ISSN 2522-6398 Vol. 3, Issue 5: 80-89, DOI: 10.25236/FER.2020.030517

Analysing students' written products in response to a TEM4 integrated task

Weilie Lu

Guangdong University of Foreign Studies, Guangzhou 510420, China
Email: 891285101@qq.com

ABSTRACT. *As integrated tasks are used in more and more testing contexts, it is necessary to investigate how test-takers approach such tasks. The present study aims to explore how English majors in the sophomore year perform in their written products in response to an interated task in TEM4. Analyses were conducted from the perspective of discource features, idea coverage the summary and the origin of ideas in the argement. Results showed that the two proficiency groups significantly differed only in grammatical accuracy, there were no significant difference between the two groups in terms of the other features: fluency and lexical sophistication. Besides, the two groups did not differ significantly in idea coverage in the summary, ideas used in the argument, ideas borrowed or ideas generated. The current study has implication for classroom teaching and test score interpretation. Directions for future research were recommended at last.*

KEYWORDS: *integrated task, written products, discourse features, summary, argument*

1. Introduction

Integrated test tasks, which require test takers to listen to, or read a stimulus, and then integrate the source information into their speaking or writing performances (Lewkowicz, 1997), is increasingly accepted in either classroom assessment or large-scale high-stakes tests throughout the world. For example, the TOEFL iBT contains integrated reading-listening-speaking and reading-listening-writing tasks, the Cambridge Assessment English's C1 Advanced and C2 Proficiency tests (Rukthong&Brunfaut, 2020) include reading-to-write tasks. The Canadian English Language Assessment (Jennings, Fox, Graves, & Shohamy, 1999) integrates listening, reading, and writing, as test takers write an essay that incorporates information from a lecture and readings. China's National Matriculation English Test, Guangdong version (NMET, GD) includes two integrated tasks: the listening-and-speaking task and the reading-to-write task, and the Test for English Majors band 4 (TEM4) also contains the reading-to-write task. It was indicated that integrated tasks could help students develop real-life communicative competence and cultivate students' multiliteracy, thus enhancing validity (Wesche, 1987; Lee, 2006). As a matter of fact, in language use in general, all the skills are often used more or less at the same time and even in teaching and testing, learners may be asked to read or listen to something before they start to interact with each other (Luoma, 2004, p.42). And in particular, writing in a university setting mostly involves integrating one's individual opinions textually with existing knowledge from sources (Ling Shi, 2004).

The popularity of integrated test tasks is easily understood and well-documented in literature. Besides being authentic, in that the task can assess abilities. corresponding to those performed outside testing situations (Rukthong&Brunfaut, 2020), it is also suggested that such a task type has some other advantages: 1) It is fairer to each candidate than an independent task. By providing candidates with source material, the integrated task reduces the impact of background knowledge, as test-takers can (partly) generate the content of their response from the input(Cumming et al., 2004;), thus test-takers are less likely to be disadvantaged due to lack of information on which to base their argument (Read, 1990;Weir, 1993); 2) It increases positive washback in classroom settings (Wesche, 1987; Esmaeili, 2002: Weigle, 2004). To elaborate on this aspect, Weigle (2004) cited an example of the introduction of integrated reading-writing test tasks in a university test of English for non-native speakers.Owing to the introduction of this test task, classroom instruction changed from teaching writing in isolation to focusing more on practising writing in combination with reading such as writing based on reading materials. Particularly, students were shown how to critically analyse source materials and appropriately integrate the texts into their writing. These abilities, as Weigle pointed out, are necessary not only to achieve success in the tests but also on academic courses (Rukthong&Brunfaut, 2020); 3) It provides better predictive capacity(Wesche, 1987); 4) It increases learner motivation, stimulation of students' learning of content

ISSN 2522-6398 Vol. 3, Issue 5: 80-89, DOI: 10.25236/FER.2020.030517

knowledge using multiple avenues (Wesche, 1987; Grant&Stoller, 2011).

The use of reading-to-write tasks for assessing academic writing in English is increasing, often replacing traditional impromptu writing-only tasks (L.Plakans, 2008), which is the case for the writing section of TEM4 in China. TEM4 was designed and developed for Engish majors to take in their sophomore year, with the aim of comprehensively checking whether the students meet the requirements specified in the syllabus. This test examines students' ability to use basic skills and their mastery of grammatical structure and word usage. It not only tests students' comprehensive ability, but also tests students' individual skills. At the same time, this test is also a means to evaluate teaching quality and promote inter-school communication. Since 2016, the old writing section in TEM4, which included both an independent writing task and a short note-writing task, has been replaced by an integrated writing task. Currently in this section, candidates will first read the direction, which instructs them how to proceed with the writing task. For example, the following is the direction from the 2019 TEM4 test paper:

Read carefullly the following excerpt and then write your responses in NO LESS THAN 200 WORDS, in which you should:

1) summarize the main message of the excerpt,and then

2) comment on Brewer’s view that parents should join in with their kids rather than limit their media consumption.

You can support yourself with information from the excerpt.

Marks will be awarded for content relevance, content sufficiency, organization and language quality. Failure to follow the above instructions may result in a loss of marks.

What follows after the direction is the source reading material or excerpt. Forty-five minutes are allotted for this part, within which candidates have to read and understand the source text, summarize the main ideas, take their positions, give their comments and opinions to argue for their positions. From the direction, we can know that to complete the task, students’ writing should contain at least two parts: the summary part---students first of all summarize the main ideas in the source material; and the argument part---students will use ideas/opinions to argument for the position they take. How do candidates perform in the summary part? Can they cover all the main ideas from the source material? How do candidates perform in the argument part? Do they borrow all the ideas from the source text or create their own ideas to support their argument? Asnswers to these questions can help understand how test takers approach integrated tasks, which is necessary to investigate as integrated tasks become more common in assessing writing for academic purposes (L.Plakans, 2009b). However, up till now, no such studies have been conducted on the integrated writing task in TEM4. Therefore, the present study is the first attempt to explore how candidates perform in response to the integrated read-to-write task in terms of discourse features and origin of ideas by analyzing their written products.

2. Literature Review

As integrated tasks become more common in assessing writing for academic purposes, it is necessary to investigate how test takers approach these tasks (L.Plakans, 2009). Tasks in different tests have different demands on production, so far the two most often investigated production demands are summary (Johns&Mayes,1990; Esmaeili, 2002; Ling Shi, 2004; Y. Asencio´n Delaney, 2008; Plakans&Gebril, 2013) and argumentative essays (S.C. Weigle, 2004; Ling Shi, 2004; Plakans, 2008,2009a,2009b; Gebril & Plakans, 2009,2013; Plakans&Gebril,2012; Weigle&Parker, 2012). To understand how test takers perform in integrated tasks, some researchers focused on analysing source text use (Johns&Mayes, 1990; Ling Shi, 2004; Cumming et al, 2005; Gebril & Plakans, 2009; Plakans&Gebril, 2012,2013; Weigle&Parker, 2012) while others focused on the analysis of discourse features (Cumming et al, 2005; Gebril & Plakans, 2009,2013). By analysing source text use, researchers may either want to explore issues relevant to textual borrowing or check whether test takers can transform the main ideas embedded in the source text into their summaries. By analysing discourse features, researchers may want to find out what features could significantly distinguish test takers of different proficiency levels, which might have implication for both teaching and test task development. Different production demands may require different abilities because they represent different degrees of cognitive complexity

In a study of the summary protocols of 'under-prepared' native-speaking university students, Johns (1985b) found that these students did not include all of the major propositions of the original in their protocols nor did they frequently combine idea units.Thus it follows that comprehension problems may lead to a failure to reproduce and combine some macro-propositions (main ideas) in summary protocols. Johns&Mayes (1990) compared idea units in summary protocols produced by university ESL students at two levels of proficiency.

ISSN 2522-6398 Vol. 3, Issue 5: 80-89, DOI: 10.25236/FER.2020.030517

Significant differences between the groups were found in two categories: replication of sentences from the original text, and combinations of idea units taken from two or more punctuated sentences in the original. In Ling Shi' (2004) study, subjects were from two different language backgrounds: 39 were native English speakers in a North American university and 48 were 3rd-year Chinese students learning English as a second language in a university in China. Half of the students in each group completed a summary task; the other half completed an opinion task. The study found that students who did the summary task borrowed more words than those who wrote the opinion essays, indicating that the summary task differs from the opinion task in terms of its dependence on source information. However, in this study, the researcher did not investigate whether there existed difference between students of different proficiency levels. Y. Asencio´n Delaney (2008) attempted to explore the construct of the the reading-to-write task. Participants performed two reading-to-write tasks—a summary and a response essay-- based on the same source text. Results indicated that the reading-to-write ability seems to be a unique construct weakly associated with reading for comprehension and disassociated from writing an essay without background reading support. In addition, it was found that language proficiency and educational level had a modest effect on the performance of the tasks. The writing task used in Plakans&Gebril's (2013) study was a comparative summary of listening and reading texts that present differing views on a topic. One of the areas the researcher focused on was the importance of source text ideas that writers included in their summary. It was found that high-scoring writers selected important ideas from the source texts and used the listening text as the task prompt instructed.

The reading-to-write construct cannot be conceived as a unitary ability, but rather as adynamic ability that interacts with task demands and individual factors (Y. Asencio´n Delaney, 2008). The ability to write a summary does not necessarily indicate an ability to write argumentative essays, which were the focused tasks in another line of study. In Ling Shi's (2004) study it was found that participants who wrote summaries used much more textual borrowing than those who wrote opinion essay. It is probable that such a difference occurred due to the different demands: in the summary task, test takers were directed to cover as many important ideas from the source text(s) as possible, while to complete the opinion task, they had to come up with ideas for their argument, textual borrowing might not help. In Plakans's (2009b) study, participants were required to write argumentative essays in the reading-to-write tasks developed for a university English placement exam. According to the results of the study, vocabulary knowledge and use, as well as stylistic concerns emerged as language difficulties for writers, indicating that vocabulary sophistication might be one factor that could distinguish students of high and low proficiency. Plakans (2009a) further found that writers from different proficiency levels used different strategies: higher scoring writers used more mining and global strategies. Plakans&Gebril (2012) attempted to answer questions with regard to writers' use of sources in their writing, the functions these sources serve, and how proficiency affects discourse synthesis. By using a mixed-method approach, the researchers found that source use served several functions including generating ideas about the topic and serving as a language repository and that score level affected text comprehension, especially at lower levels, but was not found to relate to the source use functions. S.C. Weigle, K. Parker(2012)explored the extent to which students borrowed source text language in an integrated reading/writing test. Results suggested that only a small percentage of students borrowed extensively from the source texts and that there were only minor differences in patterns of borrowing across topics, student groups and proficiency levels. Gebril&Plakans(2013) investigated the relationship between writing proficiency and discourse features in an integrated reading–writing task. Discourse features included fluency, lexical sophistication, syntactic complexity, grammatical accuracy, verbatim source use, and direct and indirect source use. Results showed that fluency was the only aspect that could differentiate all the three proficiency levels. Grammtical accuracy could only differentiated the Level 1 group and the two higher-level groups, the upper levels (level 2 and level 3) were not significantly different in this feature, based on which, researchers suspected that the writing at the higher levels might be distinguished by other aspects such as organization, content, or coherence, which were on the rating rubric. These findings confirmed what was found in Gebril & Plakans (2009).

Based on the results from the research conducted so far, we've known that higher proficiency writers produce longer essays (Cumming et al., 2005, 2006; Watanabe, 2001); grammatical accuracy is another factor that can significantly distinguish lower proficiency levels and higher levels(Plakans&Gebril, 2009, 2013); no statistically significant differences in lexical sophistication across the different proficiency levels (Plakans&Gebril, 2009, 2013). In terms of source text use, we've known that the source text can help writers generate ideas about the topic and serve as a language repository(Plakan&Gebril, 2012), thus students could borrow words or sentences from the source text(s). However, we haven't known how exactly source texts help candidates produce their ideas in support of their argument in their writing. Presented with the source text, do they just copy the ideas from the source material or do they generate their own ideas? The present study set on a further step to explore this aspect.

Most of the research reviewed above took TOEFL as the focus, no research so far investigates how

ISSN 2522-6398 Vol. 3, Issue 5: 80-89, DOI: 10.25236/FER.2020.030517

take-takers perform in the TEM4 integrated writing, which combines summary and argumentative writing into a composition. Understanding the characteristics of performances on integrated tasks at different proficiency levels will provide the field of language assessment and second-language (L2) writing with better interpretations of test scores (Chapelle, Enright, & Jamieson, 2008). Specific to the TEM4 integrated writing, first, we want to explore how test takers perform in the summary part---can they transform all the important information from the source text into the summary? Then, we want to explore how test takers perform in the argument part---do they produce their own ideas or do they just copy the ideas directly from the source text to support their argument? At last, we want to know how test takers perform in their writing in terms of discource features, including fluency, lexical sophistication and grammatical accuracy.

3. Method

3.1 Integrated writing task

The integrated writing task used in this study is the practice exercise for students preparing for the TEM4 (see Appendix A). In order to make students take the task seriously and put all their efforts to complete the task, just like how they do it in the real TEM4, the integrated writing task was embedded in the 2-hour final examination of the course called *Integrated English 4* at the end of the fourth semester. Besides this integrated writing task, the final examiniation paper for this course contained other test item types like Multiple-Choice vocabulary&grammar, reading comprehensionand and sentence paraphrasing etc. Of the two-hour-long final exam, only 30 minutes were alloted for the integrated writing part. In this task, students first read the direction, in which the key topic question *Should we restrict automobile in modern life?* was presented, then they read two source texts: one takes the positive position: we should restrict automobile in modern life and the other source text takes the negative position: we shouldn't restrict automobile in modern life. As required by the direction, students should first summarize the main ideas in the two source texts and take their positions. Then students should give their opinions in support of their argument. Altogether 30 students' written products from a whole class were collected for analysis. Based on their TEM4 scores, the students were divided into lower-level group(14, those who did not pass the TEM4) and higher-level group(16, those who passed the TEM4). Table 1 is the result of the TEM4 scores between the two groups.

Table 1: TEM4 scores between the two groups

	N	Mean	Std.Deviation	sig.
Lower-level group	14	53.57	3.32	
Higher-level group	16	65.00	4.38	.000
Total	30	59.67	6.96	

3.2 Analysis

3.2.1 Discource features

Discource features contained in this study for analysis include lexical sophistication, fluency and grammatical accuracy and the method of counting or coding these features followed the method adopted in Gebril & Plakans(2009, 2013). Lexical sophistication is mainly about the average word length. In this study we adopted the definition used by Cumming et al. (2005: 9) for average word length, "the number of characters divided by the number of words per composition." Microsoft Word was employed in this study to calculate the targeted vocabulary measure(Gebril&Plakans, 2009). Fluency is common in most studies of writing features, and it has proved successful in differentiating proficiency levels in many studies of second language writing(e.g., Cumming et al., 2005; Tedick, 1990). Following the previous research, fluency in this study was determined through word count, which can also be easily determined with the help of Microsoft Word software. To indicate grammatical accuracy, which has been a challenge in research in that it is difficult both to find precise objective measures of accuracy and to gain rater agrement (Polio, 1997), in this study, a scale from 1 to 3 was adapted from the research of Cumming et al. (2005) and Hamp-Lyons and Henning (1991). The grammatical accuracy scale has three levels:

1) Many severe errors, often affecting comprehensibility

2) Some errors but comprehensible to the reader

ISSN 2522-6398 Vol. 3, Issue 5: 80-89, DOI: 10.25236/FER.2020.030517

3) Few errors and comprehensibility seldom obscured for the reader

In this study, two raters read the written products holistically, and provide a score (1,2. or 3) representing different degree of grammatical accuracy.

The interrater reliability of r = .91 was achieved after training, which is a satisfactory value for the present research.

3.2.2 Ideas summarized

In the present integrated writing task, students were presented with two source texts and they are required to summarized the main ideas from these two source texts. To establish a standard number of main ideas in the source material, two experienced teachers were invited to read and extract the main ideas from the two texts. After discussion, they agreed with each other and at last 6 key ideas were extracted, 3 from each source text. By comparing the ideas summarized by students and those extracted by the two teachers, we can know the number of ideas they contained in the summary part, each idea covered was awarded one mark, therefore, in the summary part, students can get 6 marks in total if they cover all the 6 key ideas in their summary.

3.2.3 Origin of ideas

After summarizing the main ideas and taking their positions (showing whether they support the idea to restrict automobile or not), students had to give their opinions to support their argument. We first count the number of ideas in students written prudcuts. Then we compare students' ideas against those in the source texts(6 key ideas had been extracted from the two source texts), so we can know the number of ideas borrowed from the source text and the number of ideas created by students.

4. Results

4.1 General results

Although students were not directed to give a title for their writing, most students did. Of all the 30 students, 26 students had a title for their writing, and the length of the titles ranged from 1 word to 8 words, with 7 words being the mode(15 students), and the 7-word titles were similar in wording like *Should We Restrict Automobile in Modern life?* or use its positive or negative form, which students borrowed from the direction in this task. Of the 4 students who didn't write a title, two belonged to the lower-level group, two belonged to the higher-level group.

As to the structure of students' writing in response to the integrated task, most students (29 of the 30 students) contained three parts in their written products. In the first part, students summarized the main ideas in the two source texts and then immediately took their positions showing whether they agreed to restrict automobile or not. For all the students, the first part is the first paragraph. The second part is the argument part, in which students had to think of ideas to argue for their stance. Most students used three ideas, some ideas were borrowed from the source texts, some were created by students themselves. In this part some students wrote several paragraphes while others wrote only one paragraph. The last part (the third part) is the conclusion, in which writers restated the position they took. Of the 30 students, only one student, who belonged to the lower-level group, finished the composition without the conclusion part. After carefully examing the written products, we could find that for the three parts of their composition, most students(27) followed the 2—1—3 structure, meaning that the second part (argument for the position) was the longest, the first part (summary) was the second longest and the last part (the conclusio) is the shortest. We called this pattern the normal structure, and the other structures that did not follow this pattern are abnormal structures. Of the 30 students, only three students did not follow the normal pattern, the words contained in each part were 101-87-21(student No.23), 71-33-23(student No. 21), and 97-130-0(student No.18), all of these students who did not follow the normal pattern belonged to the lower-level group.

From Table 2, we can know that some students summarized only one main idea while some others could extract all the 6 key ideas from the two sources and that some students borrowed 3 ideas from the source text and some others created no ideas of their own. Beside, we can also get some general information about fluency, lexicial sophistication, grammatical accuracy and total number of ideas in argument.

ISSN 2522-6398 Vol. 3, Issue 5: 80-89, DOI: 10.25236/FER.2020.030517

Table 2: Overall Descriptive Statistics

	Fluency	Lexical Sophistication	Grammatical accuracy	Ideas in Summary	Ideas in argument	Ideas borrowed	Idea created
Minimum	134.0	4.68	1.00	1.00	1.00	.00	.00
Maximum	280.0	5.53	3.00	6.00	3.00	3.00	3.00
Mean	221.0	5.18	2.30	4.00	2.67	1.13	1.53
S.D.	34.3	.23	.70	1.46	.55	.86	.86

4.2 Discourse features

Discourse features analysed in this study are composed of three categories: fluecy, which refers to the length of students' composition, lexical sophistication, which is the same as the average word length and grammatical accurracy, which is indicated in terms of 1 (Many severe errors, often affecting comprehensibility), 2 (Some errors but comprehensible to the reader) or 3 (Few errors and comprehensibility seldom obscured for the reader).

According to the means in Table 3, we can see that of all the three categories, the higher-level group achieved higher fluecy(227.06 v.s. 214.14), higher lexical sophistication(5.25 v.s. 5.11) and higher grammatical accuracy(2.75 v.s. 1.79) than the lower-level group. However, the results of the independent samples Mann-Whitney U Tests showed that grammatical accuracy was the only category that could significantly distinguish the lower-level group and the higher-leve group. However, there were no significant differences between the two groups in terms of fluency and lexical sophistication. Such a result is different from that in studies conducted by Gebril & Plakans (2009, 2013). In their studies, fluency was the the only category that could significantly distinguish the three different level groups.

Table 3: Between-group comparison on discource features

Categories	group	Mean	Standard deviation	sig
Fluecy	L-group (n=14)	214.14	35.21	.289
	H-group (n=16)	227.06	33.36	
Lexical sophistication	L-group (n=14)	5.11	.23	.124
	H-group (n=16)	5.25	.22	
Grammatical accuracy	L-group (n=14)	1.79	.58	.000
	H-group (n=16)	2.75	.45	

Note: The significance level is .05.

4.3 Ideas in the summary

In the first part of the writing, students were directed to summarize the main ideas in the two source text. Therefore, the number of main ideas students could extract from the source texts might imply whether they could understand the reading material or not. Since we expect the higher-level group to understand the reading material better, we naturally expect students in this group to summarize more key ideas than those in the lower-level group. According to the mean in Table 4, we can see that the higher-level group did summarize more ideas than the lower-level group (4.25 v.s. 3.71). However, the two groups did not differ significantly in terms of ideas summarized according to the significant level of the independent samples Mann-Whitney U Test (sig.= .297).

Table 4: Between-group comparison on ideas summarized

Categories	group	Mean	Standard deviation	sig
Ideas summarized	L-group (n=14)	3.71	1.33	.297
	H-group (n=16)	4.25	1.57	

Note: The significance level is .05.

4.4 Ideas in the argument

In the first part of students' writing, after they summarized the main ideas, they showed their stance on whether they supported or opposed the idea of restricting automobile in modern life. Then in the second part, they had to argue for their stance. In this section, we aim to explore how many ideas students used in their argument. We also want to know whether they borrowed the ideas from the two source texts or they could

ISSN 2522-6398 Vol. 3, Issue 5: 80-89, DOI: 10.25236/FER.2020.030517

generate their own ideas, since the source text they were presented with also contained ideas or opinions. Results in Table 5 show that on average, the higher-level group contained more ideas in their argument part (2.81 v.s. 2.50) and borrowed more ideas from the source texts (1.31 v.s. .93), but created fewer ideas of their own than the lower-level group(1.50 v.s. 1.57), indicating that there was no connection between language proficiency and the ability to generate ideas. Further, the independent samples Mann-Whitney U Test showed that the two groups did not significantly differ in terms of ideas in argument, ideas borrowed or ideas created (sig=.138, sig.=.276, sig.=.860).

Table 5: Between-group comparison on origin of ideas

Categories	group	Mean	Standard deviation	sig
Ideas in Argument	L-group (n=14)	2.50	.65	.138
	H-group (n=16)	2.81	.40	
Ideas borrowed	L-group (n=14)	.93	.73	.276
	H-group (n=16)	1.31	.95	
Ideas Created	L-group (n=14)	1.57	.64	.860
	H-group (n=16)	1.50	1.03	

5. Discussion

Integrated English for academic purposes speaking and writing tasks involve complex texts requiring test takers to engage cognitive skills which extend beyond language proficiency skills. These cognitive skills include identifying, selecting, and combining relevant information from academic texts into oral and written performances, recognizing key relationships between source text ideas, and organizing and transforming relevant content (Brown, Iwashita, & McNamara, 2005). In this study, besides analysing the discourse features, including fluecy, lexical sophistication and grammatical accuracy, we also explore how students performed in summarizing the main ideas and how they came up with ideas to argue for their stances: whether they borrowed ideas from the source material or created ideas of their own. After conducing this study and analyzing the data, we extend our understanding of how students approached the integrated writing task in TEM4, on which so far no research has been conducted except the present one.

The general result at the beginning of the previous section showed that although students were not directed to write a title for their writing, most students (26 of the 30 students) composed their writing with a title. This may be related to their writing habbit. For most students, giving a title to a composition is not something that is required, it is something a complete composition should have. And of the 4 students who did not write a title, 2 belonged to the lower-level group and 2 belonged to the higher-level group, indicating that whether writing a title or not is not a factor to differentiate students of different language proficiency. Another finding is that most students' compositions (27 of the 30 compositions) followed the 2—1—3 normal structure, indicating that for these English majors, in their mind they had a standard of what a qualified composition should look like. This may be the result of classroom teaching or writing training. Students in S.C. Weigle's (2004) study were familiar with the five-paragraph essay format, and in the present study, students are used to the three-part structure with a normal pattern. Of the 3 students who did not follow the normal pattern, all belonged to the lower-level group, implying that following the normal writing struture may be a threshold for a qualified composition.

In the first part of the composition, students had to summarize the main ideas embedded in the two source texts. On the whole, students differed in the number of ideas they could cover in this part. Some only summarized one idea while some others could cover all the 6 main ideas. However, further investigation indicated that the lower-level group and the higher-level group did not differ significantly in the number of ideas they summarized. This finding is different from what was found in Johns&Mayes's(1990) study, in which summary protocols produced by university ESL students at two levels of proficiency significantly differed in the combinations of idea units taken from the original. The finding also differed from from that in Plakans&Gebril's(2013) research, in which researchers found that compared with low-scoring writers, high-scoring writers selected important ideas from the source texts. The occurrence of such difference may be due to the easiness of the two source text in terms of language features and topic. The topic of the two source texts is about automobile with opposing positions, which is familiar to each student, leading to their correct understanding of the texts, therefore, language proficiency is not a key factor to determine whether one could extract main ideas from the source texts or not.

In the second part, students had to argue for the position they took. Most students (21 of the 30 students) used 3 ideas, 8 students used 2 ideas, while one student belonging to the lower-level group used only one idea to argue for his/her position. When it comes to the comparison of the two groups, no significant difference exists in

ISSN 2522-6398 Vol. 3, Issue 5: 80-89, DOI: 10.25236/FER.2020.030517

the number of ideas used in this part. And further investigation shows that the two groups did not significantly differ in either ideas borrowed or ideas generated, which contradicts our earlier expectation. Before we conducted this study, we had known from literature that lower-level students borrowed more language (words or phrases) from source texts (Ling Shi, 2004; Cumming et al., 2005). Therefore, we had expected that lower-level students might borrowed more ideas from the source texts and created fewer ideas of their own. The finding in the present study indicated that for the present integrated writing task, the ability to create ideas may not be the factor to determine language proficiency. After carefully examining the data, we found a surprising contrast: in the lower-level group, each student could at least generat one idea of their own, while in the higher-level group, three students did not generate any ideas of their own, the three ideas the presented in the argument were all borrowed from the source texts. For these students, they might be clearly aware that they were taking a language test, correct language use was the priority, borrowing ideas from the source text might help them escape some mistakes in grammar or expression on the one hand, on the other hand, they could use the time others students used to generate ideas to focus on checking or improving language. This is an issue that both teachers and students have to pay attention to. As university English majors, language learning is of course a priority, but creativity in ideas is also important for the growth and development of each college student.

Analysis of discourse features was conducted in terms of fluency, lexical sophistication and grammatical accuracy. And grammatical accuracy was found to be the only feature that significantly differed between the lower-level group and the higher-level group. The two groups did not significantly differ in either fluency or lexical sophistication. The results in this study confirm the finding in Gebril & L. Plakans's (2009,2013) studies, in which the two groups did not significantly differ in lexical sophistication. What is different from their studies is that in their studies, the three different prodiciency groups significantly differed in fluency, and grammatical accuracy was found be significantly different between Level 1 and the other two higher levels, the level-2 group and the level-3 group did not differ significantly in grammatical accuracy, implying that these two higher groups might be distinguished by other features. The difference in findings between the present study and previous studies may due to the time limit. In this study, only 30 minutes were allotted for this task, with such a short time, students had to rush to write their composition with the number of words required by the task, so even students with higher proficiency might not be able to elaborate on this task. Besides, the clear requirement of a 200-word composition may also contribute to the insignificant difference between the two groups, because everyone was clearly aware of the length demand, during their writing, they would try to meet such a requirement.

In Plakans's (2009) study, it was found that language difficulties that emerged for writers were vocabulary knowledge and use, as well as stylistic concerns (L.Plakans, 2009). The present study shared such a finding. After examining their written products, we found that some students did not complete the task well just because they did not understand certain words. For example, in student No.7's writing, the position she/he took was that "we should restrict automobile". However, the three ideas presented to argue for the position were 'First of all, automobile meets our needs in our daily life' 'Secondly, automobile helps a lot in economic development' 'Lastly, electronic cars are mature enough to use now, and its technology can become better in time', all of which contradicted the position she/he took. The most possible reason for such contradiction was the misunderstanding of the word 'automobile'. Although lexical sophistication did not significantly differ between the two groups, knowledge of vocabulary did affect students' understanding of the source texts, which further affected their writing quality.

6. Conclusion

From the present research we extend our understanding of how students perform in response to an integrated task. We know that most students would write a composition with a title although there was no such a requirement in the direction. We know that students generally would follow a normal structure when completing the integrated writing in TEM4. The two groups differed significantly only in grammatical accuracy, and there existed no significant difference between the two groups in terms of fluency, lexical sophistication, the number of ideas summarized, ideas borrowed or ideas generated.

What deserves our attention is that some students in the higher-level group did not create their own ideas at all, they just borrowed the ideas to argue for their positions. Was it because the students did not have the ability to come up with their own ideas or was it because they just wanted to play safe---borrowing ideas directly from the source text may help them escape some language mistakes, leading to high marks in this part. These students may know clearly that as language learners, using language correctly is the priority, and they might also know from their learning experience that what raters would focus on was language quality, borrowing ideas from the source texts would not lead to any loss in marks awarded. Actually from the direction for this task, students were told that "Marks will be awarded for content relevance, content sufficiency, organization and language quality." ,

ISSN 2522-6398 Vol. 3, Issue 5: 80-89, DOI: 10.25236/FER.2020.030517

in which whether ideas should be created or borrowed was not mentioned. This has something to do with the construct of this task, should the ability to generate ideas be included as part of the construct? If yes, then maybe in the direction, students should be told clearly that they shouldn't borrowed all the ideas from the source texts. If yes, then in classroom teaching, teachers should teach students how to come up with ideas based on source text. If yes, then in rating, raters should not only focus on language quality, whether borrowing ideas or creating ideas may lead to difference in marks.

In this study, we try to explore how the two groups with different language proficiency---the lower-level group and the higher-level group---performed in response to the integrated task in TEM4. We analysed and compared their written products from the perspectives of fluency, lexical sophistication, grammatical accuracy. As we know, there are other discourse features like syntactic complexity that were not covered in the present study, therefore, to extend our understand, future research can be conducted by including more discourse features. As to the function of source texts, in the present study, we just focused on ideas summarized and ideas borrowed, we didn't include textual borrowing as our research focus, further research could be done by incorporating this aspect. Sample size is also one limitation. Future studies could ask more students to participate and used different research methods. In this study, we did not ask raters to award marks to each written composition, therefore, we could not know the direct relationship between the discourse features and the marks they got, which could be another direction for future research.

References

[1] Asención, Y. (2008). Investigating the reading-to-write construct. Journal of English for Academic Purposes, 7, 140–150.

[2] Brown, A., Iwashita, N., & McNamara, T. (2005). An examination of rater orientations and test-taker performance on English-for-Academic-Purposes speaking tasks (TOEFL Monograph Series MS29). Princeton, NJ: Educational Testing Service.

[3] Chapelle, C., Enright, M. & Jamieson, J. (2008). Score interpretation and use. In C. Chapelle, M. Enright, & J. Jamieson (Eds.), Building a validity argument for the Test of English as a Foreign Language (pp. 1–25). New York, NY: Routledge.

[4] Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. (2004). A teacher-verifica-tion study of speaking and writing prototype tasks for a new TOEFL. Language Testing, 21, 159–197. doi:10.1191/0265532204lt278oa

[5] Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in independent and integrated prototype tasks for the next generation TOEFL. Assessing Writing, 10, 5–43.

[6] 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 tasks for the new TOEFL (TOEFL Monograph No.MS–30 Rm 05–13). Princeton, NJ: ETS.

[7] Esmaeili, H. (2002). Integrated reading and writing tasks and ESL students' reading and writing performance in an English language test. The Canadian Modern Language Review, 58, 599–622. doi: 10.3138/cmlr.58.4.599

[8] Gebril, A., & Plakans, L. (2009). Investigating source use, discourse features, and process in integrated writing tests. Spaan Working Papers in Second or Foreign Language Assessment, 7,47–84.

[9] Gebril, A., & Plakans, L. (2013). Toward a transparent construct of reading-to-write tasks: the interface between discourse features and proficiency. Language Assessment Quarterly, 10, 9–27.

[10] Grant, L., & Stoller, F. (2001). Reading for academic purposes: Guidelines for the ESL/EFL teacher. In M. Celce-Murcia (Ed.), Teaching English as a second or foreign language (Third ed.) (pp. 187-203). Boston, MA: Heinle & Heinle.

[11] Hamp-Lyons, L., &Henning, G. (1991). Communicative writing profiles: An investigation of the transferability of a multiple-trait scoring instrument across ESL writing assessment contexts. Language Learning, 41, 337–373.

[12] Jennings, M., Fox, J., Graves, B., & Shohamy, E. (1999). The test-takers' choice: An investigation of the effect of topic on language-test performance. Language Testing, 16, 426–456.

[13] Johns, A. M. 1985b. 'Summary protocols of "under-prepared" and "adept" university students: replications and distortions of the original.' Language Learning 35/4: 497-517.

[14] Johns, A., & Mayes, P. (1990). An analysis of summary protocols of university ESL students. Applied Linguistics, 11, 253–271.

[15] Lee, Y.-W. (2006). Dependability of scores for a new ESL speaking assessment consisting of integrated and independence tasks. Language Testing, 23, 131–166.

ISSN 2522-6398 Vol. 3, Issue 5: 80-89, DOI: 10.25236/FER.2020.030517

[16] Lewkowicz, J. (1997). The integrated testing of a second language. In C. Clapham, & D. Corson (Eds.), Encyclopedia of language and education (Vol. 7, pp. 121–130).
[17] Dordrecht, The Netherlands: Kluwer Academic.
[18] Luoma, S. (2004). Assessing speaking. Cambridge: Cambridge University Press.
[19] Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. Assessing Writing, 13, 111–129.
[20] Plakans, L. (2009a). Discourse synthesis in integrated second language writing assessment. Language Testing, 26, 561–587.
[21] Plakans, L. (2009b). The role of reading strategies in integrated L2 writing tasks. Journal of English for Academic Purposes, 8, 252–266.
[22] Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. Assessing Writing, 17, 18–34.
[23] Polio, C. (1997). Measures of linguistic accuracy in second language writing research. Language Learning, 47, 101–143.
[24] Read, J. (1990). Providing relevant content in an EAP writing test. English for Specific Purposes, 9, 109–121.
[25] Rukthong & Brunfaut, (2020). Is anybody listening? The nature of second language listening in integrated listening-to-summarize tasks. Language Testing, 37(1), 31–53.
[26] Shi, L. (2004). Textual borrowing in second-language writing. Written Communication, 21, 171–200.
[27] Tedick, D. J. (1990). ESL writing assessment: Subject-matter knowledge and its impact on performance. English for Specific Purposes, 9, 123–143.
[28] Watanabe, Y. (2001). Read-to-write tasks for the assessment of second language academic writing skills: Investigating text features and rater reactions. Unpublished doctoral dissertation, University of Hawaii.
[29] Weigle, S. (2004). Integrating reading and writing in a competency test for non-native speakers of English. Assessing Writing, 9, 27–55.
[30] Weigle, S., & Parker, K. (2012). Source text borrowing in an integrated reading/writing assessment. Journal of Second Language Writing, 21, 118–133.
[31] Weir, C. (1993). Understanding and developing language tests. London: Prentice Hall.
[32] Wesche, M.B. (1987). Second language testing: The Ontario test of ESL as an example. Language Testing, 4, 28–47.

Appendix A

Writing (15') [30 MINS]

Directions:

The automobile is regarded as one of the greatest inventions in modern life. However, people also think it has caused serious problems. Should we restrict automobile in modern life? This has been an intensely discussed question for years. The following are the supporters' and opponents' opinions. Read carefully the opinions from both sides and write your response in about 200 words, in which you should first summarize briefly the opinions from both sides and give your views on the issue.

Marks will be awarded for content relevance, content sufficiency, organization and language quality. Failure to follow the above instructions may result in a loss of marks.

Write your essay on the ANSWER SHEET.

Yes	No
With the invention of the automobile, the age of transportation shifted into another gear. Quickly it became possible for people to travel more comfortably and conveniently to destinations near and far, and the figurative world moved closer together. Trucks carried cargo across countries and soon became serious competition for trains and ships. As a result, food and other consumer goods have become available even in remote areas, overall living standards have improved, and the automobile industry, which has grown fantastically over the course of the past century, employs millions of workers worldwide.	As societal reliance and global economic dependence have grown together with the automobile industry, many significant problems have surfaced. Car and truck exhausts pollute the air in metropolitan areas around the world and thus create serious health problems. The continued use of fossil fuel engines and the scarcity of oil have led to much political strife and even war, particularly in the oil rich region of the Middle East. As the powerful automobile industry remains reluctant and has yet to successfully promote an engine type that does not rely on gasoline power, the problem of fossil fuel shortage will become an even more serious problem.