Title	Japanese university students' integrated writing skills in listening-to-write tasks
Sub Title	
Author	大野, 真澄(Ōno, Masumi)
Publisher	慶應義塾大学法学研究会
Publication year	2021
Jtitle	教養論叢 (Kyoyo-ronso). No.142 (2021. 2) ,p.89- 110
JaLC DOI	
Abstract	
Notes	小屋逸樹先生・ 小瀬村誠治先生・ ロバート・ ギブソン先生退職記念特集号 論説
Genre	Departmental Bulletin Paper
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AN00062752-00000142-0089

Japanese University Students' Integrated Writing Skills in Listening-to-Write Tasks

ONO, Masumi

1. Introduction

Recently, much attention has been paid to integrated writing tasks because of their validity and authenticity in assessing learners' writing ability (Plakans, 2015; Weigle, 2004). Integrated writing tasks are also considered to be better than independent writing tasks in terms of fairness (Yang, 2009) because the test takers' background knowledge of the topic and topic familiarity are controlled to some degree; test takers are required to produce responses based on what is written in the given sources and thus they do not have to rely on their background knowledge of the topic. With these positive aspects, while integrated writing tasks are gradually becoming popular as an English proficiency test (e.g., Plakans, Gebril, & Bilki, 2016; Shin & Ewert, 2015), independent writing tasks are also commonly used in proficiency tests. However, the former task is seen as more challenging than the latter because of the inevitable element of source texts used for accomplishing the task (Ohta, Plakans, & Gebril, 2018).

In tertiary education, source use, which is involved in integrated writing tasks, is one of the key elements for academic success because several courses require students to produce writing assignments using sources. Students need to comprehend what is written in the assigned reading materials or self-selected source texts to respond to the question or the topic given in the assignment. Thus, integrated writing skills, including appropriate source use, play an essential role in accomplishing the course or the program in any discipline. Further, the use of integrated writing tasks or tests contributes to the positive washback effect (Weigle, 2004; Yang, 2009) because teachers are expected to teach and learners are encouraged to learn corresponding skills in their programs.

Types of integrated writing tasks vary depending on the purpose of the task involving writing, reading, listening, or speaking. The more the number of skills combined, the more difficult and complex the task becomes. The overall task difficulty is also affected by the genre of source texts. For instance, if the learner is required to read a short narrative passage and write a summary of it, the difficulty will be less than reading a research article to write a summary. From a different perspective, integrated writing is categorized into three types (Weigle & Parker, 2012): text-based writing (e.g., summary writing: Asención Delaney, 2008; Hijikata-Someya, Ono, & Yamanishi, 2015; Keck, 2006, 2014; Ono, Yamanishi, & Hijikata, 2019; Yamanishi, Ono & Hijikata, 2019; Yu, 2013), situation-based writing (e.g., writing responses to letters or emails: Test of English for International Communication by Educational Testing Service), and theme-based writing (e.g., opinion or persuasive essays: Asención Delaney, 2008; Ohta et al., 2018; Plakans, 2009; Plakans, 2015; Shin & Ewert, 2015). Although a wide range of reading-to-write tasks has been investigated extensively, as shown above, few studies have examined listening-to-write tasks (e.g., Cumming et al., 2005, 2006; Soleimani & Mahdavipour, 2014).

One of the studies investigating listening-to-write tasks were conducted by Cumming et al. (2005) who examined the discoursal features of integrated reading-writing tasks, integrated listening-writing tasks, both of which were developed for Next Generation TOEFL®, and independent writing tasks for the TOEFL® Essay. As one of the discoursal features, they compared the verbatim source use between the reading-writing and the listening-writing tasks at three English proficiency levels based on independent writing scores. The results indicated that the number of verbatim expressions from the source decreased for the reading-writing tasks as test takers' proficiency increased, whereas the number of verbatim expressions increased for the listening-writing tasks as test takers' proficiency increased. They also reported an interesting pattern for the group with moderate proficiency in that this group differed in verbatim source use between two topics in the reading-writing tasks, but no significant difference appeared in the use of verbatim phrases between two topics in the listening-writing tasks. Cumming et al. (2005) explained that "the extent of verbatim phrases in these tasks appears to interact in complex ways with examinees' proficiency levels, the medium of comprehension of source materials, memory factors, and task characteristics and conditions as well" (p. 28). Thus, several factors affect test takers' performances on the integrated writing tasks and that their textual borrowing

behavior seems to differ between the reading-writing and listening-writing tasks and across text takers' English proficiency levels. In addition, source texts (e.g., topic, genre, and rhetorical structure) may affect test takers' performance on integrated writing tasks, and therefore, it needs to be investigated.

Soleimani and Mahdavipour (2014) also compared reading-to-write tasks and listening-to-write tasks (i.e., summary writing based on listening or reading materials) in terms of the textual features of summaries produced by high and low proficiency groups of Iranian university students. These integrated writing tasks were modified versions of the Test of English as a Foreign Language Internet-based Test (TOEFL iBT®). In the integrated writing task in TOEFL iBT®, a test taker reads a passage and then listens to a lecture related to the same topic to summarize the main points of both source texts. However, Soleimani and Mahdavipour (2014) developed two separate tasks: reading-to-write and listening-to-write tasks. The results showed that in the listening-to-write task, the highly proficient group used verbatim phrases from the source materials more than the low-proficiency group, while in the reading-to-write task, the low-proficiency group directly borrowed more instances from the sources without appropriate textual borrowing strategies. It is, then, questionable why even the highly proficient group relied on the direct use of phrases from the sources without employing appropriate textual borrowing strategies, such as paraphrasing or quoting. The result implies that highly proficient students do not necessarily have adequate knowledge of source use and that teachers need to teach about using sources appropriately and effectively in the integrated writing tasks regardless of students' proficiency levels.

Focusing on TOEFL iBT® integrated writing tasks, where reading, listening, and writing are combined, Yang and Plakans (2012) investigated second language (L2) writers' performances. They found that even if L2 writers have a high ability to write, it does not mean that writers can employ source materials in their compositions appropriately. This result indicates that source use strategies are part of writing ability, but they need to be taught explicitly. It also suggests that the use of sources is a developmental skill, in which having knowledge of source use does not always guarantee writers' actual practice of appropriate source use. Thus, textual borrowing skills, especially paraphrasing, need to be researched in summary writing tasks.

Regarding paraphrasing, Keck's (2006, 2014) studies provide important insights. In

her (2006) study, she compared first language (L1) and L2 writers' paraphrasing strategies in a reading-to-write task in the form of summary writing. She developed four categories of paraphrasing: Near Copy, Minimal Revision, Moderate Revision, and Substantial Revision. The results showed that L1 writers' summaries contained Moderate Revision and Substantial Revision significantly more than L2 writers' summaries, while L2 writers relied on Near Copy. Furthermore, Keck (2014) found that novice writers of L1 and L2 often used verbatim phrases from the source, unlike experienced L1 and L2 writers who succeeded in paraphrasing. This result implies that the expertise of a target language could influence writers' paraphrasing behavior. It was also reported that the small number of L2 writers who relied heavily on copying strategies contributed to the negative overall result of L2 writers' misbehavior of paraphrasing attempts. While Keck's (2006, 2014) studies focused on paraphrasing in reading-to-write tasks, the number of studies on paraphrasing strategies in listening-to-write tasks is scarce. It is not clear how L2 writers use paraphrasing strategies to produce summaries in listening-to-write tasks. Therefore, the current study intends to fill this gap by investigating the paraphrasing behavior of novice L2 writers in summary writing as a listening-to-write task.

Although the TOEFL iBT® integrated writing task involves listening ability and is acknowledged widely, there is not much research that sheds light on listening-to-write tasks. Moreover, less is known about summary writing tasks, in which learners listen to a lecture to write a summary, compared to reading-based summary writing tasks. As in the TOEFL iBT® integrated writing task, listening to a lecture is one of the basic skills in tertiary education because many courses offer lectures on specific topics, and students need to take notes and understand the key ideas or concepts. Thus, listening to a lecture and writing a summary is a fundamental academic skill that university students need to master. However, due to the limited number of studies, it remains unclear whether university students can produce a summary after listening to a lecture in English.

Therefore, this study aims to investigate university students' ability to produce summaries in listening-to-write tasks. The three research questions are as follows:

1. Do ratings of summaries produced by Japanese university students differ between two topics of listening-to-write tasks?

2. To what extent do Japanese university students paraphrase the source text in listening-to-write tasks?

3. What are the different features in high-graded and low-graded summaries?

2.   Methods

2.1 Participants

The participants were 70 Japanese English as a foreign language (EFL) learners at a university in Japan, including first- and second-year students majoring in law or political science. The participants' English language proficiency was regarded as lower intermediate, equivalent to Level B1 in the Common European Framework of Reference for Languages based on the results of the Quick Placement Test (University of Cambridge Local Examinations Syndicate, 2001).

They were enrolled in compulsory English language classes in the Faculty of Law. In the classes, they were taught how to write a summary and practiced summary writing based on reading passages. Therefore, they were somewhat familiar with summary writing tasks as the reading-to-write task. They learned about the basics of source use, such as citations and quotations, and plagiarism. They wrote an essay writing assignment using sources, which is categorized as an integrated writing task, in other words, a theme-based writing task. Hence, they had experience of several types of reading-to-write tasks and knowledge of source use. While they regularly participated in listening exercises, including dictation and listening comprehension, they hardly practiced summary writing in the form of listening-to-write tasks. It was also not certain that to what extent they were familiar with listening to a lecture given in English. Their experience of listening-to-write tasks, especially in the form of summary writing, tended to be limited and varied from student to student. Thus, most of them seemed to be regarded as novice writers in this task.

2.2 Material

Two task prompts were selected from the exercise book of the TOEFL iBT® (Wadden, Hilke, & Hayakawa, 2014) for the listening-to-write tasks. One topic was "manned space flight" and the other was "successful business" (hereafter Space and Business). These topics were deemed reasonable because they were not related to students' major, and no background knowledge integral to their major was required to understand the topics. The reading passages were read aloud by a native English-speaking teacher and

Table 1 *Features of Lectures Used for the Listening-to-Write Tasks*

Topic	*N* of words	*N* of paragraphs	*N* of sentences	Flesch–Kincaid Grade Level
Space	244	5	14	10.9
Business	231	3	15	10.3

recorded to generate the listening-to-write tasks for this study: both lectures lasted approximately two minutes and were spoken at a rate of 130 words per minute.

　　Furthermore, the two materials were considered comparable, as they were similar in terms of the number of words and readability measured by the Flesch–Kincaid Grade Level (see Table 1). Both lectures contained three main points about the topic that should be included in the summary, but they were different in the structure. Space explicitly showed the lecturer's stance and key aspects in the introduction, followed by explanations of each aspect with a clear conclusion, while Business had a less clear structure because the lecturer's stance was not stated clearly at the beginning, and a conclusion was not provided. Another potential difference between the two lectures concerned with vocabulary. To identify the level of vocabulary in both lectures, the software regarding the academic word list (Smith, 2020) was used. Consequently, Business had less-frequent academic words more than Space, which could imply greater difficulty in terms of vocabulary, but the latter seemed to have more technical terms related to the topic of space exploration. Although both topics were general topics, Business could be categorized in the humanities, whereas Space was viewed as a topic in the sciences. Despite these features of the lectures mentioned above, the difficulty level of both lectures was considered reasonable, given the English proficiency of the participants.

2.3 Procedures

　　The participants were divided into Group A (*n* = 38) and Group B (*n* = 32). Group A was provided with the topic of Space, while Group B was given the topic of Business. All participants completed a listening-to-write task in a classroom, where each of them was able to use a PC. Before the task began, they were instructed to listen to the lecture twice and write a summary in English using the PC. They were also provided with a piece of blank paper, so that they could take notes while listening to the lecture. More specifically, the following instructions and prompts were given to the participants in written form:

Listen to a lecture and summarize the points made in the lecture in English. Readers of the summary will be those who have never listened to the lecture.

- You listen to a lecture twice and have 15 minutes to plan and write your response.

- Your response will be judged based on the quality of your writing and how well your response presents the points in the lecture. Typically, an effective response will be between 60 and 80 words.

- In writing a summary, include only the main points and leave out details. Use your own words and do not copy sentences from the original. You are NOT allowed to use a dictionary.

After completing the task, the typed summaries were collected.

2.4 Scoring

The summaries were graded using an analytic scoring rubric developed by Yamanishi, Ono, and Hijikata (2019). This rubric is a four-point analytic scale and has five dimensions: *content*, *paraphrase (quantity)*, *paraphrase (quality)*, *language use*, and *overall quality*. *Content* refers to whether the main ideas are grasped and developed with secondary information. *Paraphrase (quantity)* identifies the extent to which the expressions in the summary are paraphrased. *Paraphrase (quality)* identifies whether paraphrases are made effectively without using a large number of verbatim phrases from the lecture. *Language use* refers to whether the use of vocabulary and sentence structure is appropriate. *Overall quality* judges the overall quality of the summary from a holistic point of view, whether the response corresponds to the task requirements. This rubric was selected for this study because it was specifically developed for summary writing tasks and it aimed to be used for pedagogical purposes, instead of high-stake testing purposes. Although Yamanishi, Ono, and Hijikata (2019) employed summary writing as a reading-to-write task, this study used it for a listening-to-write task. Thus, the skill required in the task differed between the two studies, but the rubric was considered useful in identifying the quality of students' performance on the listening-to-write task. Furthermore, this rubric emphasizes the textual borrowing behavior comprising two different dimensions of paraphrase in terms of quantity and quality. Therefore, it seemed helpful to understand whether students use the

source text appropriately and effectively and to what extent they can paraphrase the original text in the summaries.

An experienced EFL teacher participated in this study to mark 70 summaries produced by the listening-to-write task. This rater has taught English for 21 years at the university and elsewhere and is currently a postgraduate student in an MA TESOL program. Before scoring, as a form of rater training, the rater was provided with anchor summaries that had been marked to illustrate the features of summaries with different scores assigned to each dimension. The rater was also asked to read all the materials carefully and mark several summaries with the two topics to familiarize him with the rubric, understand the appropriate scores, and adjust the severity of the scoring. The rater was advised to ask questions any time before and during the scoring process. After completing the scoring, the rater was instructed to fill out the questionnaire regarding the educational and teaching background and the scoring experience of summaries.

2.5 Data analysis

The summaries were analyzed quantitatively and qualitatively. First, the researcher examined whether the ratings of the summaries with the two topics differed using an independent samples t-test. Next, the number of verbatim phrases in the summaries was calculated using a text comparison tool named difff. This free software is useful when we need to know the number of overlapping words between two texts (Yamanishi & Ono, 2017). For summary writing, for instance, it makes it possible to compare the original text and the summary in terms of how many words overlap. In other words, we could obtain information about how many words each of the summaries contained in terms of verbatim phrases from the original text. Thus, by calculating the percentage of overlapping words, we can judge whether the summary is paraphrased effectively and the writer's textual borrowing behavior is appropriate. Since the number of words in each summary differed, the number of overlapping words was divided by the number of words in the summary to obtain a percentage of the overlapping words for each summary. The researcher also administered an independent samples t-test to reveal whether there was a difference between the two topics in terms of verbatim source use in the summaries. Finally, high-graded and low-graded summaries were compared qualitatively to identify the different textual features in the summaries. The researcher was particularly interested in gaining

insights into what challenges the students had in producing summaries in the listening-to-write tasks based on the analysis of the written summaries.

3.   Results and Discussion

3.1 Descriptive features of the summaries

Descriptive statistics of the summaries produced by the listening-to-write tasks are shown in Table 2. Summaries in Group A had 67.47 words on average, while those in Group B had 68.19 words. Both were within the word limit (i.e., 60-80 words) specified in the instructions given before the task began.

The results of the independent samples t-test showed that the two groups did not have a significant difference in terms of the average number of words in the summaries: t (68) = -0.153, p = .879, r = .02. Thus, the two groups were considered comparable based on the previously reported results.

Table 2 *Descriptive Statistics of the Summaries*

Group	Topic	*n* of Participants	Mean	*Min*	*Max*	*SD*
A	Space	38	67.47	27	109	22.061
B	Business	32	68.19	30	97	15.867
Total		70	67.80	27	109	19.345

3.2 Do ratings of summaries produced by Japanese university students differ between two topics of listening-to-write tasks?

Table 3 shows the rating results of the summaries regarding five dimensions. This indicates that in all five dimensions, mean scores in Business are higher than those in Space, which leads to a higher total score in summaries for Business than for Space. These results imply that the students summarized better with the former than the latter.

The results of independent samples t-tests showed that there was a significant difference between the two topics only in the dimension of *paraphrase (quality)*: t (68) = -3.284, $p <$ .05, r = .37. The effect size was considered to be medium. This means that the score of *paraphrase (quality)* in Business is significantly higher than that in Space. This result indicates that students' paraphrasing attempts were more successful in Business than

Table 3 *Comparison of Ratings of the Summaries Between Two Topics*

Dimension of Rubric	Topic	*n* of Participants	Mean	*SD*	*p*
Content	Space	38	2.79	0.905	0.0748
	Business	32	3.16	0.767	
Paraphrase (Quantity)	Space	38	2.79	0.963	0.1922
	Business	32	3.09	0.963	
Paraphrase (Quality)	Space	38	2.84	0.916	0.0016
	Business	32	3.47	0.621	
Language use	Space	38	2.89	0.689	0.1468
	Business	32	3.13	0.609	
Overall quality	Space	38	2.84	0.916	0.1510
	Business	32	3.16	0.884	
Total	Space	38	14.16	4.169	0.0496
	Business	32	16.00	3.408	

that in Space. Moreover, *SD* in Space is higher than that in Business, implying that individual students varied in their paraphrasing attempts in Space.

However, mean scores for the dimension of *paraphrase (quantity)* did not statistically differ between the two topics (t (68) = -1.317, $p > .05$, $r = .16$), which means that the students in both groups paraphrased almost the same amount of information in the summaries regardless of the topics. According to the rubric (Yamanishi, Ono, & Hijikata, 2019, p. 19), Score 2 indicates "fair" in that "Can paraphrase only from 25% to less than 50% of the expressions included in the summary in one's own words" while Score 3 means "good" in that "Can paraphrase from 50% to less than 80% of the expressions included in the summary in one's own words." In this study, the mean score for *paraphrase (quantity)* for Space was 2.79, whereas it was 3.09 for Business. Thus, students' paraphrasing behavior is seen as "good" for the summaries of Business, yet it is regarded as "fair" or "good" for Space. Soleimani and Mahdavipour (2014) found that highly proficient students used verbatim phrases in the listening-to-write task more than low-proficiency students. This result is interesting because the former students are supposed to have a better understanding of the content, yet they somehow used more verbatim expressions in the summaries without paraphrasing. Although the current study did not compare groups with different proficiency groups, students with lower-intermediate proficiency showed

different paraphrasing behaviors between the two topics. This finding is somewhat inconsistent with the result reported by Cumming et al. (2005) in that writers with medium proficiency did not differ in the use of verbatim phrases between two topics in the listening-writing tasks, while those with high proficiency contained more verbatim phrases in the listening-writing tasks. In order to gain more insight into students' paraphrasing behavior in this study, scrutiny is conducted in the following sections.

Furthermore, there were no significant differences in mean scores between the two topics in the dimensions of *language use* and *overall quality* while *content* was marginally significant (see Table 3). However, the total scores were statistically different between the two topics: t (68) = -1.999, $p < .05$, $r = .24$; the effect size was small. In other words, the total score of summaries for Business was higher than that for Space. This result implies that the Business lecture tended to be easier for students to summarize than that of Space. Although the two lectures had comparable features, this study found differences in scores between the two lectures. This result can be accounted for by the students' topic familiarity and the textual features of the sources. Textual features such as the rhetorical organization and cohesive devices are likely to affect difficulty in summarizing the source text, as Hirvela (2004) argues. In this sense, Space seemed to have a clearer structure, including a thesis statement and three key points and cohesive devices than Business. However, the students may have been more familiar with the topic of Business, and it may have led to their better comprehension of the content of the lecture, resulting in better summary writing performances. On the other hand, the lecture on Space had more technical terms, which may have made it difficult to understand the content of the lecture. The influence of topics and textual features needs to be investigated in future studies.

3.3 To what extent do Japanese university students paraphrase the source text in listening-to-write tasks?

Table 4 shows that both groups paraphrased nearly three-fourths of the expressions in the summaries: 73.94 percent for Group A and 72.89 percent for Group B. These results indicate that both groups actively paraphrased in summarizing the lecture in a written form and may have understood that they need to use their own words as much as possible in summary writing based on the instruction given in class and before the task. However, as the percentage of the paraphrased expressions indicates, the students could not paraphrase

Table 4 *Percentage of Paraphrased Expressions in the Summaries*

Group	Topic	*n* of Participants	No. of Paraphrased Words	% of Paraphrased Expressions	*Min*	*Max*	*SD*
A	Space	38	50.13	73.94	60.00	86.21	7.241
B	Business	32	49.91	72.89	49.02	88.75	8.399
Total		70	50.03	73.46	49.02	88.75	7.752

more than 80 percent of the expressions in the summaries, which would deserve Score 4 in *paraphrase (quantity)* in the rubric (Yamanishi, Ono, & Hijikata, 2019). Thus, their paraphrasing behavior is not viewed as "very good" and could be improved. These results support the results reported for research question 1 in the previous section.

The results of the independent samples *t*-test showed that the two groups' paraphrasing behavior did not differ significantly in terms of the quantity of paraphrased expressions in the summaries: $t(68) = 0.558$, $p = .578$, $r = .07$. In other words, the two topics did not differ in the effectiveness of paraphrasing in the summaries. These results are consistent with the findings shown in research question 1, where no significant difference was found in *paraphrase (quantity)*. Thus, it can be said that the rubric developed by Yamanishi, Ono, and Hijikata (2019) works well and is a good indicator to identify writers' paraphrasing behavior from a quantitative point of view, although only one rater participated in the scoring with rater training in this study.

3.4 What are the different features in high-graded and low-graded summaries?

Although the two groups did not differ significantly in their paraphrasing behavior from a quantitative perspective, it is necessary to closely examine the quality of high-graded and low-graded summaries to gain insight into students' summary writing behavior. An example of a high-graded summary for Space is shown in Figure 1. In this summary, which scored a full mark (i.e., Score 20), the first sentence explicitly states the lecturer's stance with three key aspects to support by saying that "A project of human's space journey must be stopped for these three reasons: dangers of activities in space, huge costs for space travel, and few practicalities of it." These three aspects are elaborated in the following sentences,

and a clear conclusion is written at the end of the summary, which is similar to the structure of the source material. Although there are minor errors in language use, the summary contains many effective paraphrased phrases, which deserve Score 4 for both *paraphrase (quantity)* and *paraphrase (quality)*.

A project of human's space journey must be stopped for these three reasons: dangers of activities in space, huge costs for space travel, and few practicalities of it. First, people have a lot of difficulties to move in non-oxygen area and some possibilities to be crashed into some space dust. The other hand, even only one trip for space costs us for tremendous money. Finally, this project is not practical. It will be push us infinite troubles eternally. In conclusion, the space travel cannot be allowed. (Th2-4: 86 words)

Note: Score 20, *content* = 4, *paraphrase (quantity)* = 4, *paraphrase (quality)* = 4, *language use* = 4, and *overall quality* = 4.

Figure 1. Example of a high-graded summary for Space

An example of a high-graded summary for Business is shown in Figure 2. In this instance, all the dimensions of the rubric also scored full marks. This summary clearly states the main ideas using the writer's own words. Language use is not perfect, yet it does not obscure the meaning of the ideas in the summaries.

The keys to sustainable success of marketing commercial products are new products and true innovation, which means the company is required to improve the existing products' nobility and quality. Moreover, the process of creating new products should include team work or interaction in the company. It's important to share their opinions and consult with experts. Besides, the new product should be produced sustainably, so the company has to create new product considering not only its appearance but also effects on the environment. (Sa2-14: 82 words)

Note: Score 20, *content* = 4, *paraphrase (quantity)* = 4, *paraphrase (quality)* = 4, *language use* = 4, and *overall quality* = 4.

Figure 2. Example of a high-graded summary for Business

On the other hand, features of low-graded summaries are as follows: summaries (a) do not fully contain the main ideas; (b) contain incorrect information or information that is not mentioned in the source material; (c) are not paraphrased substantially; (d) do not use

appropriate language; and/or (e) do not meet the word limit. An instance of a low-graded summary for Space is shown in Figure 3, which is the case of (a) and (b). In the summary in Figure 3, the wavy line represents incorrect ideas. The writer of this summary wrote a claim in the first line that "'Man space flight' is one of the kinds of program that nations should continue." This claim is the opposite of the source material. It indicates that the writer failed to comprehend the thesis statement of the lecture, although he/she managed to contain the three main points to support the claim in the summary.

"Man space flight" is one of the kinds of program that nations should continue. Space has three points. First, it is dangerous one. Space is a mysterious one, so it needs a long-time effect, which provides astronomy some dangerous. Second, it has a lot of cost. The program is corporate with many types of professionals. Finally, it is impractical. Space is the origin of human being, and it is an eternal one. to research the space leads to progress for human. (Th2-3: 81 words)

Note: Score 10, *content* = 2, *paraphrase (quantity)* = 2, *paraphrase (quality)* = 2, *language use* = 2, and *overall quality* = 2. Emphasis is mine.

Figure 3. Example of a low-graded summary for Space

Furthermore, the dotted lines show information that was not mentioned in the lecture, which means that the writer added information using his/her background knowledge or just made up information in producing the summary. This kind of summary writing behavior implies that the writer may not have understood how to write an effective summary without knowing that the writer has to use the information mentioned in the original source, rather than incorporating his/her background knowledge and opinions about the topic into the summary. Thus, this summary seems to deserve Score 2 for *content*, which eventually leads to inappropriate paraphrasing behavior in the dimensions of *paraphrase (quantity)* and *paraphrase (quality)*.

　　Another instance of a low-graded summary for Space represents the case of (c) in Figure 4. In this summary, the underlined phrase shows the verbatim expression directly borrowed from the lecture.

> Man space flight is <u>venture</u>. <u>The speaker says that Man space flight should be abandoned.</u> There are three reasons that he talk about. First, it contains several dangers. <u>No one can predict</u> what will happen in space. Second, it is not justified. Third, it is impractical.it is needed much money. So, <u>the speaker says that man space flight should be abandoned.</u> (Th2-12: 61 words)

Note: Score 14, *content* = 3, *paraphrase (quantity)* = 3, *paraphrase (quality)* = 2, *language use* = 3, and *overall quality* = 3. Emphasis is mine.

Figure 4. Example of a low-graded summary for Space

This summary also repeats the same ideas twice in the double-underlined sentences, namely, "the speaker says that man space flight should be abandoned." This reflects the writer's limited ability to paraphrase effectively resulting in Score 2 for *paraphrase (quality)*, although this claim is true to the lecture. Thus, the writer shows a correct understanding of the lecture but fails to paraphrase substantially. In addition, the writer seemed to fail to develop the main ideas fully, although the three points were vaguely mentioned. The second point was written as "Second, it is not justified." and the third point was shown as "Third, it is impractical." in the summary. However, it is unclear from the summary that what is not justified without giving explanations. In fact, in the lecture, the high cost of space flight is problematized and is not justified. The lecture also explains that impracticability is related to human beings' physical capacities in the spacecraft, instead of the cost of space flight. Thus, this summary seems to lack coherence and elaboration of the main ideas and contains misunderstanding of the information in the lecture.

The following low-graded summary for Business is mainly regarded as the case of (c) and (e) in Figure 5. The underlines denote the verbatim expressions from the source.

> <u>Commercial products are essential</u> to business success. Good companies try to have stress on true innovation and improve <u>the quality and the novelty of</u> products. True innovation requires researches and creative ideas, so it is important to work with various people and share new ideas. Finally, we should make products durable. (Th3-3: 51 words)

Note: Score 14, *content* = 3, *paraphrase (quantity)* = 3, *paraphrase (quality)* = 2, *language use* = 3, and *overall quality* = 3. Emphasis is mine.

Figure 5. Example of a low-graded summary for Business.

This summary contains most of the important ideas, but it also seems to fail to develop the main ideas substantially without explaining each point fully like the summary in Figure 4. Another challenge that the writer had in this summary is associated with paraphrasing. The use of verbatim phrases seems to contribute to Score 2 for the dimension of *paraphrase (quality)*. In this summary, the writer applied a patchwriting strategy (Pecorari, 2003), where the original expressions from the source are hardly changed in terms of vocabulary and grammar. Thus, the sentences used in the summary tend to look like those in the source. For instance, an example of patchwriting is shown below.

(1) Example
［Summary］Commercial products are essential to business success.

［Original source］Well-designed commercial products are essential for ongoing business success.

In this example, the sentence in the summary is shorter than the original, but the vocabulary and grammar of the two sentences are almost the same. This instance indicates that the writer's paraphrasing is not substantial and is regarded as Near Copy (Keck, 2006, 2014). Furthermore, this finding is consistent with Keck (2014) in that the small number of L2 writers heavily relies on copying strategies, which leads to the conclusion that, overall, L2 writers tend to use verbatim expressions from the source. The finding also partly supports Keck (2014) in that L2 novice writers show such textual borrowing behaviors compared to L2 experienced writers. Although the current study did not compare novice and experienced writers' summary writing skills, the participants varied considerably in their paraphrasing strategies, as shown in the high-graded and the low-graded summaries.

　　Another example also shows an instance of patchwriting in the same summary.

(2) Example
［Summary］Good companies try to have stress on true innovation and improve the quality and the novelty of products.

［Original source］"Never copy" is the golden rule of <u>true innovation</u>. The top <u>companies</u> always seek to <u>improve</u> both <u>the quality and the novelty of</u> their product lines.

In example (2), two sentences from the original source are combined into one sentence in the summary, which is appropriate. However, the summary uses almost similar words and phrases from the source, resulting in patchwriting. The verbatim phrase "the quality and the novelty of" has six consecutive words. In Cumming et al. (2005) and Yang and Shi (2012), three consecutive words or more copied from the source are verbatim source use, while Yamanishi, Ono, and Hijikata (2019) consider a string of more than four words as verbatim phrases. Thus, the definition of verbatim source use varies from study to study. Nevertheless, it is important for writers to understand that patchwriting could be treated as plagiarism (Marshall, 2017) or a developmental textual borrowing skill (Pecorari, 2003).

The qualitative analysis of the low-graded summaries indicates that some writers failed to grasp the lecturer's claim or develop the main ideas without elaborating them or showing incorrect information. These are related to the ability to comprehend the lecture. Although they were given a piece of blank paper to take notes during the lecture, some seemed to miss important information or failed to take notes while listening. Note-taking skills are highly important, especially in listening to long scripts. Some writers also added ideas that did not appear in the lecture or included their opinions about the topic. This case seems to be either writers do not understand that they are not supposed to do so or miscomprehend the information given in the lecture. Another challenge that writers encounter is related to source use behavior. Although they were instructed to paraphrase as much as possible instead of using verbatim phrases from the source, some failed to demonstrate adequate paraphrasing, relying on the expressions and sentence structure used in the lecture. As Keck (2006) reported, L2 writers tended to rely on copying excerpts from the source text, unlike L1 writers of English who used paraphrasing significantly more often, ranging from Moderate Revision and Substantial Revision. Thus, the result of this study partly supports Keck's (2006) study in that some L2 writers depend on copying instead of paraphrasing. In some summaries, patchwriting was used, which resulted in a low score for the dimension of *paraphrase (quality)*. Although patchwriting is thought to be a process of developing textual borrowing strategies, students need explicit instruction on

and adequate practice of source use to avoid plagiarism and acquire appropriate textual borrowing skills in integrated writing tasks and writing assignments in their courses.

4. Conclusion

This study has shed light on university students' integrated writing skills in the listening-to-write task. The findings are summarized as follows. First, the summaries' scores differed in *paraphrase (quality)* and total scores between the two topics. More specifically, the topic of Business was significantly higher in scores than that of Space. Second, with respect to paraphrasing behavior, the students paraphrased 73.46 percent of the expressions in the summaries on average, and there was no significant difference between the two topics. Third, the high-graded and low-graded summaries differed mainly in terms of *content* and *paraphrase*. Some students who wrote low-graded summaries miscomprehended the lecture's main claim, failed to develop the main points, or added information that was not presented in the lecture. They also tended to rely on verbatim phrases directly drawn from the lecture, resulting in patchwriting.

As for pedagogical implications, university students are encouraged to practice listening lecturers while taking notes, since the listening-to-write tasks require students to understand the content and transfer the information in written form. It is highly important not only to comprehend the main claim or ideas, but also to maintain the correct information as notes. However, without note-taking, it is difficult to fully remember the important information provided in the listening material, particularly when it is long and complex. In order for students to familiarize themselves with lectures, the rhetorical structure of lectures and cohesive devices such as signposting can be taught. For instance, a lecture usually begins with an introduction about a topic and a claim followed by a main body with supporting ideas, and then a conclusion is made at the end of the lecture. However, there is an exception to Business, in which a clear claim and a conclusion are not present explicitly. In addition, teaching how to take notes in an organized way is also helpful because a lecture usually has several main points. Another implication is that teachers need to teach how to write an effective summary. Although the participants in this study received summary writing instruction in advance, some students included information outside the source materials for some reasons. This may imply that

they were not accustomed to summary writing or misunderstood how to write a summary. Thus, they should be reminded that only information from the source should be used in integrated summary writing tasks unless extra instruction is provided. Finally, students need to be taught and practice paraphrasing strategies. As Keck (2014) discusses, even novice L1 writers have difficulty in paraphrasing in summary writing, novice L2 writers also need to learn and practice paraphrasing strategies. In doing so, a text comparison tool difff, which was used in this study, can be used when teaching how to paraphrase. For instance, students are instructed to listen to or read a passage and paraphrase as much as possible and then check how many words are the same as the source using the software. In this way, the text comparison tool could be a supplemental tool in finding the exact number of words or percentage each summary contains paraphrased expressions and in supporting the learning of how to paraphrase in summary writing. It is possible that some students may not realize how frequently they use verbatim phrases from the source, visualizing the comparison results between their paraphrased phrases and the source material may enhance their understanding of patchwriting and effective paraphrasing. Furthermore, Yamanishi, Ono, and Hijikata's (2019) analytic rubric may also be a useful tool in teaching paraphrasing both quantitatively and qualitatively because it clearly shows the different degree of paraphrasing. Teachers can use the rubric in summary writing instruction, and students can check their own paraphrases by looking at the rubric for self-checking before and after writing a summary. Since paraphrasing plays an essential skill not only in summary writing but also in academic writing (Keck, 2006), this skill can be taught by using supporting tools that help students learn more effectively, as discussed above.

Despite these pedagogical implications, this study has limitations. First, this study involved only students with lower-intermediate proficiency. If different proficient students had participated in this study, a comparison of different groups would have been possible. Second, this study focused on students' compositions and did not ask them to participate in a questionnaire or an interview after the task ended. These kinds of follow-up procedures would be useful in eliciting their perceptions of task difficulty and source materials such as topic familiarity and difficulty level of comprehension.

Future studies need to compare reading-to-write and listening-to-write tasks using several topics or materials to reveal whether students' performances differ between the two

task types and depending on topics. Furthermore, it is worth investigating whether different proficiency groups differ in their integrated writing strategies. Cumming et al. (2005) and Soleimani and Mahdavipour (2014) suggest that integrated writing tasks involving listening skills have a complex nature, as students' proficiency may not necessarily indicate better ability in source use. Thus, future research needs to continue paying careful attention to the aspect of textual borrowing skills integral to integrated writing tasks.

Acknowledgements

References

Asención Delaney, Y. (2008). Investigating the reading-to-write construct. *Journal of English for Academic Purposes, 7*(3), 140–150. https://doi.org/:10.1016/j.jeap.2008.04.001

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Difference in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing, 10*, 5–43. https://doi.org/10.1016/j.asw.2005.02.001

Cumming, A., Kantor, R. Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2006). *Analysis of discourse features and verification of scoring levels for independent and integrated prototype written tasks for the new TOEFL* (TOEFL Monograph No. MS-30). Princeton, NJ: ETS.

Hijikata-Someya, Y., Ono, M., & Yamanishi, H. (2015). Evaluation by native and non-native English teacher-raters of Japanese students' summaries. *English Language Teaching, 8*(7), 1–12. https://doi.org/10.5539/elt.v8n7p1

Hirvela, A. (2004). *Connecting reading & writing in second language writing instruction*. Ann Arbor: Michigan University Press.

Keck, C. (2006). The use of paraphrase in summary writing: A comparison of L1 and L2 writers. *Journal of Second Language Writing*, *15*, 261–278. https://doi.org/10.1016/j.jslw.2006.09.006

Keck, C. (2014). Copying, paraphrasing, and academic writing development: A re-examination of L1 and L2 summarization practices. *Journal of Second Language Writing, 25*, 4–22. https://doi.org/10.1016/j.jslw.2014.05.005

Marshall, S. (2017). *Advance in academic writing: Integrating research, critical thinking, academic reading and writing*. Montréal: Pearson.

Ohta, R., Plakans, L., & Gebril, A. (2018). Integrated writing scores based on holistic and multi-trait scales: A generalizability analysis. *Assessing Writing, 38*, 21–36. https://doi.org/10.1016/j.asw.2018.08.001

Ono, M., Yamanishi, H., & Hijikata, Y. (2019). Holistic and analytic assessments of the TOEFL iBT® integrated writing task. *Japan Language Testing Association Journal, 22*, 65–88. https://doi.org/10.20622/jltajournal.22.0_65

Pecorari, D. (2003). Good and original: Plagiarism and patchwriting in academic second-language writing. *Journal of Second Language Writing, 12*(4), 317–345. https://doi.org/10.1016/j.jslw.2003.08.004.

Plakans, L. (2009). Discourse synthesis in integrated second language writing assessment. *Language Testing, 26*(4), 561–587. https://doi.org/10.1177/0265532209340192

Plakans, L. (2015). Integrated second language writing assessment: Why? what? how? *Language and Linguistics Compass, 9*(4), 159–167. https://doi.org/10.1111/lnc3.12124

Plakans, L., Gebril, A., & Bilki, Z. (2016). Shaping a score: Complexity, accuracy, and fluency in integrated writing performances. *Language Testing, 36*(2), 161–179. https://doi.org/10.1177/0265532216669537

Shin, S.-Y., & Ewert, D. (2015). What accounts for integrated reading-to-write task scores? *Language Testing, 32*(2), 259–281. https://doi.org/10.1177/0265532214560257

Smith, S. (2020). AWL highlighter & gapfill. EAP FOUNDATION.COM. Retrieved from https://www.eapfoundation.com/vocab/academic/highlighter/ (August 9, 2020)

Soleimani, H., & Mahdavipour, M. (2014). The effect of variations in integrated writing tasks and proficiency level on features of written discourse generated by Iranian EFL learners. *The Journal of Teaching Language Skills*, 6(2), 131–159.

Text Comparison Tool difff (2020). Retrieved from https://difff.jp/ (August 9, 2020)

University of Cambridge Local Examinations Syndicate. (2001). *Quick placement test*. Oxford University Press.

Wadden, P., Hilke, R., & Hayakawa, K. (2014). *TOEFL® test: Writing mondai 100 [TOEFL® test: Writing exercises 100]*. Tokyo: Obunsha.

Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing, 9*(1), 27–55. https://doi.org/10.1016/j.asw.2004.01.002

Weigle, S. C., & Parker, K. (2012). Source text borrowing in an integrated reading/writing

assessment. *Journal of Second Language Writing, 21*(2), 118–133. https://doi.org/10.1016/j.jslw.2012.03.004

Yamanishi, H., & Ono, M. (2017). Youyaku shidou wo hojosuru tsuuru toshiteno ruuburikku no katsuyou ［Use of rubrics to support summary writing instructions as a tool］. In A. Mizumoto (Ed.), *ICT wo katsuyoushita eigo akademikku raitingu shidou: Shien tsuuru no kaihatsu to jissen ［Teaching academic writing using ICT: Development and practice of supporting tools］* (pp. 92–106). Tokyo: Kinseido.

Yamanishi, H., Ono, M., & Hijikata, Y. (2019). Developing a scoring rubric for L2 summary writing: a hybrid approach combining analytic and holistic assessment. *Language Testing in Asia*, *9*(13), 1–22. https://doi.org/10.1186/s40468-019-0087-6

Yang, H.-C. (2009). *Exploring the complexity of second language writers' strategy use and performance on an integrated writing test through structural equation modeling and qualitative approaches*. Unpublished doctoral dissertation. The University of Texas at Austin.

Yang, H.-C., & Plakans, L. (2012). Second language writers' strategy use and performance on an integrated reading-listening-writing task. *TESOL Quarterly, 46*(1), 80–103. https://doi.org/10.1002/tesq.6

Yu, G. (2013). From integrative to integrated language assessment: Are we there yet? *Language Assessment Quarterly, 10*(1), 110–114. https://doi.org/10.1080/15434303.2013.766744