporting my opinion. At this point, I have decided that I will write the merits and demerits of both positions, and will especially emphasize the merits of the position I am supporting. I don’t have concrete ideas of these merits or demerits yet, so I will probably write what comes to my mind as I go along (Expert 1).

This planning contrasts with “local planning,” where the participants plan what to write next, as in Example 4.

Example 4: Local Planning

After I wrote this down, I wondered what I should write next (Novice 1).

It seems that global planning has a stronger relationship with L2 writing ability than local planning in the present study, which is consistent with the results of many previous L1 and L2 writing studies (e.g., Bosher, 1998; Jones & Tetroe, 1987; Lay, 1982; Sasaki & Hirose, 1996).

For the more-skilled-versus-less-skilled student comparison, we again notice that translating ideas took a great amount of attention for both groups, but the more-skilled student writers’ behavior was closer to that of the experts in that the more-skilled students used “rhetorical refining” more often than the less-skilled students. In contrast, there was little difference between the two student groups in terms of global planning; neither of them did much global planning (see Table 6).

For the Novices I-versus-Novices II (after 6 months of instruction) comparison, we can see that translation from L1 to L2 was still a common strategy for Novices II, but we can also see that they learned to use such strategies as
“rereading” and “global planning” after the instruction. Because these strategies were emphasized during the instruction, the results again provide evidence that instruction can affect the students’ use of writing strategies.

Qualitative Results

In addition to identifying the general trends of each group’s strategy use, I also found several behavioral characteristics that could not be captured in the quantitative analyses described above. Those findings were “qualitative” in nature in that they reflected both insiders’ (i.e., the participants’) perspective and the researcher’s interpretation (Ary, Jacobs, & Razavieh, 1990). First, I noticed that both global and local planning strategies sometimes appeared to monitor or guide the participants’ writing process:

Example 5: “Local Planning” Monitoring Novice 5’s Writing

I wrote one sentence as the first reason (for supporting my idea), and then I wrote “That is the first reason.” But I felt that I needed more than one sentence for the first reason, then I remembered other sentences I had thought of for the first reason (when he made the original plan) . . .

In Example 5, Novice 5 (a more-skilled student writer) had originally planned to write more than one sentence for the first reason, and his local planning made him feel that he needed more than one sentence after he had written the first sentence. In other words, his local planning monitored his writing process, and it made him remember that he needed other sentences to support the first reason.

Evidence of such a monitoring function of planning was found not only in the planning strategies themselves, but also in the participants’ “plan retrieving” and “rereading” behaviors:

Example 6: “Plan Retrieving” Monitoring Novice 4’s Writing

And, here I was probably thinking again how I could better express what I was originally thinking to write.

Example 7: “Global Planning” Monitoring Expert 1’s Writing, Which Was Manifested In Her “Rereading”

Here, I was rereading the previous part. (Researcher: Yes.) It seems that I had difficulty organizing ideas for the last part. (Hmmm.) I was wondering how I could make this last part cohere to the previous part.
In Example 6, Novice 4 (a less-skilled student writer) originally planned to express that Japanese students may not cultivate their individuality if they wear school uniforms, but could not express that idea in English. He had written one sentence to express part of the idea, but stopped, realizing that he was not yet able to fully express his original thought. We can see Novice 4’s local plan monitoring his writing process through his comment in Example 6. In contrast, in Example 7, we can see Expert 1 monitoring how well the original plan was carried out while rereading the already written text. It is interesting that these examples all have in common the participants’ feeling that there was a “gap” between what they planned to write and what they could actually express. A similar phenomenon of using plans to monitor the writing process based on this “gap” feeling has been reported in studies of both English and Japanese L1 writing (Anzai & Uchida, 1981; Faigley, Cherry, Jolliffe, & Skinner, 1985; Uchida, 199012).

Another finding of a qualitative nature was related to the experts’ “global planning” behavior. We have seen that this behavior characterizes the writing processes of both the experts and novices II (see Table 6). However, a closer look at the protocols of their global planning revealed that the experts’ writing processes were qualitatively different from those of the novices even after the 6 months of instruction. For instance, in global planning, the experts appeared to try to find the most effective solution based on the assessment of the given working environment, treating the given writing task as a “problem” to be solved. Their “solution” entailed setting one holistic goal followed by several hierarchically structured subgoals to achieve the holistic goal as seen in Example 8 (see also Example 2).

Example 8: Expert 2’s Global Planning

Because this is a writing task, I thought I should choose one of the two opinions, so I decided on one opinion or stance, and made a plan for that. For example, I chose the opinion that we should let students choose what they wear. Then, while I was planning the overall organization, I decided to include the fact that I am a university teacher even though this is a hypothetical task . . .

After this comment, Expert 2 continued to talk in detail about how he made the overall plan, and organized the whole draft, giving many examples. Here, the writer identified the characteristics of the given task while setting a global goal for successfully completing the task. He then set several subgoals to effectively achieve the global goal, trying to make the most of his topic-related and language knowledge. As in the case of Expert 1’s writing process presented in Example 7, his global plan subsequently monitored his writing process throughout the entire progress of his writing.

Furthermore, the experts’ intention to seek the best solution under the given environment was so prevalent that it sometimes even forced the writer to change
the global planning itself, even at the level of the entire organization as seen in Example 9.

Example 9: Expert 1’s Comment Immediately After She Finished Writing

... In the beginning, I wanted to present two or three points supporting my opinion, but while I was writing, the first point got longer than I had expected, so I decided to concentrate on the first point.

It appears that Expert 1 readjusted her global plan according to the environmental change while keeping the original diagnosis in her mind. Such cases of elaborated but flexible goal-setting and assessment operating behind writing processes were not observed in the Novices’ II protocol data even though half of these novices actually made a somewhat detailed global plan before they started to write.

These characteristics of the experts’ global planning are similar to those that Flower et al. (1992, p. 181) called “constructive planning” where “writers must create a unique network of working goals and deal with the special problems of integration, conflict resolution, and instantiation this constructive process entails.” Constructive planning requires the writers not only to possess a sufficient amount of relevant topical knowledge, but also to be able to hierarchically organize ideas drawn from the existing knowledge so that these ideas will be eventually integrated into an effectively coherent whole. Flower et al. (1992) reported that only the L1 expert writers could plan in such a way. Such behavior was not observed even among their better student writers. Flower et al. (1992) explained this phenomenon as “a gulf between the experts and the novices” (p. 235). It appears that such a “gulf” also existed between the L2 writing experts and the novices of the present study, and the gulf was not filled by the 6 months of process writing instruction.

We have seen above that the novices’ compositions became a little more complex after the instruction. We have also seen that they had learned to use some of the more effective writing strategies after the instruction. However, as Grabe and Kaplan (1996, p. 129) emphasized, acquiring the writing expertise exemplified by such behavior as constructive planning would probably require “consistent practice in a variety of similar contexts to the point of proceduralization or automaticity,” which the experts of the present study must have attained through many years of experience. Furthermore, if writing expertise is a type of intelligence as Cumming (1989) speculated, we must expect a great amount of individual difference in its development in response to various types of instruction (Gardner, 1983). Because L2 proficiency, another powerful explanatory variable for L2 writing ability (Sasaki & Hirose, 1996), has also been found to be related to general intelligence among this particular type of EFL students (i.e., those who have learned English through highly controlled
formal education, see Sasaki, 1996), individual differences related to intelligence may be a critical issue for planning L2 writing curricula in this type of educational setting. In any case, the results of the present study suggest what a 6-month long writing class can possibly accomplish in the given EFL situation, and what it cannot.

Summary of the Characteristics of the Participants’ Writing Processes

The findings of the present study can be summarized as follows.

1. The experts wrote longer texts with more complex development at greater speed than the novices.
2. The experts spent a longer time before starting to write, planning a detailed overall organization, whereas the novices spent a shorter time making a less detailed plan.
3. Once the experts made their global plan, they did not stop and think while writing as frequently as the novices. In contrast, the novices tended to stop and plan what they were going to write every time they finished writing one semantically coherent chunk.
4. L2 proficiency or lack of it appears to explain part of the difference in strategy use between the experts and the novices. The novices often stopped to translate the generated ideas into English whereas the experts often stopped to refine their English expression.
5. After 6 months (two semesters) of process writing instruction, neither the quality of the students’ compositions nor their writing fluency appear to have improved. Their relatively low L2 proficiency still constrained their writing speed: They still had to stop to translate often. Although the number of strategies they used decreased by half for some reason, some of the students started to use skilled writers’ strategies such as “rereading” and “global planning.”
6. Both “global planning” and “local planning” monitored/guided the participants’ writing processes.
7. The experts’ global planning and partial adjustment of such planning while writing was based on their elaborated but flexible goal-setting and assessment of the characteristics of the given task for successfully achieving the task. Such behavior appears to be a manifestation of writing expertise that cannot be acquired over a short period of time.

LIMITATIONS AND SUGGESTIONS FOR FURTHER STUDIES

The present study investigated, using multiple data sources, the cognitive processes of various types of Japanese EFL learners while writing an argu-
mentative exposition in English. The method employed has proved to elicit meaningful data from learners with relatively low L2 writing abilities, who have not been able to participate in this type of process writing study before. Although the results of the present study are important as a first exploratory step, they should be complemented by those of further studies. First, the study should be replicated with a larger sample population for the purpose of confirming the results. Moreover, replications are also necessary with different types of writing under different conditions. Using other topics and types of tasks, such as writing a complaint letter or a project report to be completed within a longer span of time with the help of dictionaries, may reveal somewhat different processing phases.

Second, future studies should investigate affective or emotional factors that were not treated in the present study. Affective factors such as motivation or attitude have not been given much attention in previous writing models, but have begun to be recognized as crucial for understanding writing behavior (e.g., Hayes, 1996). If Schumann (1998, p. xv) is correct in saying that “variable success in second language acquisition (SLA) is emotionally driven,” L2 writing should not be an exception. Attempts to add the “affect” component would make the target writing process model more comprehensive.

Finally, related to the second point, the study should also be replicated with other sample populations of different L1, L2, and educational/cultural backgrounds. If we assume that writing cannot be conducted in a social vacuum and that writing is “social construction” (Cumming, 1998, p. 61), we cannot ignore the social/cultural contexts where the writing takes place. We should extend the scope of our research, and investigate such issues as how differently writing is treated in different L1 and L2 cultures, or how such cultural factors affect the learners’ and teachers’ expectations (e.g., Porte, 1997; Shi & Cumming, 1995). Incorporating the results of these studies will surely contribute to enriching the content and usefulness of a comprehensive L2 writing process model we can eventually hope to build.

Acknowledgments: An earlier version of this article was presented at the third Pacific Second Language Research Forum at Aoyama Gakuin University, Tokyo, on March 28, 1998. I would like to thank Keiko Hirose for collecting part of the data, Setsuko Miyamoto for coding part of the data, Junko Okabe, Mariko Nishimoto, and Ryu Itoh for scoring the compositions, and Alister Cumming, Carol Rinnert, and William J. Herlofsky, the editors and anonymous reviewers of the Journal of Second Language Writing for their valuable comments and suggestions. Needless to say, the positions taken and any errors that may remain are solely my responsibility. The preparation of this manuscript has been aided by a research grant for the 1998 academic year from Nagoya Gakuin University.
NOTES

1. Although several studies such as Cumming (1989) included writers with professional experiences, they were experts in L1 writing instead of in L2 writing.

2. This study investigated the L1 writing processes of 10 fifth graders and 10 second-year junior high school students using the method employed by Anzai and Uchida (1981). The results showed that there were no clear age differences in terms of writing strategies used, and that during the pauses the participants repeatedly used five strategies of Planning, Organizing, Translating, Reviewing, and Activating. Based on these results, Handa (1985) proposed a preliminary model of L1 Japanese writing processes.

3. These two raters were EFL writing specialists independent of the present study.

4. In Sasaki and Hirose (1996), the mean CELT score of the “good” writers was 201.15, and that of the “weak” writers was 145.91.

5. Researchers in the field of L1 writing have also found that it was difficult to collect think-aloud data. For example, Uchida (1989) reported that only 2 out of 10 potential participants produced analyzable think-aloud data in a study investigating children’s revising processes in L1 Japanese (see also Uchida, 1986 for discussion of this issue).

6. If these two prompts had been alternated with half of the eight students before and after the instruction (i.e., half of them receiving Prompt 1 before the instruction and Prompt 2 after the instruction with the other half receiving Prompt 2 before the instruction and Prompt 1 after the instruction), I could have avoided possible topic effects on the students’ composition scores and their use of writing strategies. However, if I had alternated the prompts for the Novice I group, I would also have had to alternate the prompts for the Expert group for a fairer comparison. In such a case, I would have had to consider the possible effects of these two different topics on the participants’ use of writing strategies especially when the experts’ writing strategies were compared with those of the novices for one composition written on only one occasion. Previous studies (e.g., Carter, 1990; Cumming, 1989; Flower, Schriver, Carey, Haas, & Hayes, 1992) suggest that writers may change their writing strategy use according to different topics. Thus, I decided not to alternate the prompts for the novices for the particular design of the present study.

7. Performing such a test-like task was not uncommon to these participants, however. Informal interviews after the session revealed that both the expert and novice writers often encountered the task of “writing in a relatively short-period of time” (for the novices, taking in-class tests or the Test of Written English given by the Educational Testing Service; for the experts, summarizing the main points of a paper). Tannenbaum et al. (1996) also reported that Japanese EFL learners would encounter similar tasks.
8. Anzai and Uchida (1981) used pauses longer than 2 s instead of 3 s. Longer pauses were used in the present study because it was concluded that they were the shortest possible pauses that could be handled based on the pilot study results.

9. Unfortunately, one of the novices took her text for Prompt 1 home and lost it. Thus, the remaining three texts were analyzed for Novices I in the results.

10. In order not to confound the planning time with the “time-on-task” (Hayes & Nash, 1996, p. 49) as in many other L1 studies, the ratio of the planning time to the total time spent was calculated for each group.

11. Among these group differences, however, only the difference between the Experts and the Novices I was large enough to be statistically significant ($z = -2.62, p = .009$ by the Wilcoxon Mann Whiney test).

12. This chapter used several observational reports and the results of experiments, and explained how children learn to tell stories and eventually start to write them on paper. The author maintained that the ability to tell stories in L1 is supported by the growth of many cognitive abilities such as the ones to plan, monitor, and evaluate. She also described the differences between speaking and writing processes, and the influence of writing on speaking.

13. These examples were identified as writing tasks that might be required in the field of English for international communication by Tannenbaum et al. (1996).

REPRESENTS


**APPENDIX A.**

Questions Novice Students Were Told to Consider When Revising Their Drafts

**About The Whole Paragraph**

1. Have you written everything you wanted to say?
2. Does your paragraph have unity (are there any sentences that deviate from the main idea)?
3. Does your paragraph have cohesion (are the sentences in the paragraph connected by cohesive devices such as connectors or pronouns)?

**About each sentence**

1. Is it easy for your readers to understand?
2. Is it grammatically correct?
3. Is the content convincing (would it be better if you gave more examples or explanation)?
4. Is it directly related to the main idea of the paragraph?
5. Is it appealing to the readers?
6. Are there any spelling errors?
APPENDIX B.

Original Japanese of the Two Prompts

Prompt 1: ある英字新聞の『読者のページ』で、中学・高校での制服について論争が続いています。論争では、主に、「制服はあった方がよい」という立場と、「学校での服装は自由な方がよい」という立場が対立しています。この2つの立場のうち、どちらかの立場を擁護する投書を、30分程度で英語で書きなさい。

Prompt 2: ある英字新聞の『読者のページ』で、クリスマスにパーティーを開いたり贈り物をするという日本の習慣について論争がついています。論争では、主に、「これは、良い習慣である」という立場と、「この習慣は止めるべきである」という立場が対立しています。もしあなたが、この2つの立場のうち、どちらかの立場を擁護する投書をするとしたら、どのように書きますか。30分以内で意見をまとめて、英文で書きなさい。

APPENDIX C.

Encoded Categories for Students’ Retrospective Accounts of Their Writing Processes

<table>
<thead>
<tr>
<th>Category</th>
<th>Definition</th>
<th>Example (English translation by M. Sasaki)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning</td>
<td>Detailed planning of overall organization</td>
<td>(After saying that she had made a general plan of what she would write in the beginning, in the middle, and at the end of the composition): Well, I will write the “advantages” and “disadvantages” (of wearing school uniforms). And, ah, when I first read this (the prompt), I first wondered which stance I should take, and decided to take the position that the students should be given freedom to choose what they wear. (Researcher: Oh, so ...) So, I will try to write more disadvantages (of requiring school uniforms) than advantages. Then, considering the trends of present-day society, I should generally take this position (supporting the students’ free-dom to choose what they wear). I now feel that I should go with the idea that students should be given freedom to choose what they wear (Expert 3).</td>
</tr>
</tbody>
</table>

*Original Japanese of the Two Prompts: Prompt 1: ある英字新聞の『読者のページ』で、中学・高校での制服について論争が続いています。論争では、主に、「制服はあった方がよい」という立場と、「学校での服装は自由な方がよい」という立場が対立しています。この2つの立場のうち、どちらかの立場を擁護する投書を、30分程度で英語で書きなさい。

Prompt 2: ある英字新聞の『読者のページ』で、クリスマスにパーティーを開いたり贈り物をするという日本の習慣について論争がついています。論争では、主に、「これは、良い習慣である」という立場と、「この習慣は止めるべきである」という立場が対立しています。もしあなたが、この2つの立場のうち、どちらかの立場を擁護する投書をするとしたら、どのように書きますか。30分以内で意見をまとめて、英文で書きなさい。
<table>
<thead>
<tr>
<th>(2) <strong>Thematic planning</strong></th>
<th>Less detailed planning of overall organization</th>
<th>I wanted to write that it would be better for students to wear uniforms, so I wanted to write opinions supporting that idea first . . . (Uh, huh. Have you decided what you are going to write at the end?) No. (Not yet?) No. (Have you decided what you are going to write in the middle?) I didn’t think about it at all (Novice 8).</th>
</tr>
</thead>
<tbody>
<tr>
<td>(3) <strong>Local planning</strong></td>
<td>Planning what to write next</td>
<td>After I wrote this down, I wondered what I should write next (Novice 1).</td>
</tr>
<tr>
<td>(4) <strong>Organizing</strong></td>
<td>Organizing the generated ideas</td>
<td>I wondered how I could synthesize the opposing arguments that came up in my own mind (Expert 1).</td>
</tr>
<tr>
<td>(5) <strong>Conclusion planning</strong></td>
<td>Planning of the conclusion</td>
<td>Yes, I gave up (on giving a concrete example), and so, I decided to conclude the composition (Novice 1).</td>
</tr>
</tbody>
</table>

**Retrieving**

| (1) **Plan retrieving** | Retrieving the already constructed plan | I thought about writing “Generally speaking, uniforms deprive students of freedom,” but I deleted the sentence because I remembered that I was supporting the position that students should wear uniforms (Novice 7). |
| (2) **Information retrieving** | Retrieving appropriate information from long-term memory | I was trying to think of one more reason (to support my opinion) (Novice 5). |

**Generating ideas**

| (1) **Naturally generated** | Generating an idea without any stimulus | There was something that just came up in my mind (Novice 5). |
| (2) **Description generated** | Generating an idea related to the previous description | After the last paragraph, I thought of the thing I could write next (Novice 8). |

**Verbalizing**

| (1) **Verbalizing a proposition** | Verbalizing the content the writer intends to write | I wanted to say something more general, but I was wondering how to write it (Novice 2). |
| (2) **Rhetorical refining** | Refining the rhetorical aspect(s) of an expression | I was wondering whether I should use “choosing” or “deciding” here (Expert 3). |
| (3) **Mechanical refining** | Refining the mechanical or (L1/L2) grammatical aspect(s) of an expression | Yes, I was wondering how I spell “chaotic” (Expert 4). |
| (4) **Sense of readers** | Adjusting expression(s) to the readers | I was thinking what would be the best way to present the reason to the readers (Novice 1). |

**Translating**

| | Translating the generated idea into L2 | I wrote down in Japanese that we should ask students’ opinions, but I could not translate it into English (Novice 3). |
APPENDIX D.

Examples of Episode Boundaries That Appeared in One Skilled Student Writer’s Composition

I’m infavor of Uniform System. 

In Japan, almost junior high school and high school introduce Uniform system and students obey the system. 

There are many merit “Uniform system.”

First, uniform is symbol of students and the school. Every school has own uniform, pattern, color and so on. It’s fun. Second, students’ object is studying.

Especially, when they are junior high or high school students, they should study many things and much time. Therefore, but for uniform, they would have passed choosing clothes, and they can’t study any more. Next, uniform remember me of student age. When I was a high school student, I had our own uniform. Because of the dark color, I didn’t like it. But whenever I watch the picture of the uniform, my happy life is coming back. Now I like it very much. Uniform has many many memories.

This is why, I’m infavor of uniform system.

Notes:

a The example preserves the original indentation, capitalization, spelling, and wording.
b Each boundary is marked by a slash.