

Vocabulary Change in Process Writing: Effect of Text Structure Instruction

Sachi Oshima
Chuo Gakuin University, Chiba, Japan
oshima.s@mc.cgu.ac.jp

ABSTRACT

The purpose of this study, conducted as a follow-up study of Oshima (2020), was to examine whether Japanese EFL students' use of vocabulary changed after being given lessons on explicit instruction on text structure and process writing. Two groups of college students—the beginner-level group writing a descriptive essay and the advanced-level group writing an argumentative essay—wrote an outline, the first draft (D1), the second draft (D2), and the final draft (FD), and I examined the differences in lexical richness between students' D1 and FD with New Word Level Checker (Mizumoto, 2021). The results showed that both groups' drafts had changed in the number of words used (tokens), the number of unique words used (types), and the number of lower frequency words used. This study's finding also supports the importance of choosing an appropriate measurement to analyze students' vocabulary levels. For Japanese students, 1K-word bands, which have been widely used in previous literature, seem too broad to capture their small vocabulary improvement.

KEYWORDS: Vocabulary change, Text structure instruction, Process writing, New Word Level Checker, SEWK-J

1. INTRODUCTION

Many English teachers in Japan have been faced with the need to help beginner-level college students improve their English writing skills. The Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT) (2015) reported that 31.90% of senior high school graduates had reached the A2 level of the CEFR, which means that 69.10% were A1-level basic users of English even after ten or more years of instruction before entering a university. Teachers frequently find students' writing lacks coherence

and basic organization. Their writing also contains many grammatical errors, and vocabulary is often inappropriately used. Some previous researchers (e.g., Nishigaki et al., 2007; Sasaki, 2000) have examined Japanese students at the beginner's or low-intermediate levels; however, more studies are needed to respond to a considerable number of Japanese EFL college students with low English proficiency.

To address this need, I conducted a study (Oshima, 2020) by using an action research design (Wallace, 1998), in which I carried out the research while teaching students simultaneously. The participants were two groups of Japanese EFL college students: a group of advanced-level students who wrote an argumentative essay and the other group with beginner-level students who wrote a descriptive essay. I considered that the Japanese EFL college students' poor writing performance would be partially due to the differences in the text structures between Japanese and English. Therefore, while introducing process writing in lessons, I provided the students with explicit instruction on how an English text should be organized. Based on the positive changes found in both groups' drafts in terms of the essay's organization and their awareness of the text's readers, I concluded that explicit instruction on text structure can be a useful means of developing English writing skills regardless of students' English proficiency levels.

The above study revealed some positive changes brought by explicit instruction on text structure during process writing; however, one area remained uninvestigated: vocabulary. In the study, the students in both groups answered an open-ended questionnaire after submitting their final drafts. Both groups reported that their ability to organize ideas improved the most, which had been expected. However, an unexpected result was that the students cited an improvement in word choice or vocabulary, even though the treatment involved an explicit instruction on text structure, not on vocabulary. Did the students' self-perceptions of vocabulary development really occur through revising drafts? If so, to what extent did their vocabulary change? The purpose of this follow-up study is to examine whether students' vocabulary use changed after being given explicit instruction on text structure and process writing.

2. LITERATURE REVIEW

2.1 Incidental Vocabulary Learning and Writing

In the view of applied linguistics, incidental vocabulary learning can be defined as the by-product of a meaning-focused task (Webb, 2019). Comparing it with intentional learning by which learners are engaged in word-focus activities, Webb (2019) explained that incidental vocabulary learning takes place through meaning-focused tasks in which learners obtain input for the purpose of interest, information, and enjoyment.

Many researchers have examined incidental vocabulary learning through reading, listening, and viewing (e.g., Brown, 2021; Dang et al., 2022; Feng & Webb, 2020). For example, Brown (2021) employed a classroom-as-lexical-environment approach and examined incidental vocabulary learning in a regular classroom environment. In his study, 29 Japanese university students at high-intermediate to intermediate L2 English proficiency levels took 90-minute English classes twice over a 15-week course. He recorded the students' orthographic exposures to 40 vocabulary items in these regular

classes and conducted pre-course and post-course vocabulary tests (word-recognition and meaning-recall tests). The students' scores on both tests significantly increased with a large effect size, and Brown (2021) suggested that the frequency of occurrence in the classroom and word form variation were considered to have significant positive effects on vocabulary gains. Dang et al. (2022) examined incidental vocabulary learning by watching the video of an academic lecture. They hypothesized that academic lectures are likely to be an important source of incidental vocabulary learning for L2 learners because lecturers repeat topic-related words and learners pay attention to the input. Fifty-five postgraduate students at a university in China watched a 50-minute video (without captions) of a lecture in an introductory undergraduate course in algorithms at the Massachusetts Institute of Technology. In the video, the lecturer did not use PowerPoint slides but wrote key information and formulas on the blackboard; in other words, the video did not contain much orthographic input, and the students learned target vocabulary (50 single words and 19 collocations) mainly from auditory input or listening. As a result of the pretest, posttest, and delayed posttests—meaning recall test for single-word items and form recognition test for collocation—of target vocabulary, Dang et al. (2022) suggested that vocabulary learning is likely to occur through viewing an academic lecture. They also pointed out that frequency of occurrence is likely to affect the learning of single words, but not collocations.

Contrary to studies on incidental vocabulary learning through reading, listening, and viewing, such as above, fewer studies have been focused on incidental vocabulary learning through writing. For instance, in Uchiyama et al.'s (2019) meta-analysis examining 26 studies on incidental vocabulary learning, none of these studies employed writing as a mode of input; more specifically, the modes of input in these 26 studies were either reading, listening, reading while listening, or viewing.

Incidental vocabulary learning through writing has been examined from the perspective of the Involvement Load Hypothesis (ILH; Laufer & Hulstijn, 2001). ILH proposes that retention of L2 words is conditional upon each task's involvement load (IL) determined by one motivational component (*need*) and two cognitive components (*search & evaluation*). Here, *need* refers to whether unknown words are needed to complete a task, *search* means the attempt to find the L2 form of a word or its meaning, and *evaluation* refers to comparing an unknown word's L2 form or meaning with other possible words or meaning (Yanagisawa & Webb, 2021). In Yanagisawa and Webb's (2021) meta-analysis, which examined 42 empirical studies on ILH and incidental L2 vocabulary learning, writing tasks such as sentence writing and composition writing were employed in 36 out of the 42 studies. Yanagisawa and Webb showed that ILH was significantly predictive of learning. Pichette et al. (2012) also compared the relative effectiveness of reading and writing sentences for incidental L2 vocabulary learning of 16 target items. French-speaking ESL students at intermediate and advanced levels ($N = 203$) read sentences containing target items in the reading task, whereas they wrote three sentences per item in the writing task. The researchers employed "cued recall test" (p. 71) in which the students were provided with a clue (L1 French definition) and required to produce the L2 English word corresponding to the clue. The results obtained by a surprise immediate recall test showed a superior recall for the writing task over the

reading task, although the delayed recall test suggested that this superiority of writing disappears over time. Based on the ILH, Pichette et al. (2012) concluded that “writing a text may lead to significantly higher recall than reading” (p. 77) because writing requires greater cognitive demands or IL than reading.

2.2 Vocabulary Development in Process Writing

In process writing, students generate and organize their ideas, write multiple drafts, get teacher and peer feedback, revise drafts in response to the feedback, and complete the final draft (White & Arndt, 1991). Muncie (2002) examined the vocabulary development of 30 intermediate and upper-intermediate Japanese university students after process writing. Students first wrote a timed composition of self-introduction and then three drafts about “friendship” through process writing. A Lexical Frequency Profile (LFP; Laufer & Nation, 1995) was employed to analyze the students’ timed composition of self-introduction, and the first and final drafts of the “friendship” composition. The LFP showed the percentage of words used in the texts in three categories: the first 1000 most frequent words of English (below 1000), the next 1000 most frequent words (1000 to 2000), and the words which cannot be included in the two categories (above 2000). Muncie (2002) compared the LFP ratio among three groups (timed composition, the first draft, and the final draft), and an ANOVA showed that there was no significant difference among them; in other words, the students’ vocabulary did not change from the first draft to the final draft. Responding to this result, he explained that “this result is not surprising, since one could argue that an improvement in vocabulary could not reasonably be expected over the course of writing one composition” (p. 230).

Muncie’s (2002) results showed no significant improvement in vocabulary from the first to the final drafts; however, several points should be addressed before concluding. First, it seems necessary to consider whether LFP was appropriate for measuring the vocabulary improvement in process writing. As Muncie (2002) admitted, the difficulty in interpreting the results of LFP, the comparison of the three categories’ percentages (below 1000, 1000-2000, and above 2000) might not be a suitable measurement for vocabulary improvement because “the ratio of above 2000 words would increase not only when a new word in the above 2000 range is introduced, but also when a word in the below 2000 or 1-2000 range is removed” (p. 232). Second, he excluded five students’ texts from the analysis because their initial timed compositions were short and “the LFP is only stable on compositions of over 200 words” (p. 228). This omission of the five students’ texts not only reduced the statistical power but also hindered the chance to investigate how students who had written less than 200 words improved their drafts due to process writing. Moreover, students’ performances on the timed writing of self-introduction and those on the process writing of “friendship” would not always be the same because of the differences in several variables, such as the topics, time pressure, and perhaps, genres. Finally, in order to employ the LFP, Muncie (2002) could not analyze each student’s text as it was, which might have affected the results. He needed to ensure that only the correctly used words were entered in the LFP, proper nouns were deleted, and correct forms of misspelled words were added to the analysis. To put it simply, while he tried to examine the improvement from the students’

first draft to the final draft, he modified or rather improved the students’ texts, especially vocabulary, beforehand by correcting spelling or ensuring the correct use of words. If he had analyzed the students’ drafts without any modification, the results might have been different.

2.3 Research Questions

The purpose of this study is to examine whether Japanese EFL students’ use of vocabulary changes after being given lessons on explicit instruction on text structure and process writing. Considering several points to be addressed in Muncie’s (2002) study, I employed a novel online vocabulary profiling application, New Word Level Checker (NWLC; Mizumoto, 2021) and examined all students’ drafts without any prior modification or omission. The performances of two groups of students at different English proficiency levels are also compared. The following two research questions are examined:

- RQ1. Does students’ vocabulary change after being given lessons on explicit instruction on text structure and process writing? If so, how does it change?
- RQ2. Is the change in vocabulary while process writing, if any, applicable to students at different levels?

3. METHOD

3.1 Participants and Types of Essays Used

The students’ written texts were obtained from Oshima’s (2020) action research targeting two groups of Japanese EFL college students: one group of advanced-level students who wrote an argumentative essay and the other group of beginner-level students who wrote a descriptive essay. Table 1 is a summary of the participants and types of essays used.

Table 1
Participants and essay

	Advanced-level Group	Beginner-level Group
Number of Students	33	13
Grade	2nd year	2nd year
Majors	Non-English Majors	Non-English Majors
Number of Lessons	4 Lessons (90 min. each)	4 Lessons (90 min. each)
English Proficiency	High-intermediate or Advanced (TOEIC scores between 600-800)	Beginners (TOEIC Scores below 300)
Essay Genre	Argumentative	Descriptive
Essay Theme	Media	Three subjects studied in the university

Note. Adapted from Oshima (2020).

3.2 Treatment: Explicit Instruction on Text Structure in Process Writing

The lessons were conducted over four weeks. The students in both groups went through the following four phases: (1) Week 1—writing an outline (90 minutes), (2) Week 2—writing draft 1 (D1, 90 minutes), (3) Week 3—writing draft 2 (90 minutes), and (4) Week 4—writing the final draft (FD, 90 minutes). In each phase, the teacher explicitly introduced the structure of an essay—argumentative for advanced-level students and descriptive for beginners.

After each lesson, the students completed their writing at home and submitted the assignment by email before the next lesson. The teacher wrote comments and feedback on each draft and brought them to the next lesson. The feedback on outline, D1, and draft 2 was focused on coherence and whether each student logically organized their idea. In the phase of writing FD, the teacher made additional comments in terms of whether the text would be clearly understood by the readers. If students had any questions or concerns during these four phases, they could always consult the teacher in class or via email.

3.3 Data Analysis Method

Research questions asked whether explicit instruction on text structure and process writing changed vocabulary in students' written texts. To examine changes over the four-week course, first, the NWLC (Mizumoto, 2021) was used to analyze the vocabulary used in the students' D1 and FD. Each student's D1 and FD were entered into NWLC, first with the Scale of English Word Knowledge—Japanese (*SEWK-J*) word list and secondly with the *SEWK-J: Fine-grained* word list. Then, the results obtained by NWLC were analyzed by the paired-samples *t*-test to identify whether there were any changes in vocabulary use from D1 to FD.

SEWK-J and *SEWK-J: Fine-grained* were chosen for the following four reasons:

First, as “the *SEWK-J* was developed to estimate the difficulty that the vocabulary in a text presents to Japanese learners of English” (Mizumoto et al., 2021, p. 31), NWLC, together with *SEWK-J* or *SEWK-J: Fine-grained*, is optimized for Japanese EFL learning contexts and perfectly fits this study's context.

Secondly, *SEWK-J* and *SEWK-J: Fine-grained* are designed to represent learners' actual vocabulary knowledge. Mizumoto et al. (2021) suggested that “when matching learners with texts, we should consider basing test and lexical profilers on what learners *do* know, rather than...what learners *should* know” (p. 39). In this study, the purpose of analyzing the students' texts is to gauge what vocabulary the students actually used; therefore, *SEWK-J* and *SEWK-J: Fine-grained* fit this study's aim.

Thirdly, *SEWK-J* and *SEWK-J: Fine-grained* are flemma-based word lists, which was considered appropriate for Japanese EFL contexts (McLean, 2018; Mizumoto et al., 2021). Flemma consists of a word's base form and its associated inflectional forms regardless of part of speech. On the other hand, the word family, which many previous researchers have employed, consists of base word and its derived forms from Level 2 to 6 on Bauer and Nation's (1993) word family hierarchy (Kyle, 2019; McLean, 2018). The difference between the flemma and the word family is obvious when looking at one word, for example, *develop*. The flemma of a word *develop* consists of only four forms

(*develop*, *develops*, *developed*, and *developing*), whereas the word family of the same word *develop* consists of 19 forms including *development*, *undevelopmentable*, and *developmentwise* (see McLean, 2018, p. 826, for more details). McLean (2018) pointed out that Japanese EFL learners, especially those at the beginner's level, have a very limited ability to comprehend associated derivational forms. He, therefore, suggested that the *flemla* was found to be a more appropriate word counting unit for Japanese EFL learners.

Finally, by using *SEWK-J* and *SEWK-J: Fine-grained*, the students' vocabulary can be analyzed by finer bands than 1K-word bands used by other profilers, such as VocabProfiler (<https://www.lextuor.ca/vp/comp>) and LFP (Laufer & Nation, 1995). By VocabProfiler, researchers can analyze written texts by 1K-word bands according to frequency. Muncie (2002) used LFP and analyzed the words in three categories: the first 1000 most frequent words, the next 1000 most frequent words, and the words which cannot be included in the two categories. On the other hand, *SEWK-J* provides the results in 500-word bands, and *SEWK-J: Fine-grained* shows the results in 250-word bands. These finer bands are expected to provide a more detailed assessment of vocabulary used, as even trivial, vocabulary improvement can be captured.

4. RESULTS

4.1 The Number of Tokens and Types

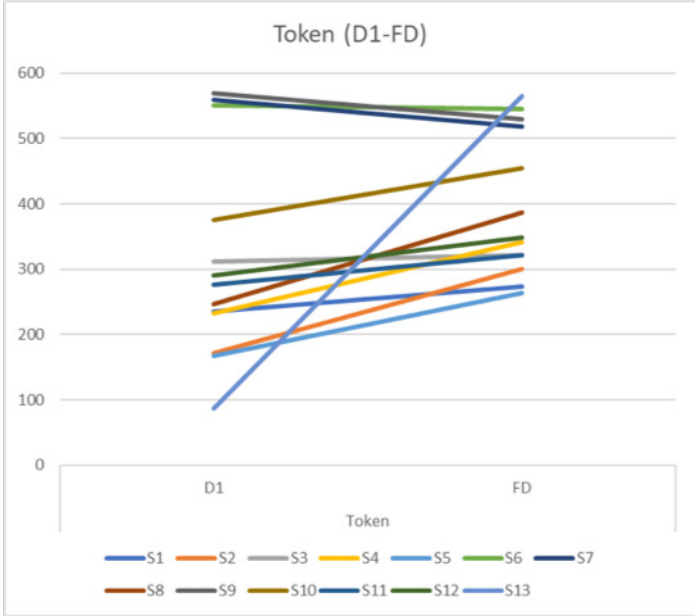
When analyzing vocabulary, the definitions of lexical items are important because they can significantly affect the results of a study (Kyle, 2019; Mizumoto et al., 2021). In this study, lexical items are defined and operationalized based on Kyle (2019). First, *tokens* refer to running words in a text, whereas *types* refer to unique words in a text. For example, the following sentence, "I am going to see cherry blossoms with my friends who live near my house," consists of 15 tokens (words), but it includes 14 types (unique words) because there are two "my."

Beginner-Level Group

First, comparing the number of tokens and types counted in students' D1 and FD, both tokens and types showed statistically significant difference with a middle to large effect size (Loerts et al., 2020); for tokens, $t(12) = -2.30, p = .041, d = -0.64$; for types, $t(12) = -2.54, p = .026, d = -0.71$. The students' essays became longer through revising drafts, and accordingly, the number of tokens and types increased.

Although the students' essays generally increased in the number of tokens and types, there were also individual differences (Figure 1).

Figure 1
Beginner-level learners' change in written tokens

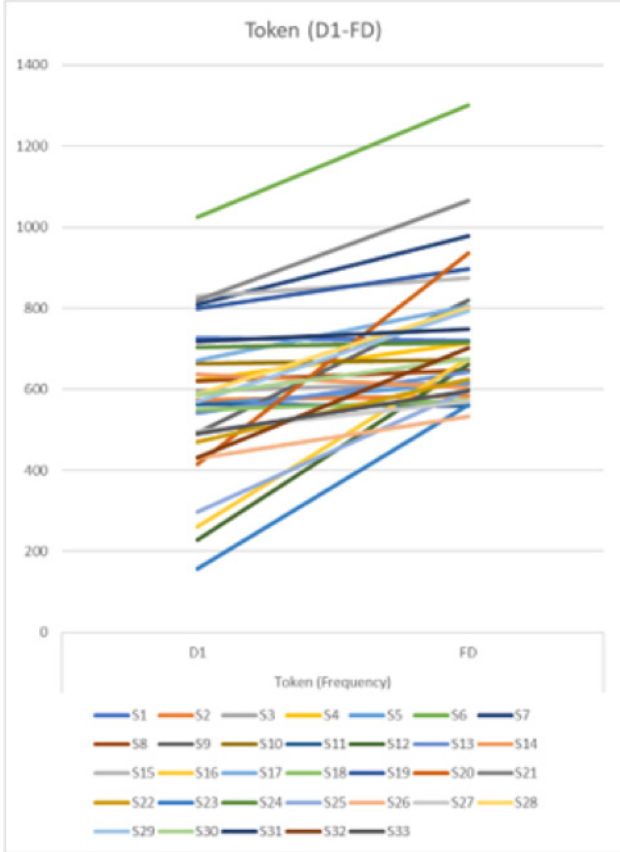


Above all, what seemed unique in Figure 1 was that the students who had written more than 500 tokens in their D1 (S6, S7, and S9) slightly decreased the number of tokens and types in their FD, whereas those who had written less than 500 tokens in their D1 increased the number of tokens and types in their FD. Figure 1 suggested one student’s (S13) sharp increase in the number of tokens because S13 had forgotten to write their D1 by the due date and submitted the draft in a great hurry. As a result, S13’s D1 looked like a memo containing only a few sentences and phrases. S13 spent time writing their FD so that the number of tokens substantially improved.

Advanced-level Group

Similar to the beginner-level group, statistically significant differences with large effect sizes (Loerts et al., 2020) were observed in the advanced-level group; for tokens, $t(32) = -6.08, p < .001, d = -1.06$; and, types, $t(32) = -4.61, p < .001, d = -0.80$. Individual differences were found; however, as a whole, students’ data showed a similar tendency—their essays became longer, and accordingly, the number of tokens and that of types increased (Figure 2).

Figure 2
Advanced-level learners’ change in written tokens



4.2 The Word Level Analyzed with *SEWK-J*

Each student’s D1 and FD were entered into NWLC with the *SEWK-J* word list, which provides the results with 500-word bands (see Tables 2 and 3) according to frequency. The percentage of each level was analyzed by the paired-samples *t*-test to see the change from D1 to FD.

Beginner-Level Group

Descriptive statistics (Table 2) showed that mean values at each level did not change (less than 1.0). Paired-samples *t*-test also indicated that there was no significant difference between each level (*p* > .05). Although the beginner-level students’ essays became longer, they did not use more lower frequency vocabulary (i.e., more difficult or complex vocabulary) from D1 to FD.

Table 2*Descriptive statistics (SEWK-J, beginner-level)*

	Proper Nouns		L1 (1-500)		L2 (501-1000)		L3 and above (1001-)	
	D1	FD	D1	FD	D1	FD	D1	FD
Valid	13	13	13	13	13	13	13	13
Missing	0	0	0	0	0	0	0	0
<i>M</i>	10.06	10.31	68.69	69.10	10.54	10.31	10.66	10.17
<i>SD</i>	3.68	3.71	4.98	4.574	2.25	1.31	2.11	2.23
<i>Min</i>	3.09	4.04	62.79	62.77	5.75	8.00	7.73	6.90
<i>Max</i>	16.09	16.37	77.49	79.12	12.79	11.78	15.48	14.78

Advanced-Level Group

Similar to the beginner-level group, there was almost no difference (less than 1.0) in the mean values at each level (Table 3). Paired-samples *t*-tests also indicated no significant difference in each level ($p > .05$). The advanced-level students' texts did not improve from D1 to FD in terms of vocabulary level, either.

Table 3*Descriptive statistics (SEWK-J, advanced)*

	Proper Nouns		L1 (1-500)		L2 (501-1000)		L3 and above (1001-)	
	D1	FD	D1	FD	D1	FD	D1	FD
Valid	33	33	33	33	33	33	33	33
Missing	0	0	0	0	0	0	0	0
<i>M</i>	5.03	5.14	68.71	68.08	12.41	12.68	13.74	13.98
<i>SD</i>	2.91	2.65	5.57	4.91	2.67	2.17	3.76	3.61
<i>Min</i>	0.46	0.69	58.8	60.48	4.82	8.32	8.56	7.86
<i>Max</i>	11.90	11.25	80.87	79.20	17.83	17.25	21.75	21.44

4.3 The Word Level Analyzed with SEWK-J: Fine-grained

Analyses using SEWK-J with 500-word frequency bands (Tables 2 and 3) did not provide any significant evidence of the students' vocabulary development. Thus, SEWK-J: Fine-grained which reported on 250-word frequency bands (see Tables 4 and 5) was employed to capture more incremental changes in vocabulary use.

Beginner-Level Group

Descriptive statistics (Table 4) showed that the mean value at each 250-word band was not very different between D1 and ED; however, the paired-samples *t*-tests showed that the Band 3 (501-750) was statistically significant, $t(12) = 2.97, p = .012$,

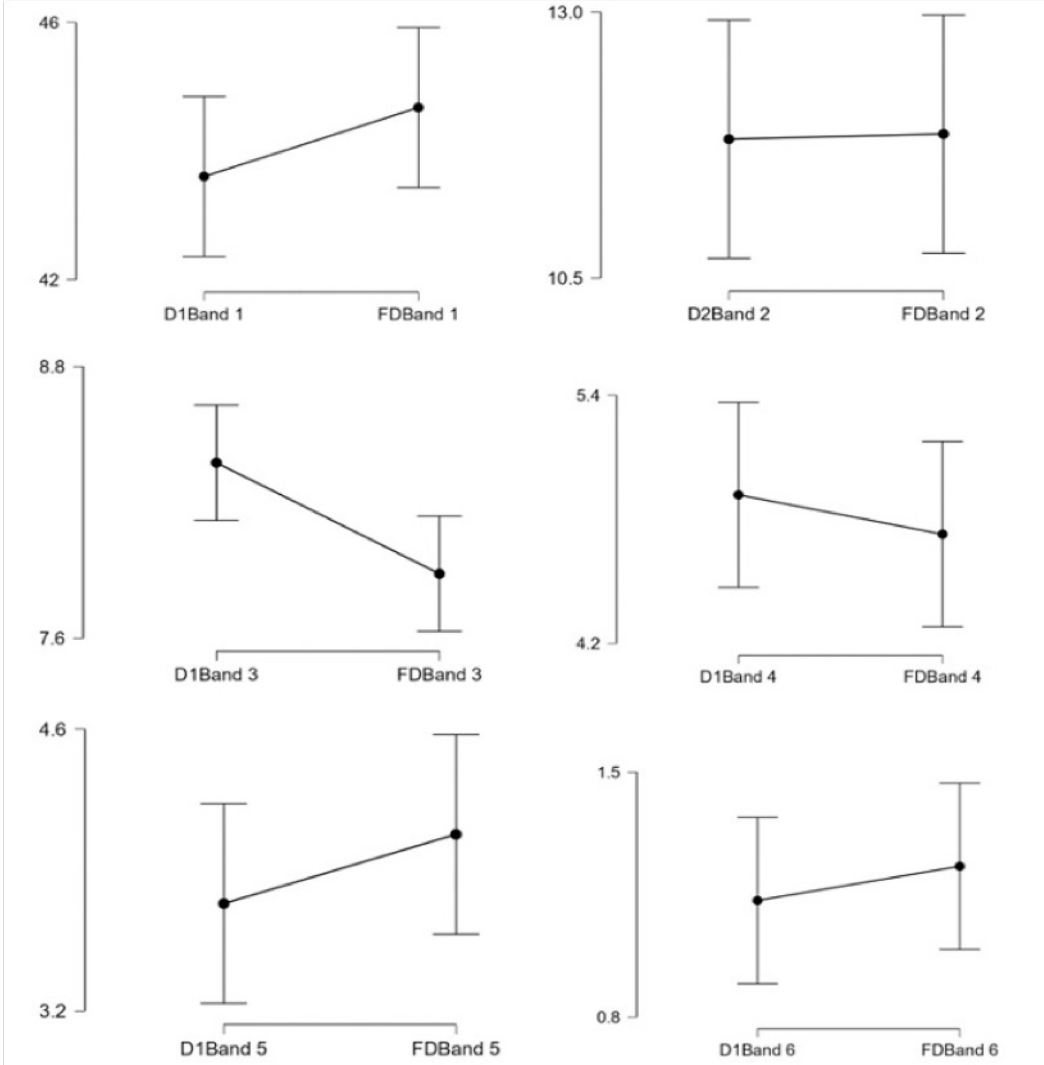
$d = 0.82$, indicating that Band 3 words declined from D1 to FD with a large effect size (Loerts et al., 2020).

Table 4
Descriptive statistics for SEWK-J: Fine-grained (beginner-level)

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>
D1 Band 1 (1-250)	13	43.60	3.93	1.09
FD Band 1	13	44.68	3.89	1.08
D2 Band 2 (251-500)	13	11.81	2.94	0.81
FD Band 2	13	11.86	2.20	0.61
D1 Band 3 (501-750)	13	8.38	1.83	0.51
FD Band 3	13	7.89	2.19	0.61
D1 Band 4 (751-1000)	13	4.92	2.10	0.58
FD Band 4	13	4.73	1.578	0.44
D1 Band 5 (1001-1250)	13	3.73	1.13	0.31
FD Band 5	13	4.08	1.08	0.30
D1 Band 6 (1251-1500)	13	1.13	0.69	0.19
FD Band 6	13	1.23	0.82	0.233

Visual inspections (Figure 3) also suggested changes from D1 to FD; the number of tokens in Bands 1, 5, and 6 increased, whereas those in Bands 3 and 4 decreased. Above all, looking at Band 3 (510-750), it is clear that the students’ use of Band 3 tokens dropped from D1 to FD.

Figure 3
Visual Inspection (SEWK-J: Fine-grained, beginner)



Advanced-Level Group

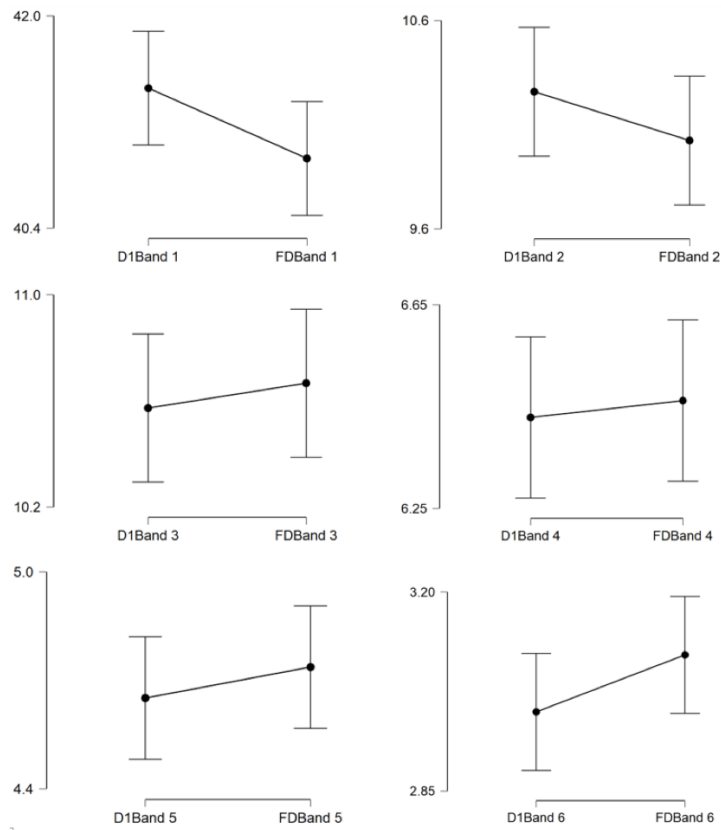
Descriptive statistics (Table 5) showed that the mean value at each 250-word band was not very different between D1 and FD. Because the Shapiro-Wilk test showed that normality was violated in several word bands, a non-parametric Wilcoxon signed-rank test was conducted instead of a paired-samples *t*-test. As a result, Band 1 (1-250) was statistically significant, $w = 394.00$, $p = .043$, indicating that Band 1 words significantly declined from D1 to FD.

Table 5
Descriptive statistics (SEWK-J: fine-grained, advanced)

	<i>N</i>	<i>M</i>	<i>SD</i>	<i>SE</i>
D1 Band 1 (1-250)	33	41.46	3.75	0.65
FD Band 1	33	40.93	3.12	0.54
D1 Band 2 (251-500)	33	10.26	2.40	0.42
FD Band 2	33	10.02	1.97	0.34
D1 Band 3 (501-750)	33	10.57	2.73	0.48
FD Band 3	33	10.67	2.37	0.41
D1 Band 4 (751-1000)	33	6.43	2.34	0.41
FD Band 4	33	6.46	2.33	0.41
D1 Band 5 (1001-1250)	33	4.65	1.62	0.28
FD Band 5	33	4.74	1.28	0.22
D1 Band 6 (1251-1500)	33	2.99	1.27	0.22
FD Band 6	33	3.09	1.29	0.23

Visual inspection (Figure 4) suggested a wide variance in each individual’s data; however, the words in Band 1 (1-250) significantly decreased from D1 to FD, and those in Bands 5 and 6 increased.

Figure 4
Visual inspection (SEWK-J: fine-grained, advanced)



5. DISCUSSION

The purpose of this study was to examine whether Japanese EFL college students’ drafts changed in terms of vocabulary use. More specifically, I examined whether the students’ use of vocabulary changed from D1 to FD.

Although the ILH (Laufer & Hulstijn, 2001) and some previous researchers such as Pichette et al. (2012) and Yanagisawa and Webb (2021) have supported incidental L2 vocabulary learning through writing, Muncie (2002) showed that his participants’ vocabulary did not change from D1 to FD because “an improvement in vocabulary could not reasonably be expected over the course of writing one composition” (p. 230). As mentioned in the introduction of this paper, the motive of this follow-up study was an unexpected result of the students’ post-treatment questionnaire in Oshima (2020); that is, both beginner- and advanced-level groups reported an improvement in word choice or vocabulary, even though they were given an explicit instruction on text structure. As a result of this follow-up study, both beginner- and advanced-level students’ vocabulary changed from D1 to FD in terms of: (1) the number of tokens (words) written, (2) the number of types (unique words) used, and (3) the number of low frequency vocabulary used. It appears that these changes have the students perceive that they had improved word choice or vocabulary in their essays.

5.1 Tokens and Types

First, the number of tokens significantly increased from D1 to FD in both beginner-level and advanced-level groups. Accordingly, the number of types also significantly increased in both groups. These two measures (tokens & types) suggest that, through revising their drafts, the students produced longer FD than D1.

One main reason why they produced longer FD than D1 is the enhancement of the students’ awareness of readers (Hyland, 2003; White & Arndt, 1991). When the students wrote something unclear to the readers, the teacher wrote comments such as “What is XX?” and “Add more explanation.” This feedback seemed to encourage the students to elaborate their explanations, add more information, and describe the content in more detail, which resulted in more sentences produced in FD (see Table 6 for an example).

Table 6
Excerpt from an advanced-level student’s texts

D1	...Thus, people have growing doubts about the way Japanese media reports about violent crimes. For example, in Zama case , the victims were...
FD	...Thus, people have growing doubts about the way Japanese media report violent crimes. For example, in [the] Zama case in which nine people were killed in Zama City in 2018 , the victims were...

Note. Adapted from Oshima (2020). Emphasis added.

5.2 The Vocabulary Level

Secondly, the vocabulary level measured by NWLC with *SEWK-J: Fine-grained* showed changes from D1 to FD in both beginner- and advanced-level groups.

As for the beginner-level group, Band 3 (word 501-750) word use declined from D1 to FD with a large effect size. After a closer analysis of each student’s drafts, the change in Band 3 seemed to be caused by a specific word: will (in Band 3). Many students mistakenly used the future tense of will in their D1; however, through revising drafts, they became able to use the correct tenses, such as present or past tense (see Table 7).

Table 7
Excerpt from a beginner-level student’s texts

D1	First subject is Music...Four teachers in charge will take turns on a weekly basis. The grade evaluation is decided by attendance and test. In lesson, the students will sing the ninth. The students will sing the ninth....
FD	First subject is Music...Four teachers in charge shift turns on a weekly basis. The evaluation is decided by attendance and test. In lesson, the students sing the Beethoven’s Ninth. Students sing the ninth....

Note. Emphasis added.

As for the advanced group, Band 1 (1-250) declined from D1 to FD with a statistically significant difference. The close analysis of each student’s draft suggested that one possible cause of this decline was the word *it* (in Band 1). As shown in Tables 8 and 9, the advanced-level students often used the pronoun *it*, which sometimes makes the sentence unclear to the readers. Through revising drafts, especially through enhancing the readers’ awareness, the students reduced the use of *it* and elaborated on their explanations with lower-frequency words.

Table 8
Excerpt from an advanced-level student’s texts

D1	To sum up there it is undeniable that SNS creates isolation by several kinds of dreadful outcome. It is distinct that <u>it must make movements to protect people from those bad effects</u> . Although it is impossible to human beings live without SNS, it is unrealistic to ban using SNS.
FD	To sum up, it is undeniable that SNS creates isolation by several kinds of dreadful outcome. It is distinct that <u>making movements to protect people from those bad effects is indisputable</u> . Live without SNS is impossible. Therefore, it will be unrealistic to ban using SNS.

Note. Emphasis added.

Table 9*Excerpt from an advanced-level student's texts*

D1	Many have argued that it costs to buy newspapers. While it is true that people can watch news cheaper or for free on the internet and TV, information in newspaper is reliable, and it is worth buying.
FD	Many have argued that it costs to buy newspapers. While it is true that people can watch news cheaper or for free on the internet and TV, information in newspapers is <u>more reliable than that in new technologies, and newspapers are worth buying.</u>

Note. Emphasis added.

6. CONCLUSION

This study was conducted as a follow-up study of Oshima (2020), and its aim was to examine whether EFL learners' use of vocabulary changed after explicit instruction on text structure and process writing. I employed the New Word Level Checker (NWLC; Mizumoto, 2021) and examined the students' drafts without any prior modification or omission. The results suggested that, even without explicit instruction on vocabulary, the students' drafts had changed from D1 to FD in terms of the number of tokens (words), the number of types (unique words), and the number of lower frequency words used. It is true that the teacher's feedback was provided through process writing, which might have some influences on the students' vocabulary improvement. However, the purpose of this feedback was to improve logical organization and raise the awareness of readers, not to have the students focus on vocabulary. As Muncie (2002) suggested, teachers should "introduce activities which would stimulate the use of more advanced vocabulary" (p. 234) in process writing, it would be worth examining whether and to what extent explicit vocabulary instruction can improve students' vocabulary use in process writing.

It cannot be denied that there are several limitations in this study. First, the number of participants was small, which reduced the statistical power and hinders the generalizability of the results. Above all, in the beginner-level group, one student (S13) had forgotten to write their D1 by the due date and submitted it in a great hurry, making their D1 considerably shorter. Among the 13 beginner-level students, this outlier student's data substantially affected the statistical analysis. Secondly, obtaining the number of tokens and types might yield the motivation to calculate the type-token ratio (TTR), which can be calculated by the number of types divided by that of tokens (type/token); however, simply comparing the TTR of D1 and FD was considered to provide misleading results because the lengths of D1 and FD were showing a significant difference (Figures 1 & 2) (Covington & McFall, 2010; Kyle, 2019). Alternatives of TTR, such as VocD and Measure of Textual Lexical Diversity (MTLD), could be employed in future studies. Thirdly, the interpretation of the effect size (Cohen's *d*) in this paper has room to be further examined. I employed Loerts et al.'s (2020) benchmark, suggesting *d* value of 0.20 as small, 0.50 as medium, and 0.80 as large effect size (p. 78). On the other hand, Plonsky and Oswald (2014) proposed an alternative

one for the pre-post or within-group effect size: “*d* value of .60 as generally small, 1.00 as medium, and 1.40 as large” (p. 889). As Plonsky and Oswald called for further meta-analyses to examine within- and between-group contrasts separately, alternative benchmark should be also considered in future studies.

As final remarks, it should be emphasized that this study’s finding supports the importance of choosing an appropriate measurement to analyze students’ texts. Muncie (2002) employed LFP (Laufer & Nation, 1995) and could not find improvement from D1 to FD. In this study, I first employed the *SEWK-J* word list, which provides the results within 500-word bands, and could not find the students’ improvement. However, by employing *SEWK-J: Fine-grained*, which shows the results with 250-word bands, I could observe significant changes between learners’ D1 and FD. Considering the length of students’ essays—in most cases, less than 1,000 words for advanced-level students and less than 500 words for beginners, 1K-word bands seem too broad to capture the students’ vocabulary improvement. From this, analyzing students’ texts with finer bands should be recommended so that their small, even trivial, vocabulary improvement can be captured.

Oshima (2020) suggested that explicit instruction on text structure can be a useful means of developing students’ writing in terms of logical organization and the awareness of readers. Based on the result of this study, it can be concluded that explicit instruction on text structure also affects learners’ vocabulary use.

Acknowledgments

I would like to thank Stuart McLean for his helpful suggestions and guidance to conduct this study. I also thank the editorial team of *Englishes in Practice (EiP)* and the anonymous reviewers for their insightful comments and feedback on an earlier version of this paper.

References

- Bauer, L., & Nation, I. S. P. (1993). Word families. *International Journal of Lexicography*, 6(4), 253–279.
- Brown, D. (2021). Incidental vocabulary learning in a Japanese university L2-English language classroom over a semester. *TESOL Journal*, 12(4), e595. <http://doi.org/10.1002/tesj.595>
- Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type–token ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94–100. <http://doi.org/10.1080/09296171003643098>
- Dang, T. N. Y., Lu, C., & Webb, S. (2022). Incidental learning of single words and collocations through viewing an academic lecture. *Studies in Second Language Acquisition*, 44(3), 708–736. <https://doi.org/10.1017/S0272263121000474>

- Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, 42(3), 499–523. <https://doi.org/10.1017/S0272263119000494>
- Hyland, K. (2003). *Second language writing*. Cambridge University Press.
- Kyle, K. (2019). Measuring lexical richness. In S. Webb (Ed.), *Routledge handbook of vocabulary studies* (pp. 454–476). <https://doi.org/10.4324/9780429291586>
- Laufer, B., & Hulstijn, J. (2001). Incidental vocabulary acquisition in a second language: the construct of task-induced involvement. *Applied Linguistics*, 22(1), 1–26. <https://doi.org/10.1093/applin/22.1.1>
- Laufer, B., & Nation, P. (1995). Vocabulary size and use: lexical richness in L2 written production. *Applied Linguistics*, 16(3), 307–322. <https://doi.org/10.1093/applin/16.3.307>
- Loerts, H., Loweie, W., & Seton, B. (2020). *Essential statistics for applied linguistics* (2nd ed.). Red Globe Press.
- McLean, S. (2018). Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, 39(6), 823–845. <https://doi.org/10.1093/applin/amw050>
- Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT). (2015). *Seito no eigo-ryoku koko suishin plan [Plan for improving students' English ability]*. https://www.mext.go.jp/a_menu/kokusai/gaikokugo/__icsFiles/afieldfile/2015/07/21/1358906_01_1.pdf
- Mizumoto, A. (2021). *New Word Level Checker* [Web application]. <https://nwlc.pythonanywhere.com/>
- Mizumoto, A., Pinchbeck, G. G., & McLean, S. (2021). Comparisons of word lists on new word level checker. *Vocabulary Learning and Instruction*, 10(2), 30–41. <https://doi.org/10.7820/vli.v10.2.mizumoto>
- Muncie, J. (2002). Process writing and vocabulary development: Comparing lexical frequency profiles across drafts. *System*, 30(2), 225–235. [https://doi.org/10.1016/S0346-251X\(02\)00006-4](https://doi.org/10.1016/S0346-251X(02)00006-4)
- Nishigaki, C., Chujo, K., MacGoldrick, S., & Hasegawa, S. (2007). A cross-sectional contrastive analysis of Japanese students' English composition skills. *The Journal of Asia TEFL*, 4(1), 27–54. http://journal.asiatefl.org/main/main.php?inx_journals=11&inx_

contents=306&main=1&sub=3&submode=3&PageMode=JournalView&s_title=A_Cross_Sectional_Contrastive_Analysis_of_Japanese_Students_English_Composition_Skills

- Oshima, S. (2020). Effects of explicit instruction on text structures in process writing. In P. Clements, A. Krause, & R. Gentry (Eds.), *Teacher efficacy, learner agency* (pp. 433-442). JALT. <https://doi.org/10.37546/JALTPCP2019-50>
- Pichette, F., de Serres, L., & Lafontaine, M. (2012). Sentence reading and writing for second language vocabulary acquisition. *Applied Linguistics*, 33(1), 66–82. <https://doi.org/10.1093/applin/amr037>
- Plonsky, L., & Oswald, F. L. (2014). How big is “big”? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878-912. <https://doi.org/10.1111/lang.12079>
- Sasaki, M. (2000). Toward an empirical model of EFL writing processes: An exploratory study. *Journal of Second Language Writing*, 9(3), 259-291. [https://doi.org/10.1016/S1060-3743\(00\)00028-X](https://doi.org/10.1016/S1060-3743(00)00028-X)
- Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, 69(3), 559–599. <https://doi.org/10.1111/lang.12343>
- Wallace, M. J. (1998). *Action research for language teachers*. Cambridge University Press.
- Webb, S. (2019). Incidental vocabulary learning. In S. Webb (Ed.), *Routledge handbook of vocabulary studies* (pp. 225–239). <https://doi.org/10.4324/9780429291586>
- White, R., & Arndt, V. (1991). *Process writing*. Longman.
- Yanagisawa, A., & Webb, S. (2021). To what extent does the involvement load hypothesis predict incidental L2 vocabulary learning? A meta-analysis. *Language Learning*, 71(2), 487–536. <https://doi.org/10.1111/lang.12444>